*J Forensic Sci*, July 2008, Vol. 53, No. 4
doi: 10.1111/j.1556-4029.2008.00751.x
Available online at: www.blackwell-synergy.com

**TECHNICAL NOTE**

*Ann H. Ross,[1] Ph.D. and Shanna Williams,[2] Ph.D.*

# Testing Repeatability and Error of Coordinate Landmark Data Acquired from Crania*

**ABSTRACT:** This study evaluates issues of precision, repeatability, and validation in three-dimensional (3D) landmark coordinates. Two observers collected 19 homologous cranial landmarks from three skulls during three separate digitizing sessions. Statistical analysis was conducted on the 171 interlandmark distances (ILDs) derived from the original coordinate data. A mixed model ANOVA detected significant within-subject error in 54 of the 171 ILDs (i.e., 32%). A GLM procedure revealed significant between-observer variation in 14 ILDs and significant observer-by-session differences in 13 ILDs. The majority of these differences involved ILDs with type 3 landmarks as endpoints, such as euryon and alare. Unlike type 1 and 2 landmarks which are biologically informative in all directions, type 3 landmarks contain a substantial arbitrary component. Thus, it is not surprising type 3 landmarks displayed significant digitizing error. Given these results, we caution researchers to be mindful of type 3 landmarks measurement discrepancies when selecting landmarks for coordinate data evaluation.

**KEYWORDS:** forensic science, cranial landmarks, repeatability, geometric morphometrics

Historically, size and shape analyses have relied on the application of multivariate statistical methods to caliper-derived linear distances and/or angles (1–4). One of the limitations of traditional caliper-derived metric data is that the measurements are confined to the positions of the caliper endpoints, which are defined by anatomical locations or landmarks (5,6). The end result is a linear distance measure that encodes incomplete information about the relative positions of these landmarks in space (5,6).

Modern methods of geometric morphometrics address many of the shortcomings associated with traditional metrics by focusing on the analysis of landmark coordinates (2,6). Unlike traditional metrics, coordinate data can better capture in multiple directions the geometric information available in anatomical structures (6). These newer three-dimensional (3D) methods have gained popularity in physical anthropology over the last decade and have been adopted by many evolutionary theorists, clinicians, and forensic anthropologists. Data capturing techniques range from direct digitization of landmarks via 3D digitizers to point extraction from scanned images.

Only a handful of studies have tested the precision, repeatability, and validation of anthropometric landmarks extracted from computed tomography (CT) and other optical surface imaging methods. They were all found to be highly repeatable (7–9). However, despite finding CT scans to be internally consistent, Richtsmeier cautioned against the use of CT data in combination with other direct means of measurement (8). Corner et al. (10) examined measurement error from 11 landmarks on a single macaque skull digitized 20 times by two observers and found that both observers were consistent in locating the landmarks. They concluded that since both observers were experienced, measurement error depends on the digitizer and the type of landmarks included in the study. Furthermore, Corner and colleagues suggested the amount of error is directly related to the linear distance between any two landmarks, with a greater proportion of error existing over small distances (11). In a recent study, Slice and coworkers (12) examined the landmark-specific error for data collected by multiple observers for multiple specimens. They found that type 1 landmarks, such as nasion and bregma, were the most reproducible, while type 3 landmarks, such as opisthocranion, were the least reproducible. They also found that measurement variability was a function of the interaction between landmark, skull, and observer.

More than a decade has passed since the "revolution" in morphometrics, however, these newer modalities of data acquisition and analyses have undergone little systematic testing for accuracy, particularly in regards to direct skull digitization. Thus, the purpose of this pilot study is to evaluate the repeatability and error associated with the collection of coordinate cranial landmark data via direct digitization from dry skulls.

## Materials and Methods

### *Research Design*

The coordinates for 19 standard homologous cranial landmarks were collected using a Microscribe 3DX® and G2X® digitizer (Immersion Corporation, San Jose, CA) and the software ThreeSkull written by Steve Ousley (Table 1 and Figs. 1 and 2). Detailed landmark descriptions are found in Howells (13). The landmarks were chosen to include standard type 1 and 2 landmarks (e.g., bregma and subspinale), as well as standard caliper-derived type 3 anatomical points (e.g., euryon). The caliper-derived measurements were first located with calipers and marked with a pencil. All pencil marks were erased before transferring the skulls to the second observer. Due to software incompatibility, the subtense points were

[1]Department of Sociology and Anthropology, NC State University, CB 8107, Raleigh, NC 27695-8107.

[2]C.A. Pound Human Identification Laboratory, Department of Anthropology, University of Florida, P.O. Box 103615, Gainesville, FL 32610.

*Based on a paper presented at 59th Annual Meeting of the American Academy of Forensic Sciences, San Antonio, TX, February 2007.

Received 16 April 2007; and in revised form 4 Nov. 2007; accepted 23 Dec. 2007.

TABLE 1—*List of landmarks used with abbreviations.*

| |
|---|
| 1. Alare left (alarl) |
| 2. Alare right (alarr) |
| 3. Bregma (brg) |
| 4. Dacryon left (dacl) |
| 5. Dacryon right (dacr) |
| 6. Euryon left (eul) |
| 7. Euryon right (eur) |
| 8. Lambda (lam) |
| 9. Metopion (met) |
| 10. Occipital subtense point (ocspt) |
| 11. Opisthocranion (opg) |
| 12. Parietal subtense point (paspt) |
| 13. Radiometer point left (radptl) |
| 14. Radiometer point right (radptr) |
| 15. Subspinale (ssp) |
| 16. Zygion left (zygl) |
| 17. Zygion right (zygr) |
| 18. Zygoorbitale left (zygool) |
| 19. Zygoorbitale right (zygoor) |

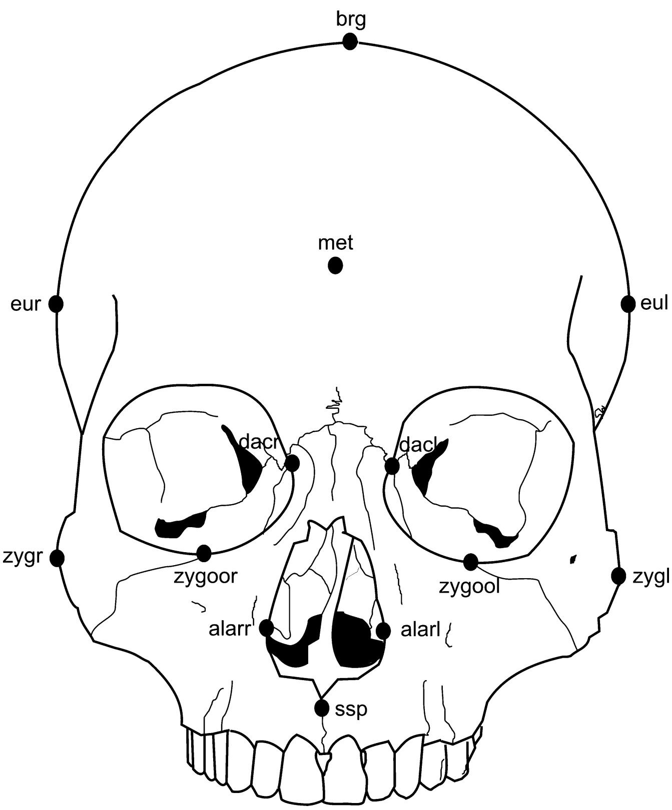


FIG. 1—*Anatomical landmarks in anterior view.*

collected differently between observers. The first author located the points with coordinate calipers and marked them with a pencil. The second author used the "Arcs" feature within ThreeSkull to derive subtense points from continuous stream data collection.

Three skulls were randomly selected for this study from teaching specimens housed at the C.A. Pound Human Identification Laboratory. Each skull underwent three separate digitization sessions by two separate observers for a total of six digitizations for each skull. Ideally, you would want to acquire coordinate data from the skulls fixed in a common coordinate system between digitizing sessions. However, in this study the skulls were digitized at two separate locations, by the second author at the C.A. Pound Human

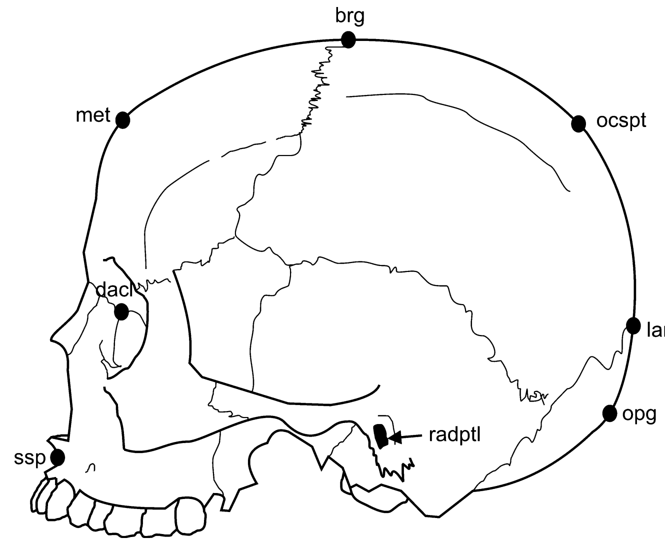


FIG. 2—*Anatomical landmarks in lateral view.*

Identification Laboratory and by the first author at NC State. In this case, a Generalized Procrustes Analysis (GPA) is not recommended, as any error due to repeatability would be masked by the "fitting" process, i.e., the process of translating and rotating the raw coordinate data into a common coordinate system and scaling to a common size (8,14,15). Because the skulls were not "fixed" in a common coordinate system between digitizing sessions, interlandmark linear distances or ILDs were derived using the software PAST, which is available for downloading (16). The number of pairs is calculated as $N(N-1)/2$ for $N$ landmarks, creating a total of 171 ILDs (16). ILDs allow for multivariate analysis of distance data as they are not sensitive to rotation or translation. This, in turn, makes Procrustes fitting of the data unnecessary.

### *Statistical Analysis*

The newly derived ILDs were used for subsequent statistical analyses. Digitization error, the within-subject error or the proportion of the total variance explained by multiple digitizing sessions of the same skull, was tested using a mixed model analysis of variance (ANOVA) for random effects. Repeatability, defined here as the between-observer variation, was tested with ANOVA using the general linear model (GLM) routine. These analyses were performed using SAS system for Windows Version 9.1.3 (17).

## Results

The mixed model ANOVA detected significant error due to digitizing in excess of 5% of the total variance in 32% or 54 out of the 171 ILDs (Table 2). Twenty of these interlandmark distances have endpoints on euryon, 15 on alare, eight on opisthocranion, six on radiometer point, and seven on left dacryon. Some of these ILDs included both endpoints.

The GLM procedure detected significant between-observer difference for 14 ILDs (Table 3). The results show that the landmarks which are not highly repeatable are those involving type 3 landmarks, specifically, euryon, alare, and radiometer point. In addition, parietal and occipital subtense points appear to be problematic. However, the inconsistency with both subtense points may reflect the differential acquisition of these points, as one observer used a coordinate caliper to locate the points prior to digitization, while

TABLE 2—*ILDs showing >5% error (see Table 1 for landmarks).*

| | Observer 1 | | Observer 2 | |
|---|---|---|---|---|
| ILD | Min | Max | Min | Max |
| 1 (1–2) | 23.88 | 27.59 | 23.99 | 27.26 |
| 2 (1–3) | 139.12 | 151.42 | 139.53 | 150.16 |
| 5 (1–6) | 117.43 | 136.98 | 108.91 | 126.24 |
| 6 (1–7) | 123.18 | 138.52 | 129.45 | 139.18 |
| 8 (1–9) | 95.50 | 106.98 | 96.31 | 107.20 |
| 10 (1–11) | 162.32 | 193.40 | 164.09 | 188.57 |
| 13 (1–14) | 97.40 | 107.07 | 96.52 | 105.03 |
| 17 (1–18) | 39.60 | 47.33 | 39.90 | 48.22 |
| 22 (2–6) | 128.62 | 145.21 | 119.36 | 136.48 |
| 23 (2–7) | 109.78 | 126.14 | 115.41 | 126.07 |
| 27 (2–11) | 160.86 | 192.05 | 165.44 | 187.11 |
| 30 (2–14) | 85.89 | 95.82 | 86.19 | 93.90 |
| 31 (2–15) | 16.79 | 19.48 | 18.46 | 20.00 |
| 34 (2–18) | 21.87 | 26.99 | 17.10 | 24.61 |
| 37 (3–5) | 105.83 | 115.20 | 104.75 | 114.39 |
| 38 (3–6) | 79.17 | 122.03 | 91.58 | 101.36 |
| 43 (3–11) | 125.47 | 146.10 | 115.90 | 143.15 |
| 53 (4–6) | 94.61 | 125.75 | 90.26 | 107.86 |
| 54 (4–7) | 102.15 | 120.10 | 111.69 | 118.30 |
| 55 (4–8) | 151.38 | 176.84 | 151.98 | 175.58 |
| 56 (4–9) | 66.45 | 71.15 | 61.70 | 70.79 |
| 57 (4–10) | 148.93 | 170.98 | 149.12 | 174.01 |
| 58 (4–11) | 149.83 | 177.61 | 151.45 | 175.38 |
| 61 (4–14) | 91.35 | 100.41 | 93.86 | 99.55 |
| 63 (4–16) | 67.35 | 77.19 | 68.03 | 73.32 |
| 67 (5–6) | 106.97 | 132.37 | 102.19 | 118.19 |
| 68 (5–7) | 87.09 | 109.72 | 96.46 | 108.78 |
| 76 (5–15) | 43.21 | 52.63 | 43.41 | 52.12 |
| 80 (5–19) | 67.31 | 75.06 | 68.55 | 73.00 |
| 81 (6–7) | 115.47 | 134.18 | 115.50 | 132.60 |
| 82 (6–8) | 71.98 | 106.18 | 91.27 | 105.54 |
| 83 (6–9) | 92.81 | 139.35 | 97.40 | 116.43 |
| 89 (6–15) | 130.59 | 147.87 | 121.10 | 142.13 |
| 91 (6–17) | 95.42 | 122.36 | 87.67 | 110.40 |
| 92 (6–18) | 124.89 | 142.72 | 118.52 | 135.30 |
| 94 (7–8) | 90.33 | 114.51 | 87.37 | 105.41 |
| 97 (7–11) | 96.29 | 116.44 | 86.54 | 110.83 |
| 100 (7–14) | 47.51 | 65.49 | 40.92 | 59.54 |
| 101 (7–15) | 124.04 | 145.92 | 128.99 | 143.80 |
| 102 (7–16) | 131.31 | 146.50 | 131.47 | 142.84 |
| 103 (7–17) | 120.18 | 140.38 | 127.40 | 137.34 |
| 105 (7–19) | 51.38 | 69.02 | 50.21 | 63.69 |
| 108 (8–11) | 16.74 | 45.85 | 22.52 | 34.03 |
| 114 (8–17) | 159.41 | 187.42 | 160.06 | 186.66 |
| 118 (9–11) | 152.65 | 185.24 | 144.14 | 183.87 |
| 127 (10–11) | 1.49 | 49.81 | 14.98 | 20.25 |
| 129 (10–13) | 79.69 | 97.84 | 79.87 | 105.24 |
| 136 (11–12) | 83.54 | 102.36 | 80.16 | 97.11 |
| 139 (11–15) | 166.42 | 203.42 | 173.44 | 197.65 |
| 140 (11–16) | 119.60 | 149.18 | 127.06 | 148.45 |
| 141 (11–17) | 153.97 | 186.33 | 156.94 | 182.75 |
| 157 (14–15) | 98.77 | 109.34 | 98.50 | 107.07 |
| 159 (14–17) | 101.24 | 109.81 | 102.30 | 108.49 |
| 167 (16–18) | 95.42 | 107.39 | 93.97 | 105.80 |

the second observer used contour data to calculate the points. The differential acquisition of data was necessary because the data acquisition software ThreeSkull was not compatible with the collection of contour data when using a G2X® digitizer.

Interestingly, three of the ILDs with significant between-observer variation had endpoints on dacryon and two had endpoints on zygoorbitale. Occasionally these landmarks may be difficult to locate, especially if they are ill-defined due to suture obliteration, which may reflect the present results.

The GLM procedure also revealed significant observer-by-session difference for 13 ILDs (Table 4). These results display the same pattern as the significant between-observer differences in that the majority of these differences involved ILDs with type 3 landmarks as endpoints, specifically euryon, alare, opisthocranion, and radiometer point. In other words, the observers were having difficulty locating type 3 landmarks across digitizing sessions for the same skull. Notably, locating opisthocranion across different sessions of the same skull seems to be more problematic than between observers.

TABLE 3—*Between-observer variation.*

| ILD | DF | Type III SS | MS | F-Value | Pr > F |
|---|---|---|---|---|---|
| Alarl-dacr | 1 | 3.89 | 3.89 | 29.98 | 0.03 |
| Alarl-brg | 1 | 36.38 | 36.38 | 70.85 | 0.01 |
| Alarr-paspt | 1 | 63.13 | 63.13 | 19.39 | 0.05 |
| Brg-radptl | 1 | 13.23 | 13.23 | 27.03 | 0.04 |
| Dacl-radptl | 1 | 5.05 | 5.05 | 24.24 | 0.04 |
| Dacl-zygool | 1 | 6.02 | 6.02 | 22.78 | 0.04 |
| Dacr-zygool | 1 | 6.14 | 6.14 | 233.91 | 0.004 |
| Eul-radptl | 1 | 196.67 | 196.67 | 167.09 | 0.006 |
| Eur-lam | 1 | 111.36 | 111.36 | 44.53 | 0.02 |
| Eur-ocspt | 1 | 700.79 | 700.79 | 36.1 | 0.03 |
| Met-ocspt | 1 | 71.72 | 71.72 | 29.98 | 0.03 |
| Paspt-radptr | 1 | 9.93 | 9.93 | 20.01 | 0.05 |
| Paspt-ssp | 1 | 50.97 | 50.97 | 69.64 | 0.01 |
| Ssp-zygool | 1 | 0.58 | 0.58 | 30.88 | 0.03 |

TABLE 4—*Observer* session variation.*

| ILD | DF | Type III SS | MS | F-Value | Pr > F |
|---|---|---|---|---|---|
| Alarl-opg | 2 | 10.05 | 5.02 | 4.54 | 0.05 |
| Alarr-ssp | 2 | 1.15 | 0.58 | 5.15 | 0.04 |
| Brg-opg | 2 | 46.33 | 23.16 | 6.91 | 0.02 |
| Dacl-ocspt | 2 | 4.25 | 2.12 | 17.13 | 0.001 |
| Eurl-opg | 2 | 53.76 | 26.88 | 14.51 | 0.002 |
| Eurl-paspt | 2 | 7.15 | 3.57 | 7.55 | 0.01 |
| Eurl-radptr | 2 | 19.74 | 9.87 | 5.11 | 0.04 |
| Lam-opg | 2 | 120.29 | 60.15 | 6.08 | 0.03 |
| Met-opg | 2 | 13.85 | 6.93 | 8.31 | 0.01 |
| Ocspt-zygool | 2 | 3.78 | 1.69 | 7.99 | 0.01 |
| Opg-paspt | 2 | 118.35 | 59.17 | 6.78 | 0.02 |
| Opg-zygool | 2 | 4.74 | 2.37 | 5.79 | 0.03 |
| Zygl-zygoor | 2 | 11.28 | 5.64 | 5.05 | 0.04 |

## Discussion

Fred Bookstein identified three types of landmarks, or anatomical points, commonly used in geometric morphometric analyses (5). Type 1 landmarks are defined as discrete juxtapositions of tissues such as at the intersection of three sutures, for example, dacryon and asterion (5,6). Type 2 landmarks are curvature maxima or other local morphogenetic processes, usually with a biomechanical implication such as a muscle attachment site. Ectoconchion and prosthion would be examples of type 2 landmarks (5,6). Type 3 landmarks, on the other hand, are extremal points, like the endpoints of maximum cranial length and breadth (5,6). These points are considered to be deficient and are rarely meaningful as landmarks. Their deficiency arises from the fact that the meaningful information they contain is only in the direction parallel to the remote defining structure and this information has a substantial arbitrary component (6). Hence, though you may have three coordinates, only a single linear combination of them is meaningful. Types 1 and 2 landmarks, meanwhile are viewed as biologically informative in all directions (6).

Given the limitations of type 3 landmarks, it is not surprising that significant digitizing error resulted from the use of these landmarks in this study. For example, the exact location of the anatomical point euryon is only determined by directly measuring maximum cranial breadth with a set of spreading calipers. Although maximum cranial breadth can be recorded with a high level of accuracy in traditional craniometrics, the exact anatomical position of the coordinates for euryon cannot be easily located, because the span of the maximum cranial breadth measurement may encompass a relatively large area. For example, a skull may measure 138 mm in a broad general location, and thus the exact location of the coordinates cannot be easily replicated between observers (Fig. 3). The same holds true for the landmarks alare and opisthocranion, which also have to be determined by direct measurement. The ILDs that included left dacryon, which showed error in excess of 5%, are most likely the product of an ill-defined landmark such as the result of suture obliteration and are probably atypical.

Concurring with this study, Slice and colleagues (12) found that type 1 landmarks (e.g., nasion and bregma) were the most reproducible, while type 3 landmarks (e.g., opisthocranion) were the least reproducible. They also found that measurement variability was a function of the interaction between landmark, skull, and observer. Corner et al. (10) found that measurement error was greatest at anatomical points that were located on an osseous edge at extreme ends of sutures such as zygomaxillare superior. They recommend repeatedly collecting ''problematic'' landmarks four to five times and using the average in the analyses. However, this would be impractical when under strict time constraints.

Extremal or type 3 landmarks are fairly accurate when instrumentally derived as linear distances in traditional morphometrics. However, these results demonstrate that they are highly variable when attempting to archive their exact anatomical location. Moreover, type 3 landmarks are associated with a sizeable degree of error, both between and within observers. Thus we caution the use of type 3 landmarks in morphometric coordinate data evaluations and recommend utilizing only types 1 and 2 landmarks, which are considered biologically significant.

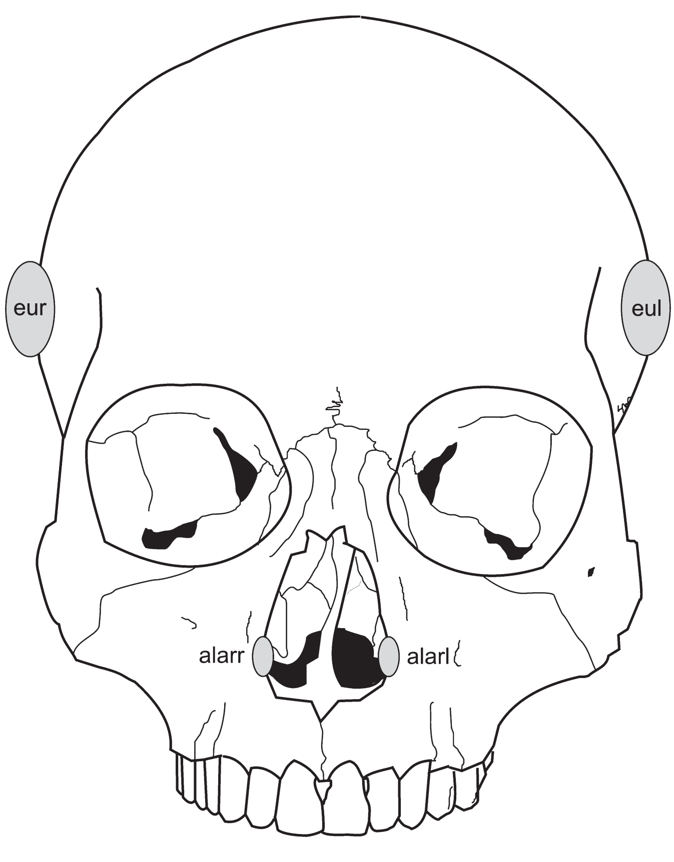


FIG. 3—*Broad general location for landmarks euryon and alare.*

*Acknowledgment*

This project was funded by a National Institute of Justice grant (# 2005-MU-BX-K078).

## References

1. Lynch JM, Wood CG, Luboga SA. Geometric morphometrics in primatology: craniofacial variation in *Homo sapiens* and *Pan troglodytes*. Folia Primatol 1996;67:15–39.
2. Rohlf FJ, Marcus LF. A revolution in morphometrics. Tree 1993;8:129–32.
3. Ross AH, McKeown AH, Konigsberg LW. Allocation of crania to groups via the ''new'' morphometry. J Forensic Sci 1999;44:584–7.
4. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the ''revolution.'' Ital J Zool 2004;71:5–16.
5. Bookstein FL. Morphometric tools for landmark data. Geometry and biology. Cambridge: Cambridge University Press, 1991.
6. Slice DE. Modern morphometrics. In: Slice DE, editor. Modern morphometrics in physical anthropology. New York: Kluwer Academic, 2005;1–45.
7. Kohn LA, Cheverud JM, Bhatia G, Commean P, Smith K, Vannier MW. Anthropometric optical surface imaging system repeatability, precision, and validation. Ann Plast Surg 1995;34:362–71.
8. Richtsmeier JT, Paik CH, Elfert PC, Cole TM, Dahlman HR. Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans. Cleft Palate Craniofac J 1995; 32: 217–27.
9. Aldrige K, Boyadjiev SA, Capone GT, DeLeon VB, Richtsmeier JT. Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. Am J Med Genet 2005; 138A: 247–53.
10. Corner BD, Lele SB, Richtsmeier JT. Measuring precision of three-dimensional landmark data. Quant Anthropol 1992;3:347–59.
11. Lele SB, Richtsmeier JT. An invariant approach to statistical analysis of shapes. Boca Raton: Chapman & Hall/CRC, 2001.
12. Slice DE, Unteregger C, Bookstein FL, Schäfer K. Modeling the precision of landmark data. Am J Phys Anthropol 2004;123(S38):183.
13. Howells WW. Cranial variation in man. Papers of the Peabody Museum of Archaeology and Ethnology Harvard University. Vol. 51. Cambridge: Harvard University, 1973.
14. Bookstein FL. Combining the tools of geometric morphometrics. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE, editors. Advances in morphometrics. New York: Plenum Press, 1996;131–51.
15. Rohlf FJ, Slice DE. Methods for comparison of sets of landmarks. Syst Zool 1990;39:40–59.
16. Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological Statistics Software Package for Education and Data Analysis [computer program]. Palaeontol Electronica 2001;4: 9pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm. http://folk.uio.no/ohammer/past/download.html. Accessed on April 28, 2008.
17. SAS. System for Windows [computer program]. Version 9.1.3. Cary (NC): SAS Institute, Inc., 2003.

Additional information and reprint requests:
Ann H. Ross, Ph.D.
Department of Sociology and Anthropology
North Carolina State University
CB 8107
Raleigh, NC 27695-8107
E-mail: ann_ross@ncsu.edu