J Forensic Sci, May 2008, Vol. 53, No. 3
doi: 10.1111/j.1556-4029.2008.00705.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Angi M. Christensen,[1] Ph.D.; and Adam D. Sylvester,[2] Ph.D.

Physical Matches of Bone, Shell and Tooth Fragments: A Validation Study*

ABSTRACT: The following study examines the reliability of physically matching fragments of bone and other mineral-based biological materials such as shells and teeth. Participants with varying education, training, and experience were asked to complete a matching exercise consisting of intentionally fragmented specimens. Success rates were very high; the positive association (correct match) rate was 0.925, while the nonassociation (overlooked match) rate was 0.075, and negative associations (incorrect matches) occurred at a rate of just 0.001. Results also indicate that those with more education and related experience tended to have higher positive association rates, although not significant statistically. Experienced osteologists, however, completed the matching exercise in significantly less time. Low error rates among both experienced and inexperienced individuals support the reliability and validity of performing physical matches of these materials, and suggest that performance may also be related to an individual's aptitude for spatial tasks or other factors.

KEYWORDS: forensic science, forensic anthropology, physical match, *Daubert*, error rate

Physical matches are routinely used in forensic examinations as a way to determine whether two or more pieces of evidentiary material such as glass, paper, metal, paint, plastic, wood, tape, and fabric were formerly one piece of material. Bone fragments are often physically matched in the reconstruction of skeletal elements as part of forensic anthropological examinations, as well as in paleoanthropological and archaeological contexts. Although routinely performed and widely regarded as intuitively evident, the reliability of physically matching fragments of bone and other mineral-based biological materials such as shells and teeth has never been empirically tested. While physical matches of conjoinable fragments are a significant concern of archaeologists (1) and computer programs have even been designed to assist in the reconstruction process [see (2) for an example], no tests have been performed to estimate the probability of an incorrect match. As a forensic matter, a well-designed study addressing this issue may be beneficial given the Daubert guidelines regarding scientific testing and documented error rates (3,4). The following study examines the reliability and validity of physically matching bone, shell, and tooth fragments, as well as the influence of experience level.

Materials and Methods

The fragments used for this study included human bones, nonhuman bones, nonhuman teeth, turtle shells, and mollusk shells. Fragments were created by deliberately fracturing larger specimens using a combination of static and dynamic loading until structural failure. In most cases, specimens were secured using a C-clamp and tapped with increasing force with a hammer until failure. In other cases, the specimens were tightened in the C-clamp until failure. It is recognized that fracture dynamics may be somewhat different in the "green" versus "dry" state (5), but trauma such as cranial fracturing has been reconstructed in the dry state (or at least de-fleshed state) in many forensic cases. It is thus assumed here that fracturing the specimens while already in the dry state has little to no negative impact on the extrapolation of the results to cases where fracturing occurred in the green state.

Fifty-seven of the fragments generated were used in the study. Each fragment was labeled with a randomly assigned number between 1 and 100, and all pairs of matching fragments were recorded. The matching exercise contained a total of 40 correctly matching edges, and six fragments with no corresponding match. The exercise was administered, as pictured in Fig. 1, to individuals with varying levels of education, experience, and training in osteology and physical matching. Participants were instructed to identify, and affix together with tape (provided), all physical matches they believed to be present among the fragments, and were advised that some fragments may have multiple matches, while others may have no matches.

In addition, participants were asked to answer questions pertaining to their expertise or education, previous experience performing physical matches, any education or training they have had in osteology, and criteria used to identify matches. They were also asked to record the time required to complete the exercise. Completion of the exercise was defined by the participant as the point at which they believed all possible physical matches had been identified and affixed, or when they simply decided to stop.

Upon completion of the matching exercise, the test was scored by one of the authors by disassembling the test while recording the correctly matched pairs and errors. A *positive association* refers to a correct match, a *negative association* refers to an incorrect match, and a *nonassociation* refers to an overlooked or rejected match. The term "false negative" was intentionally avoided, because it has a rather specific meaning in a forensic context, indicating that an examiner compared two fracture edges and concluded that they either did not match or that there was insufficient information to

[1]Laboratory Division, Federal Bureau of Investigation, 2501 Investigation Parkway, Quantico, VA 22135.
[2]Center for Musculoskeletal Research, 301 Perkins Hall, The University of Tennessee, Knoxville, TN 37996.
*Presented in oral format at the 59th Annual Meeting of the American Academy of Forensic Sciences, February 19–24, 2007, in San Antonio, TX.

Received 18 March 2007; and in revised form 13 Oct. 2007; accepted 13 Oct. 2007.

FIG. 1—*Fragments in matching exercise.*

make a conclusion, when in fact the two pieces match. The authors did not watch participants performing the exercise, so the basis for failing to associate a match (overlooked or rejected) is not known. A nonassociation was therefore defined simply as the failure to affix a fracture match, regardless of the reason.

The positive association rate was calculated as the number of correct matches over the total number of possible correct matches. The nonassociation rate was calculated as the number of unassociated matches divided by the total number of possible correct matches. For these two rates, the denominator was thus always 40. In probability terms, the question of interest was: what is the probability of correctly identifying (or overlooking) a match given that it exists in the exercise? The negative association rate was calculated as the number of incorrect matches divided by the total number of matches identified; thus, the denominator changes depending on the number of associations the participant made. In probability terms, the question of interest was: what is the probability of an incorrect match given that a match was made?

Ninety-six individuals participated in the study, with backgrounds varying from no higher than high school education and no professional forensic experience, to forensic scientists with decades of experience, and board certified forensic anthropologists. One participant was eliminated from analysis because it was strongly suspected that the instructions provided with the exercise were not properly followed, leaving 95 analyzed participant results. Participants were analyzed by two experience categories (Fig. 2): experience in osteology and experience in performing physical matches. Osteology experience was broken down into those with: no experience, some experience, and expert. The "some experience" group includes those who have had any coursework, training, or professional experience in osteology or physical anthropology, and the "expert" group includes those holding or currently pursuing doctoral degrees in physical anthropology or who are employed as osteologists or physical anthropologists. For physical matching experience, because no participants in the study solely and routinely perform physical matches, no expert category was assigned and participants were grouped by those with no experience and those with some or any professional experience.

To ensure that the two grouping classes are unrelated (i.e., that membership in the osteology group was not affected by membership in the physical match group), we performed a chi-squared test of independence on the grouping variables. To compare positive association rates, we performed separate ANOVA tests, using each of the two grouping variables to divide the sample. In addition, because it appeared that the positive association rates were not normally distributed, we also compared groups using the nonparametric Kruskal–Wallace test. Exercise completion time for each participant was compared using ANOVA tests. Significance tests comparing osteology groups were followed by multiple tests to isolate differences between individual groups. The significance level for all tests was set at $\alpha = 0.01$.

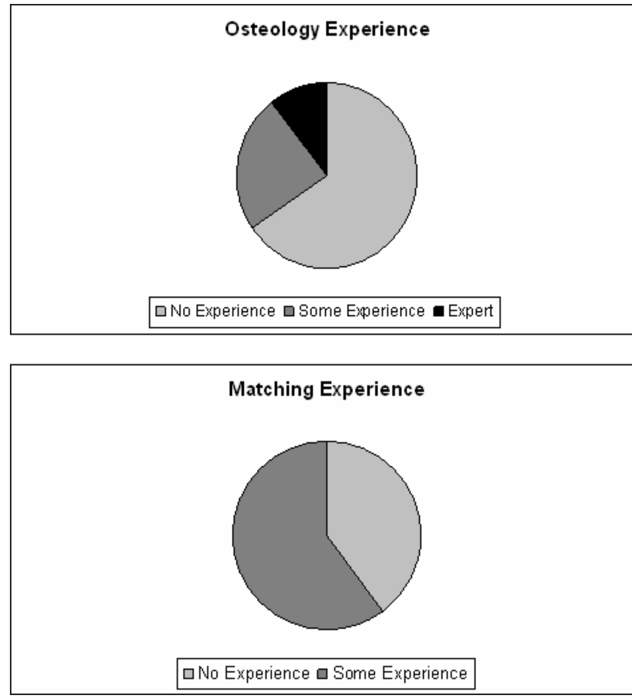

FIG. 2—*Participant experience.*

Results

The positive association rate among the pooled 95 participants was 0.925, with a highly skewed, no normal distribution, because of the large number of perfect or nearly perfect scores on the exercise (Fig. 3). Error rates did show a trend between groups based on osteology experience. On average, experts performed better than those with some experience, and those with some experience performed slightly better than those with no experience (Table 1). These differences, however, were not statistically significant (ANOVA $p = 0.73$, Kruskal–Wallace $p = 0.37$). For matching experience, those with some experience performed slightly better than those with no experience (Table 2), but again, the differences were not

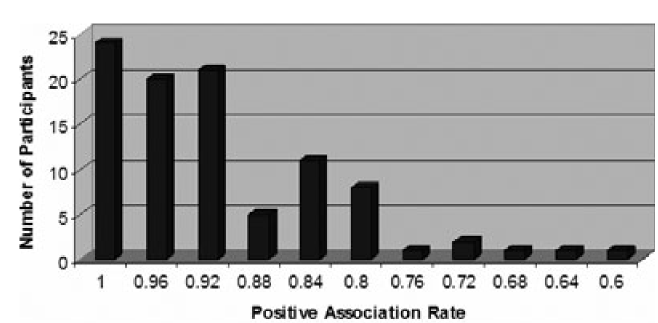

FIG. 3—*Nonassociation rates (all participants).*

TABLE 1—*Positive associations by osteology experience.*

Group	Positive association rate	SD
Expert (*N* = 10)	0.945	0.093
Some experience (*N* = 23)	0.924	0.083
No experience (*N* = 62)	0.923	0.081

ANOVA: *F*-value = 0.32, *p*-value = 0.73.
Kruskal–Wallace: x^2 = 1.95, *p*-value = 0.38.

TABLE 2—*Positive associations by physical matching experience.*

Group	Positive association rate	SD
Some Experience (*N* = *57*)	0.93	0.081
No Experience (*N* = *38*)	0.918	0.084

ANOVA: *F*-value = 0.44, *p*-value = 0.51.
Kruskal–Wallace: x^2 = 0.39, *p*-value = 0.53.

statistically significant (ANOVA $p = 0.51$, Kruskal–Wallace $p = 0.53$).

Figure 4 illustrates several commonly identified correct matches. Figure 4*a* shows three correctly matched bone fragment pairs that were identified by almost all participants. Figure 4*b* shows one of the correct matches between two mollusk shell fragments. In addition to the fit of the fracture edges, the match can be further tested by observing the microscopic structure of the shell on the inside of the fractured edge, which can be seen using reverse lighting and a comparison microscope. (Note that microscopic views are just for illustration; participants were not provided a microscope to perform the exercise.) Figure 4*c* shows a similar example of a tooth fracture match.

The single most common nonassociation was a human distal fibula and shaft fragment, which was missed by 44% of all participants (Fig. 5*a*). There was, however, a large difference in the success rate of this particular match by osteology experience. Whereas, nearly half of the nonexpert osteologists (the combined "some experience" and "no experience" groups) missed this match, this match was missed only once among expert osteologists. It is possible that this is because the experts recognized the fragments as originating from a specific human bone. The difference in the overall success rate for matches of human materials between expert and nonexpert osteologists, however, was not statistically significant. Among expert osteologists, the positive association rate for human material was 0.9313, while among nonexperts, the positive association rate for human material was 0.9309 (ANOVA $p = 0.99$; Kruskal-Wallis $p = 0.554$).

Among the most frequent nonassociations were matches between the four fragments of a nonhuman rib (Fig. 5*B*). Participants often succeeded in pairing up the four pieces into two halves (Fig. 5*B* a-to-b and c-to-d), but failed to fully assemble the rib, as shown in Fig. 5*B* bottom right. Thirty-eight per cent of nonexpert osteologists, and 30% of expert osteologists missed one or more of these matches, which is not significantly different (x^2 = 0.2245, $p = 0.6354$).

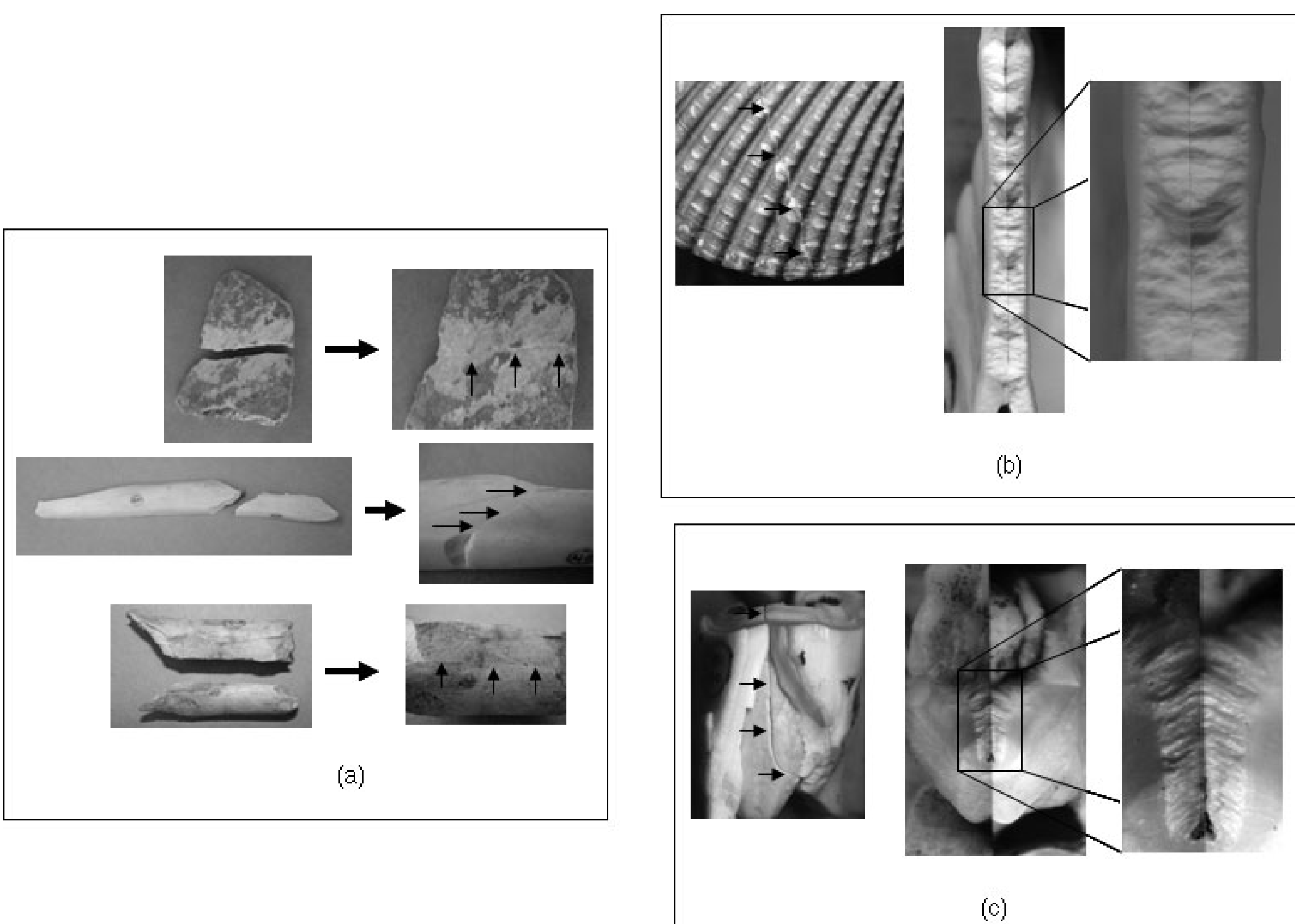

FIG. 4—*Examples of correct associations. (a) Bone matches (b) Shell match macro- and microscopically (c) Tooth match macro- and microscopically.*

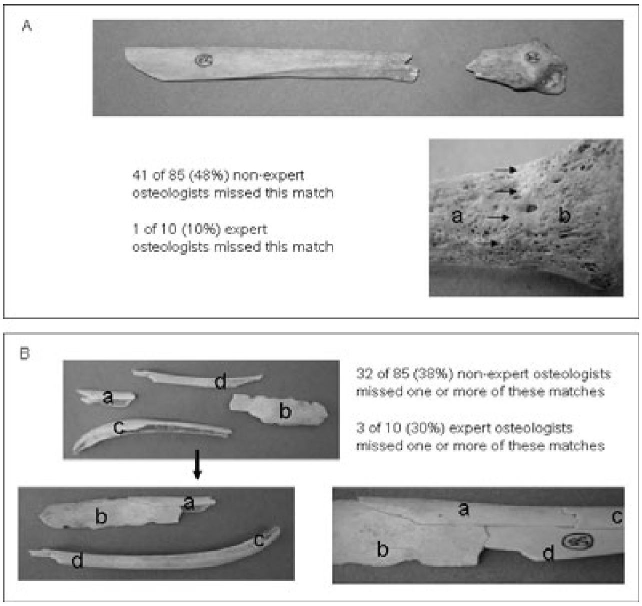

FIG. 5—*Frequently nonassociated matches.*

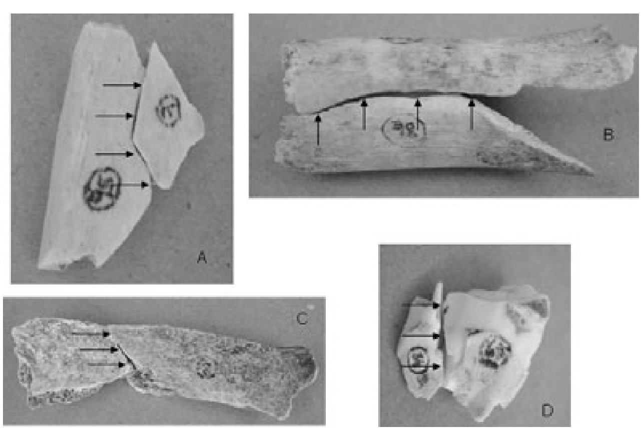

FIG. 6—*Negative associations.*

A total of four negative associations were made in the study, three by one participant and one by another (Fig. 6). In two of the cases (Fig. 6*A* and 6*B*), the fragments were actually from the same bone, but were affixed along the wrong edge. In the remaining two cases (Fig. 6*C* and 6*D*), the incorrectly associated fragments were from the same material class, but not from the same original element. The total negative association rate was 0.001. Because only two participants made negative associations, inter group comparisons were not performed.

The time to complete the exercise ranged from 21 to 150 min with a mean of 60 min (Fig. 7). In regard to completion time by experience level, a significant ANOVA test ($p = 0.005$) reveals a group difference. Subsequent ANOVA tests between groups reveal that the "expert" osteologists, who averaged just less than 39 min (Table 3), finished the task significantly more quickly that those in the "no experience" group, but the differences were not significant in either the "expert"–"some experience" or the "some experience"–"no experience" comparisons.

By far, the most commonly reported criterion used to identify matches was color, followed by texture, shape, and size (Fig. 8).

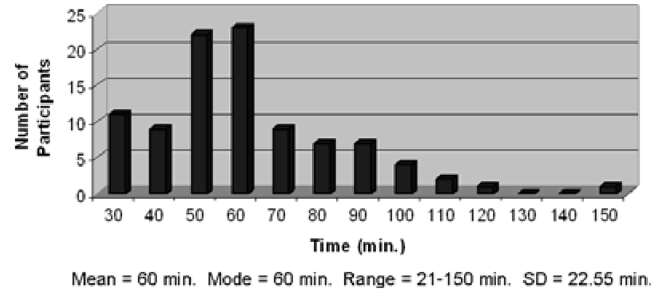

FIG. 7—*Exercise completion time (all participants).*

TABLE 3—*Exercise completion time by osteology experience (minutes).*

Group	Mean	Range	SD
Expert (N = 10)	38.7	21–60	3.65
Some experience (N = 23)	57.87	30–150	5.84
No experience (N = 62)	63.24	30–120	20.18

ANOVA: F-value = 5.58, p-value = 0.005.

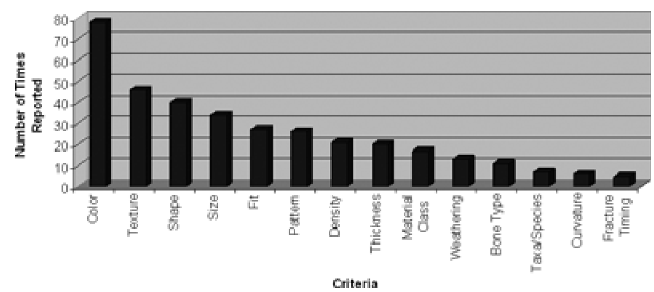

FIG. 8—*Criteria reported by participants for identifying matches.*

Also reported were fracture fit, pattern, density, thickness, material class, weathering, bone type, taxa, curvature, and fracture timing. Interestingly, although not surprising, only expert osteologists reported taxa as a criterion, and all but one expert osteologist reported taxa as a criterion.

Discussion and Conclusions

There was a trend (but without statistical significance) for those with more experience to perform slightly better on the exercise, especially when it came to expert osteologists and matches involving certain human bones. Those with more experience in osteology also completed the exercise in significantly less time. The positive association rate among all participants was 0.925, and negative associations were extremely infrequent, occurring at a rate of just 0.001.

Note that the positive association rates calculated here are likely somewhat deflated because of the way the exercise was scored. Each nonassociated fracture edge counted against the participant even though the available number of associations actually decreases with each correct match made. For example, in the rib pictured in Fig. 5*B*, failure to associate the two rib halves, although technically just one association once the other two matches have been made, counts for three missed fracture edge matches (a-to-c, a-to-d, and b-to-d) in the scoring system used.

In this study, nonassociations were significantly more likely to occur than negative associations. Nonassociation errors, however, are generally viewed among forensic experts as less harmful because they nearly always work in the accused's favor (possibly falsely absolving a defendant who is actually guilty), as compared with false positive errors that wrongfully associate two evidentiary items (possibly falsely accusing or convicting a defendant who is actually innocent).

Osteologists who can draw on their education and experience to correctly identify, anatomically orient, and reassemble bone fragments may be at an advantage in more complex cases involving human remains. Even individuals inexperienced in both physical matching and osteology, however, were able to locate and identify about 92% of all correct matches, supporting the reliability and validity of performing these matches. Perhaps identifying physical matches often is intuitively evident, at least for most people, and the ability to locate matches may be related to an individual's aptitude for spatial tasks or other factors rather than their education or professional experience. Although the authors are unaware of any specific challenges that bone reconstruction has faced in a forensic context, physical matches could potentially arise as *Daubert* issues for anthropologists in areas such as minimum number of individuals estimates as well as trauma reconstruction, and by providing quantifiable reliability data and error rates in advance, such challenges can be easily addressed.

Acknowledgments

We sincerely appreciate the time and interest of those who participated in the study. We would also like to thank the Department of Anthropology, University of Tennessee for providing some of the specimens used, individuals from the FBI Laboratory's Firearms and Toolmarks Unit for valuable comments during the development of the study and this manuscript, and the anonymous reviewers whose comments and suggestions greatly improved the paper.

References

1. Villa P. Conjoinable pieces and site formation processes. Am Antiq 1982;47:276–90.
2. Papaioannou G, Karabassi E, Theoharis T. Virtual archaeologist: assembling the past. IEEE Comput Graph Appl 2001;21:53–9.
3. *Daubert V. Merrell-Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
4. Bradley MJ, Keagy RL, Lowe PC, Rickenback MP, Wright DM, LeBeau MA. A validation study for duct tape end matches. J Forensic Sci 2006;51:504–8.
5. Villa P, Mahieu E. Breakage patterns of human long bones. J Hum Evol 1991;21:27–48.

Additional information — reprints not available from author:
Angi M. Christensen, Ph.D.
FBI Laboratory, Trace Evidence Unit
2501 Investigation Parkway
Quantico, VA 22135
E-mail: angi.christensen@ic.fbi.gov