Precision, Repeatability, and Validation of the Localization of Cranial Landmarks Using Computed Tomography Scans

JOAN T. RICHTSMEIER, PH.D.
CHUL H. PAIK, PH.D.
PETER C. ELFERT, R.T.
THEODORE M. COLE III, M.A.
HOLLY R. DAHLMAN, B.A.

Computed tomography (CT) has brought to the craniofacial surgeon a three-dimensional representation of internal structures. CT scans provide visualization of anatomy for preoperative planning and postoperative evaluation. Beyond visualization, however, a CT scan enables assessment of measurements useful to clinicians and basic scientists. All measurement systems used with CT require the ability to accurately locate regions of interest on the image (i.e., areas, volumes, outlines, curves, surfaces, points). This study evaluates the precision and repeatability of locating anatomic landmarks in three dimensions on CT slice images, and validates these locations using an established measurement system. The average error of landmark position is always less than 0.5 mm and for some landmarks error is negligible. Repeatability studies show that less than 2% of the total variance in our data is due to measurement inaccuracy. Although data collected from CT scans are internally consistent, validation results caution the use of CT data in combination with data collected using calipers or other direct means of measurement.

KEY WORDS: *biologic landmarks, computed tomography, craniofacial, precision, repeatability, validation*

Computed tomography (CT) imaging has had a vast and dramatic effect on the field of medicine. Neuroradiologic CT images allow the clinician to visualize and classify craniofacial malformations and to plan therapeutic and/or reconstructive surgery (Vannier et al., 1984; Leboucq et al., 1991, 1993; Fernbach and Feinstein, 1992). Neuroradiologic CT images are also becoming a common source of data for clinical and basic science research (e.g., Genitori et al., 1991-1992; Gault et al., 1992; Fok et al., 1992; Posnick et al., 1992; Richtsmeier et al., 1994). Methods developed to examine repeatability, measurement error, or validation of data produced by other systems (e.g., Speculand et al., 1988; Corner et al., 1992; Kohn and Cheverud, 1992) are not routinely applied to CT data, but previous work has been done to validate cranial metrics obtained from standard CT images (Christiansen et al., 1986; Hildebolt and Vannier, 1988; Matteson et al., 1989; Waitzman et al., 1992a, b; Richardson et al., 1993).

A three-dimensional object, complete with surface, volume, and density information is too complex to statistically analyze as a whole. Consequently, in order to study anatomic structure, biologic variation, morphologic dimensions, or biologic processes such as growth using data from CT images, relevant details must be extracted from the images for analysis. Depending upon the focus of the investigation, the form is characterized using selected details. The definition and quantification of these details is the first step of the data-reduction procedure. Several methods are available for the analysis of surface, outline, and landmark data.

Landmarks have been used for over a century by anthropometrists interested in quantifying cranial variation (e.g., Broca, 1875; Boas, 1905; Papillaut, 1906; Martin, 1928) and a field of study, morphometrics, has grown around the statistical analysis of landmark data (e.g., Rohlf, 1990; Rohlf and Bookstein, 1990; Richtsmeier et al., 1992; Rohlf and Marcus,

---

Dr. Richtsmeier is Associate Professor, Department of Cell Biology and Anatomy and Cleft and Craniofacial Center, Division of Plastic, Reconstructive and Maxillofacial Surgery, The Johns Hopkins University. Dr. Paik was Research Associate and Director of Computer Programs in Neuroradiology, The Johns Hopkins University and is now Director, Medical Imaging Laboratory, Samsung Biomedical Research Institute, Seoul, Korea. Peter Elfert is a Neuroradiology Technologist, The Johns Hopkins Medical Institutions. Tim Cole is Research Assistant, Department of Cell Biology and Anatomy, The Johns Hopkins University School of Medicine. Holly Dahlman is a Medical Student, The Johns Hopkins University School of Medicine, Baltimore, Maryland.

This study was supported in part by grants from the National Science Foundation (DBS-9209083), the National Institute of Dental Research (1 P50 DE11131-01), and the National Institutes of Health (Biomedical Research Service Grant SO-7-RR05378), and by The Johns Hopkins School of Medicine Fund (M551077OE).

Submitted October 1994; Accepted January 1995.

Reprint requests: Dr. Joan T. Richtsmeier, Department of Cell Biology and Anatomy, The Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205.

1993). A group of landmarks defined on a form provides a repeatable, geometric representation of homologous structures, but preserves the geometry of a form only up to a point; information pertaining to surfaces and outlines are lost when landmarks are used to define a form (Bookstein et al., 1982; Read and Lestrel, 1986). In exchange for that limitation, the use of landmark data assures the investigator of homology of the biologic structures to be compared between individuals or within an individual over time (see Hall, 1994), and a data type that can be expressed free of any particular coordinate system (Lele, 1993; Richtsmeier and Lele, 1993).

Anatomic cranial landmarks are defined as biologically meaningful loci that can be unambiguously defined and repeatedly located on a biologic structure. The most basic requirement of a landmark is that it can be easily identified and located with a high degree of accuracy and precision. Classes of anatomic landmarks include foramina that contain neurovascular bundles (e.g., foramen spinosum, foramen ovale); the intersection of sutures (e.g., nasion, bregma); and bony processes (e.g., jugular notch, anterior clinoid process). Particular landmarks are chosen based on prior biologic knowledge (e.g., knowledge of general anatomy, of growth patterns, of disease processes, or of a particular physiologic process such as mastication). This knowledge may help to identify landmarks that have direct bearing on a research problem. Our laboratory has been using landmark data from CT scans, and other sources to analyze the changes that occur in craniofacial morphology during growth, and the role of growth in the production of craniofacial dysmorphology (e.g., Wong et al., 1991; Corner and Richtsmeier, 1992a, b; Richtsmeier, 1993; Richtsmeier et al., 1993a, b; Richtsmeier and Lele, 1993; Dufresne and Richtsmeier, 1995). In these studies, we have tried to use landmarks that adequately define the form of various anatomic regions of the craniofacial skeleton.

In this paper, we apply an analytical model proposed by Kohn and Cheverud (1992) to landmark data collected from high-resolution (1.5-mm slice thickness) CT images of dry skulls. The purpose is to validate our method of collecting three-dimensional landmark coordinate data from CT slice images and to determine the accuracy and repeatability of these measures. These data provide an evaluation of CT based measurements for use in morphometric research.

Research Design

We have designed our study to evaluate the types of potential errors in imaging systems outlined by Kohn and Cheverud (1992). The three types of errors include: (1) error in producing a digital image from a living subject (checking the ability of the scanner to produce a good quality image); (2) error in detecting landmarks in digitized images (checking the ability of the observer to see the landmark on an image); and (3) error in locating and recording landmarks on the images by the anthropometrist (checking the ability of the observer to correctly use specific hardware, software, and interface tools, such as a mouse-driven cursor). Following Kohn and Cheverud (1992), we define precision as the average absolute difference between repeated measures of the same image and repeatability as the precision of the measurement relative to the differences among the images contrasted in any given study. Repeatability is measured as the proportion of total variance due to differences between images using a repeated-measures analysis of variance design. One minus the repeatability is the proportional error due to measurement inaccuracy. Although measures of precision and accuracy are calculated in absolute terms, they are generally evaluated in relative terms (Kohn and Cheverud, 1992). In a practical sense, more error is expected and can be tolerated in a study of images representing a population with significant variation between individuals than in a study of a morphologically homogeneous population.

Even if studies of precision and repeatability show acceptable replication of measures captured from images, the measurements may still be inaccurate or biased (Kohn and Cheverud, 1992). This type of error can be evaluated by comparing measures taken on one system with measures taken using an already validated tool. In this study, we use data collected using the Polhemus 3Space table-top digitizer as the "gold standard."

Ten skulls (9 adults, 1 juvenile) from The Johns Hopkins University School of Medicine teaching collection were chosen for this study. Each skull underwent two separate episodes of computed tomography scanning resulting in two sets of CT images for each skull. A set of 35 landmarks were collected from each image during two separate data collection episodes. Using this design, four sets of landmarks were recorded for each skull (Fig. 1). Differences in data collected from the two CT images of a single skull are due to error in digital recording by the CT scanner. Differences in data collected at separate times from the same image are due to recording error. This design enables us to separate error due to the imaging device from error in locating the landmarks on the images. These quantities determine the replicability of measurements taken using CT slice images. To measure potential bias, the locations of the landmarks were also recorded directly from the skulls using the 3Space table-top digitizer. Differences between the data recorded from images of the skulls and those recorded directly from the skulls using the 3Space table-top digitizer measure error in the way that the skull is represented by a CT image and in the tool used to collect data from the CT images (in this case the Macintosh screen and mouse-driven cursor).

Statistical analysis of the repeated measures taken from CT images allows estimation of the precision and repeatability of the images. Precision of three-dimensional coordinates taken repeatedly from a single image can be measured by superimposing repeated measures of the single image in such a way that the squared distance between homologous landmarks is minimized (i.e., using the generalized procrustes algorithm following Gower, 1975; Goodall, 1991). Superimposition is acceptable when landmark data are collected two or more times from an object fixed in a particular coordinate system (e.g., data collected repeatedly from a single CT scan or from a skull that is fixed motionless on the digitizer table-top; see Corner

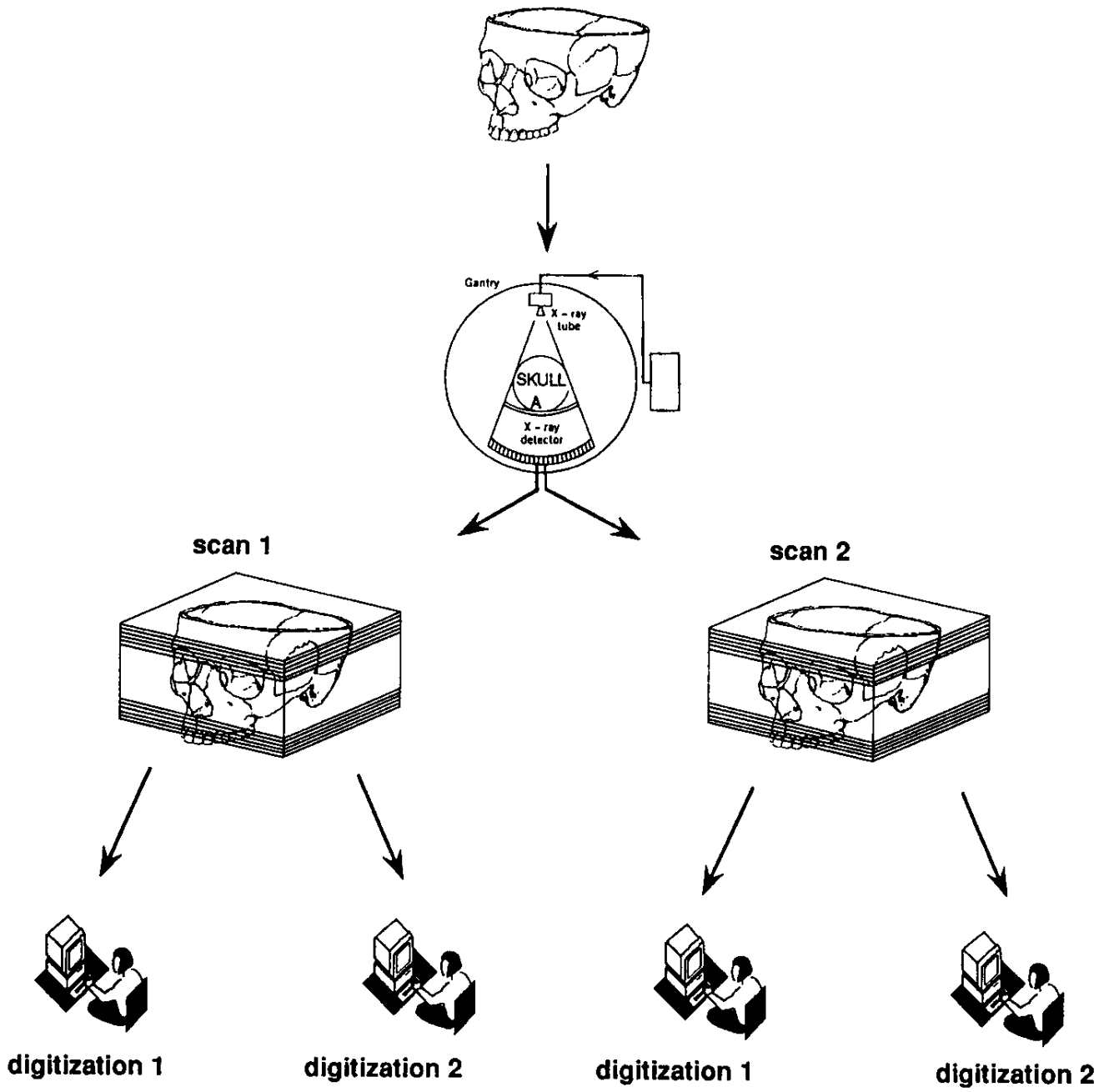

FIGURE 1 The research design (adapted from Kohn and Cheverud, 1992) used for testing our data collection system. This design allows us to test whether structures can be reliably located on the images produced by the imaging system, whether an imaging system can produce repeatable images, and whether structures can be reliably located on individual specimens. The validation portion of the research design is not included in the diagram.

et al., 1992). However, when objects are measured in different coordinate systems and these measures are superimposed according to the generalized procrustes algorithm (Gower, 1966, 1975; Goodall, 1991), error is introduced which makes it impossible to accurately measure the precision of landmark location. This is because the variance-covariance structure obtained by using the procrustes estimator is wrong (Lele, 1993). Below we will explain how this impacts the measure of error.

In Lele's (1993) example, 130 two-dimensional forms with four landmarks were generated with known variation around the four landmarks. When these 130 forms were scaled, rotated, and translated according to the procrustes algorithm, the resulting variance-covariance structure is very different

from that used in the design of the experiment. This means that the data generated by superimposing the 130 forms according to the procrustes fit does not represent the original data accurately. Lele (1993) shows that the fitting criteria of the generalized procrustes estimator (i.e., to minimize the sum of the squared distances between like landmarks on the forms considered) dictates the error. For the generalized procrustes algorithm, the mathematical gain by fitting those landmarks that are farther from the centroid is larger than the cost of not fitting those landmarks that lie closer to the centroid (Lele, 1993). Consequently, the method tends to better match landmarks further from the centroid of the form than the landmarks that are close to the centroid, irrespective of the true distribution at each landmark. In an application of the procrustes algorithm for measuring error as distances between like landmarks, error would appear to be less for those landmarks further from the centroid, while error for those landmarks close to the centroid would be artificially inflated (Lele, 1993).

Alternate superimposition criteria like resistant-fit procrustean approaches (e.g., Siegel and Benson, 1982; Rohlf and Slice, 1990) that superimpose forms under a different minimization criterion (resistant-fit attempts to match a subset of the points exactly) will provide a different, but still erroneous localization of error. Superimposition methods should not be used for error detection in the comparison of landmark data collected from two separate CT images of a single object, in the comparison of data from a single object that has been moved between digitization episodes (using the 3Space digitizer or any scanning device), or for the comparison of landmark data collected from images with data collected using a digitizer or scanner. To avoid these inadequacies of superimposition methods, we present alternate methods for estimation of error between data collected from separate images and for validating data collected from CT images.

The two separate digitizations of a single CT image (see Fig. 1) are compared by calculating differences in the three-dimensional location of the landmarks in the two digitizations. The linear distance between the locations of a pair of homologous landmarks (one from digitization 1, one from digitization 2) is calculated. If landmarks are located in exactly the same three-dimensional site during the two digitizations, this distance will be zero. There is no need for fitting the two digitizations using any superimposition scheme, because they share an identical coordinate system defined by the images. Differences in landmark location are averaged across images and across individual skulls to provide a measure of error due to the digitizing process. Error at this level, quantified as the distance between homologous points digitized from a single image, is the result of interaction between the data collection device and the operator.

To obtain a measure of error due to the imaging device, the separate images of a single skull must be compared. We do this in the following way: (1) the two digitizations of landmark locations collected from an image are averaged using the method proposed by Lele (1993: p. 581); (2) the average location of the 35 landmarks on each image are expressed as a matrix of inter-landmark distances; (3) the absolute difference between like linear distances measured on the two images is calculated. These three steps are accomplished using portions of Euclidean Distance Matrix Analysis (Lele, 1991; Lele and Richtsmeier, 1991, 1992), which is invariant to rotation and translation and is consequently free of problems associated with superimposition. In the conventional application of Euclidean Distance Matrix Analysis (EDMA), the ratios of all homologous linear distances are calculated to describe the relative difference between two forms. In this study, we seek a measure of absolute error in data collected from two images produced from the same form. This is why we calculate the absolute difference between like linear distances measured on the two images. Since there are 595 unique linear distances among the 35 landmarks, there are 595 absolute differences that describe the relationship between the digitizations of the images of each skull.

Repeatability can also be measured with our data. First, an average form is calculated using the four sets of landmark coordinates from the two images of each skull using the algorithm of Lele (1993). This average skull form is expressed as the set of all possible linear distances among the landmarks. The average X, Y, and Z coordinate locations of the landmarks can then be calculated from this overdetermined set of linear distances (Richtsmeier et al., 1993b; Richtsmeier and Lele, 1993; Richtsmeier and Walker, 1993). Following Kohn and Cheverud (1992), we used the X,Y, and Z coordinates as dependent variables in a nested analysis of variance with two levels: individual skull and image nested within individual skull. The separate digitizations of each image serves as the residual term. The variance due to each of these random effects is derived from the analysis of variance (we used the NESTED procedure from SAS [1989]). The proportion of the total variance due to differences between individual *skulls* is the repeatability. The proportion of the total variance due to differences between the two *images* of the same skull represents error due to the imaging device. The proportion of the total variance due to the residual represents the error due to *digitization* of the images.

Translation and rotation of forms using any superimposition scheme will mask local errors for reasons discussed previously. One way to avert superimposition is to fix the object that is being measured (Corner et al., 1992). Fixing the skull to the scanning table during the two scanning episodes for a skull maintains a constant relationship between landmark location in the XY plane and the gantry during separate scanning episodes. When magnification is held constant, this arrangement ensures that coordinates along the X and Y axes can be directly compared between images of the same skull without superimposition. However, fixing the skull to the table does not necessarily ensure that the position of the skull along the axis oriented orthogonal to the plane of the slice images (in our case the Z axis) will be constant

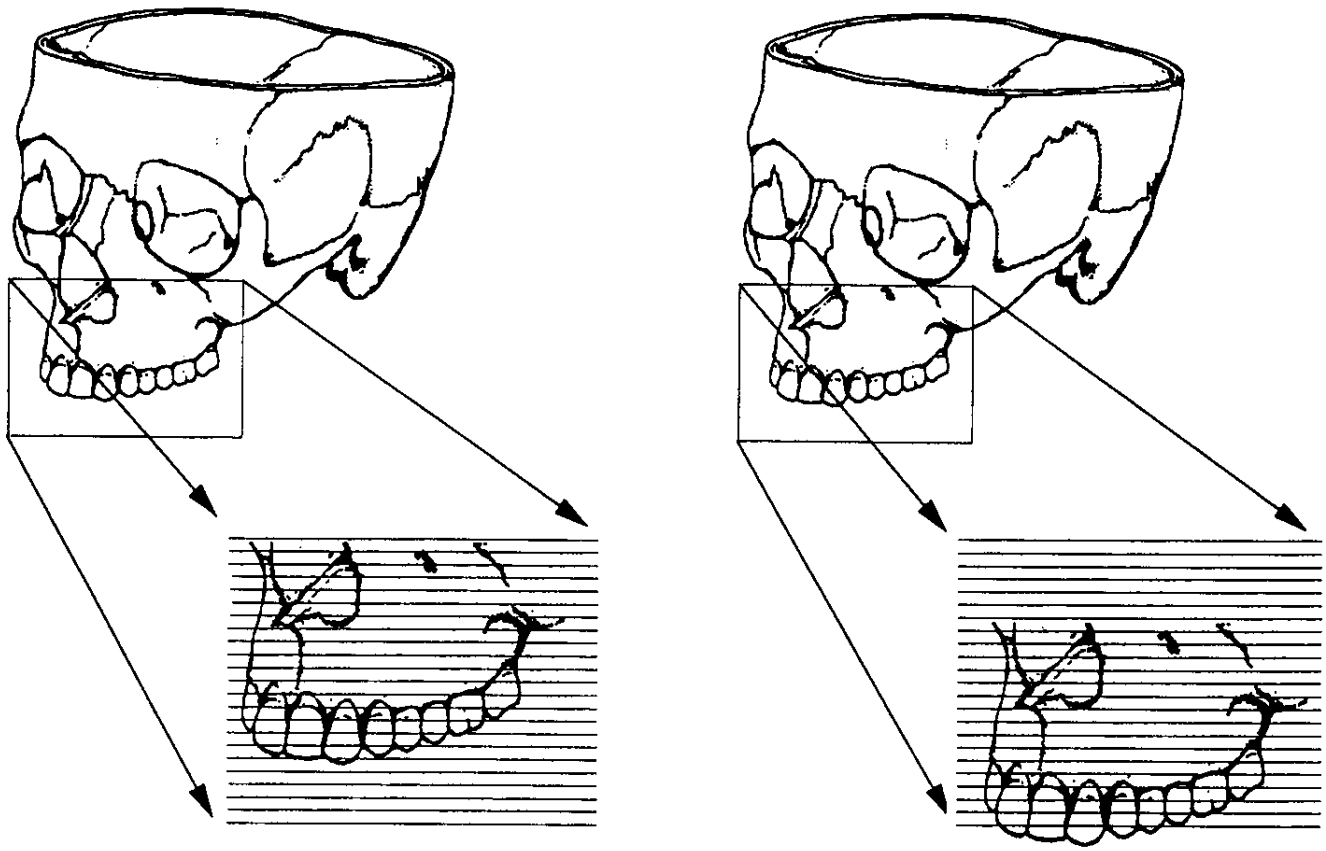

FIGURE 2 Illustration of the potential lack of correspondence between two CT scanning episodes of the same skull. Superimposition (registration) of the two images on the basis of anatomic or extrinsic markers would require translation of the forms along the Z axis. See text for explanation of why superimposition methods are inappropriate when attempting to localize measures of error or difference between forms.

between any two scanning sessions (Fig. 2). Control over the Z axis would require that the x-ray beam of the first slice image be placed at exactly the same location on the skull during each scanning session. We did not control the location of the x-ray beam of the first slice image with relation to anatomic location. Therefore, calculation of repeatability of the Z coordinate of landmark locations would require the superimposition of forms. For this reason, the nested ANOVA design of Kohn and Cheverud (1992) was used to measure repeatability along the X and Y axes only. We report a measure of precision along the Z axis as the absolute difference in the Z coordinate for landmarks located during separate digitizations of a single scan.

Lastly, validation of data collected from CT images is accomplished by comparing these data to data collected from a previously validated method. The average linear distances between all landmarks were calculated from the two digitizations of the two images of each skull (totalling four digitizations). These average linear distances are compared with those measures obtained using the 3Space table-top digitizer. Differences between homologous linear distances measured using the two devices indicate error in data collected from the CT images.

Materials and Data Collection

Specimens

Ten skulls were chosen from The Johns Hopkins University, Department of Cell Biology and Anatomy skeletal teaching collection. Skulls were chosen on the basis of completeness and absence of breakage. The calvaria (skull cap) of each skull had been previously removed by sawing around the circumference of the vault. Nine of the ten skulls were adult specimens. The tenth skull represented an immature individual of approximately 6 years of age.

The skulls were scanned using the GE CT9800 housed in the Neuroradiology Department, The Johns Hopkins Hospital. Each skull underwent CT scanning twice using the same protocol (DFOV=20 cm; KV= 120; mA=100; 1.5-mm slice thickness). Once placed on the scanner table, a skull was not moved between the two scanning sessions. Images were sent across the Department of Radiology network to a Silicon Graphics workstation in The Johns Hopkins Outpatient Center. At this point, header and image data were separated and the slice image files (in raw data format) were sent electronically using file transfer protocol to an optical drive (Pinnacle

Micro, Irvine, California) attached to a Macintosh IIci located in JTR's laboratory in The Johns Hopkins School of Medicine.

Landmark Data Collection

Thirty-five craniofacial landmarks were identified and defined (Table 1, Fig. 3). The location of all landmarks was recorded three times directly from the skulls using the Polhemus 3Space table-top digitizer (Polhemus, Inc., Colchester, Vermont). This instrument has been subjected to previous validation (Hildebolt and Vannier, 1988) and precision (Corner et al., 1992) studies. Using these three sets of 35 landmark locations, an average form was created for each of the dry skulls following the algorithm given by Lele (1993). The average form represents the relative location of the 35 landmarks on a skull. These data were considered the gold standard for this study.

The CT slice image data were viewed on a Macintosh IIci using the program Image (version 1.47, Wayne Rasband, National Institutes of Health). Because our study was conducted to validate landmark data collected from CT images, every effort was made to correctly locate the landmarks on the images. As the landmarks were located on slice images, the coordinates were written to a spreadsheet. The X coordinate corresponds to pixel row, the Y coordinate corresponds to pixel column, and the Z coordinate is recorded as slice number representing the relative table position (Fig. 4). The Z coordinate for a landmark represents the center of the thickness of that slice. When a landmark appeared to lie within two slices, its Z coordinate was estimated as halfway between the center of the contiguous slices. These coordinate values were transformed to a metric scale using the following information: pixel size = 0.39 mm (X and Y coordinates) and slice thickness = 1.5 mm (Z coordinate). Resolution along the X and Y axes is approximately four times greater than along the Z axis.

Results

The results of the analysis of precision in locating landmarks on the same image are reported in Table 2. The average difference in landmark locations was less than 0.5 mm for all landmarks and was less than or equal to 0.2 mm for 13 of the 35 landmarks. The superior pterygopalatine fissure, foramen ovale, occipital condylar foramen, opisthion, and the jugular notches show the greatest amount of error. These results do not suggest that any particular landmark type (foramina, sutural intersections, bony protuberances) is disproportionately prone to error. In fact, the measure of precision in locating left and right expressions of the same landmark are often disparate.

Precision of the position of landmarks located on two separate images are reported in Table 3. These measures of the absolute difference in two images of one skull represent error due to imaging device. There are 595 unique linear distances among 35 landmarks. We present an average of the difference in homologous linear distances measured on the two scans (error for each of the linear distances are available upon

TABLE 1 Landmark Identification: Abbreviation and Description

Landmark Name	*Landmark Abbreviation*	*Landmark Definition*
Intradentale superior	IDS	point on alveolar edge between the two upper central incisors
Right pterygoid hamulus	RPTYHAM	inferior tip of the pterygoid hamulus, right side
Left pterygoid hamulus	LPTYHAM	see RPTYHAM, left side
Right greater palatine foramen	RPALFOR	center of greater palatine foramen on bony palate, right side
Left greater palatine foramen	LPALFOR	see RPALFOR, left side
Posterior nasal spine	PNS	most posterior point on the hard palate
Right inferior pterygo-palatine fissure	RINPTYFI	superior point of the intersection of the lateral pterygoid plate and palatine bones at the base of the pterygopalatine fissure, right side
Left inferior pterygo-palatine fissure	LINPTYFI	see RINPTYFI, left side
Basion	BAS	most anterior point on the foramen magnum located on the sagittal plane
Right condylar intersection with foramen magnum	RCONDISX	intersection of the occipital condyle with the edge of foramen magnum, right side
Left condylar intersection with foramen magnum	LCONDISX	see RCONDISX, left side
Right superior pterygo-palatine fissure	RSUPTYFI	intersection of the superior portion of the pterygoid plate and maxilla at the superior aspect of the pterygopalatine fissure, right side
Left superior pterygo-palatine fissure	LSUPTYFI	see RSUPTYFI, left side
Vomer-sphenoid junction	VSJ	intersection of the posterior aspect of the alae of the sphenoid with the basisphenoid
Right hypoglossal canal	RHYPOGL	center of the medial aspect of the hypoglossal canal, right side
Left hypoglossal canal	LHYPOGL	see RHYPOGL, left side
Right foramen ovale	RFOROVAL	center of foramen ovale, right side
Left foramen ovale	LFOROVAL	see RFOROVAL, left side
Right carotid canal	RCAROTID	center of the inferior opening of the carotid canal, right side
Left carotid canal	LCAROTID	see RCAROTID, left side
Right occipital condylar foramen	ROCCOND	center of occipital condylar foramen, right side
Left occipital condylar foramen	LOCCOND	see ROCCOND, left side
Opisthion	OPISTHIO	most posterior point on the foramen magnum located on the sagittal plane
Right jugular notch	RJUGNOTC	apex of bony protuberance of the pars lateralis of the occipital bone which makes up the lateral wall of the jugular foramen, right side
Left jugular notch	LJUGNOTC	see RJUGNOTC, left side
Right inferomedial point on the superior orbital fissure	RIMSORBF	intersection of the greater wing of the sphenoid, just above and anterior to foramen rotundum with the body of the sphenoid, right side
Left inferomedial point on the superior orbital fissure	LIMSORBF	see RIMSORBF, left side
Right petrous intersection	RPETISXN	intersection of the apex of the petrous temporal with the body of the sphenoid posterior to foramen lacerum, right side
Left petrous intersection	LPETISXN	see RPETISXN, left side
Right internal acoustic meatus	RINTACOU	center of the aperature for the internal acoustic meatus, right side
Left internal acoustic meatus	LINTACOU	see RINTACOU, left side
Right anterior clinoid process	RANTCLIN	most posterior projecting portion of the anterior clinoid process, right side
Left anterior clinoid process	LANTCLIN	see RANTCLIN, left side
Right superolateral point of the superior orbital fissure	RSLSORBF	intersection of the greater wing of the sphenoid with the less wing of the sphenoid, right side
Left superolateral point of the superior orbital fissure	LSLSORBF	see RSLSORBF, left side

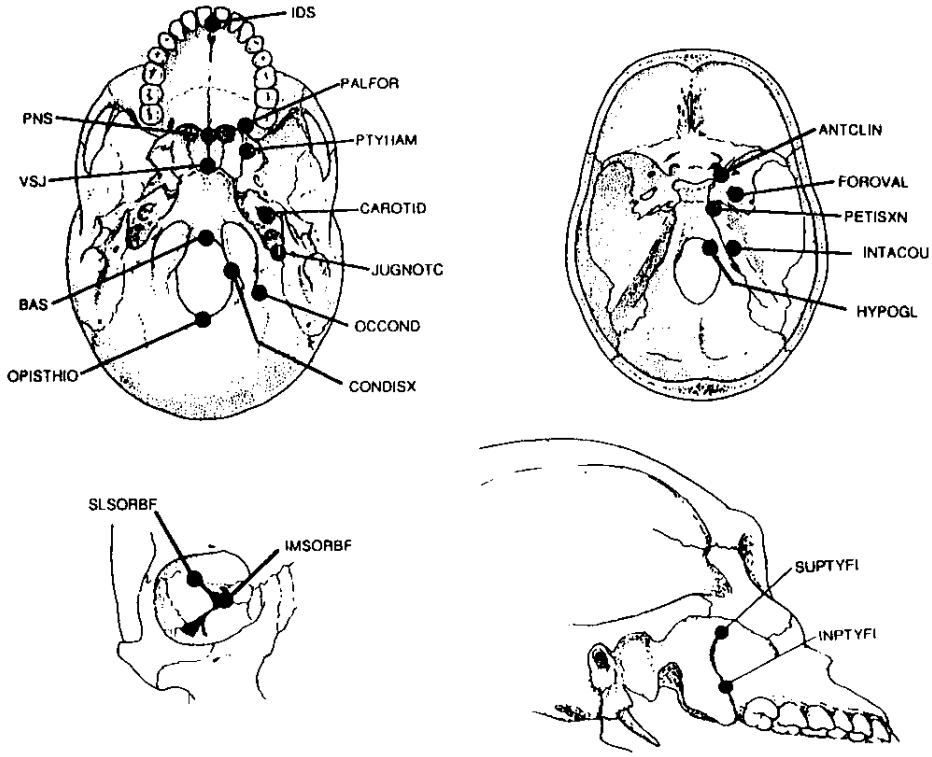

FIGURE 3 Location of landmarks used in study. Abbreviations for landmarks are used in Figure and further defined in Table 1. Views are defined as follows: upper left is an inferior view of the human skull; upper right a superior view of the cranial base, the calvaria has been removed; lower left is an anterior view of the right orbit showing landmarks that are located on the more internal aspect of the orbit associated with the cranial base; lower right is an oblique view, roughly inferolateral, to show the pterygopalatine fissure and associated landmarks.

request). An error of 0.2 mm would mean that like linear distances differ by 0.2 mm between scans. This could be interpreted in several ways; e.g., the position of one of the two landmarks varies by 0.2 mm, or that both landmarks differ in location by 0.1 mm between scans. Table 3 shows that the average error between scans is quite low, but the maximum error suggests that certain landmarks are more problematic than others. Close inspection of the data indicate that in those skulls for which the standard deviation of the error is large, outstanding errors for one or two landmarks are in large part responsible for the elevated measurement error. The raw data (available upon request) indicate that a disproportionate amount of error is associated with the following landmarks on a skull: intradentale superior, occipital condylar foramen, jugular notch, superior orbital fissure. Intradentale superior is the only landmark that showed a disproportionately large degree of error in more than one skull.

Repeatability of our data is summarized in Table 4. Results of our nested ANOVA summarize the average proportion of total variation in locating the X and Y coordinates explained by: (1) digitizing the image; (2) scanning the skull; and (3) anatomic variation among the skulls. In locating landmarks on CT scans along the X axis, an average of 0.32% of the total variation is due to differences in digitizing, and 1.21% of the total variation is due to difference in scanning. The average proportion of total variation along the X axis due to differences among skulls (repeatability) is 98.5%. Repeatability in locating landmarks along the Y axis is relatively higher (> 99.5%). Disparity between error along the X and Y axes can be explained by increased error along the X axis in locating a single landmark: intradentale superior. For one of our skulls, the first slice imaged by the CT scanner was initiated superior to the most inferior portion of the maxillary alveolus. For this reason, intradentale superior is not included in the most inferior scan taken for that skull and its location was estimated from the available data. Estimation of intradentale superior contributed more to error in the X than in the Y direction. Overall, the proportion of variance due to digitizing and imaging is remarkably low.

We could not use the ANOVA model to evaluate repeatability along the Z axis for reasons of registration discussed previously. Instead, we measured the precision of locating landmarks along the Z axis by comparing the Z coordinate of homologous landmarks collected during the two digitizations of each scan. The results are shown in Table 5. Forty-five percent of the scans (9/20) indicate that there is no error in locating landmarks along the Z axis for any of the 35 landmarks. Fifty-five (11/20) percent of the scans show errors of one half slice (0.75 mm) or more for one or more landmarks. The error of 3 mm for intradentale superior on skull 133 reflects the failure to capture that landmark in the image of skull 133. Landmarks associated with the pterygopalatine fossa, the jugular notch, and the pterygoid hamulus appear to have more frequent error in the Z direction. However, more striking is that of the 350 paired homologous points compared, 93% showed no error along the Z axis. Ignoring the error due to estimating intradentale superior on the second scan of skull 133, our greatest error in the Z direction was 1.5 mm (one slice).

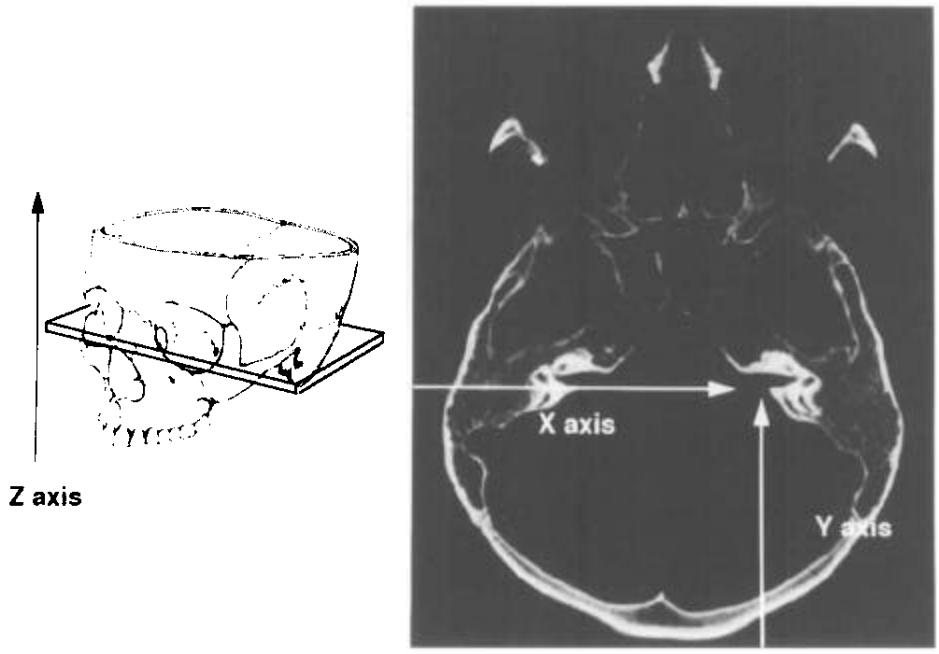

FIGURE 4 A single CT slice image through the cranial base. To locate a landmark in three-dimensional space using CT slice images, the slice image containing the landmark of interest is identified. Each slice image measures 512 × 512 pixels and is 1.5-mm thick. X and Y coordinates of the landmark are located according to pixel column and row on the image (for the CT scans used in this study a pixel = .039 mm). The Z coordinate for the landmark is assigned according to the table position of the slice in which the landmark is located. The landmark being located in this example is the left internal acoustic meatus.

Validation of data collected from CT images was computed by comparing the average linear distances between all possible pairs of landmarks (averaged over two digitizations of two scans for each skull) to the same linear distances calculated using coordinate data from the 3Space table-top digitizer. Since the 3Space digitizer is an already validated system, differences between homologous linear distances measured using the two devices indicate error in the landmark data collected from the CT images. Results of the validation analysis are summarized in Table 6.

On average, linear distances calculated from landmarks located on CT slice images differ from those calculated using the 3Space digitizer by 1 to 2 mm. Given the pixel size of our scans (pixel = 0.39 mm), linear distances measured using the CT slice images differ from those measured using the 3Space digitizer by two to four pixels. This could mean an error of two to four pixels in locating one of the landmarks defining

TABLE 2 Precision of Landmark Location between Digitizations within an Image. Differences Are Expressed in Millimeters and Averaged Over Images and Skulls

Landmarks	*Digitizing Error (mm)*
IDS	0.44
RPTYHAM	0.28
LPTYHAM	0.26
RPALFOR	0.12
LPALFOR	0.16
PNS	0.29
RINPTYFI	0.25
LINPTYFI	0.20
BAS	0.20
RCONDISX	0.20
LCONDISX	0.27
RSUPTYFI	0.37
LSUPTYFI	0.48
VSJ	0.22
RHYPOGL	0.15
LHYPOGL	0.39
RFOROVAL	0.32
LFOROVAL	0.30
RCAROTID	0.24
LCAROTID	0.20
ROCCONDF	0.28
LOCCONDF	0.38
OPISTHIO	0.39
LJUGNOTC	0.48
RJUGNOTC	0.31
RIMSORBF	0.14
LIMSORBF	0.16
RPETISXN	0.16
LPETISXN	0.18
RINTACOU	0.17
LINTACOU	0.31
RANTCLIN	0.23
LANTCLIN	0.20
RSLSORBF	0.35
LSLSORBF	0.32

TABLE 3 Precision of Landmark Location Measured as the Absolute Difference in Homologous Linear Distances Measured on the Two Images of a Single Skull (in Millimeters)

Skull	*Average Error (mm)*	*Maximum Error (mm)*	*Minimum Error (mm)*
66	0.17	1.17	0.00
103	0.53	2.39	0.00
133	0.45	3.66	0.00
111	0.27	2.33	0.00
141	0.22	0.99	0.00
147	0.44	2.50	0.00
151	0.23	1.66	0.00
29	0.21	0.82	0.00
40	0.21	1.60	0.00
5*	0.22	1.08	0.00

Among 35 separate landmarks, 595 unique linear distances are defined. We show the average, minimum, and maximum difference in linear distances defined on the two scans. Measures specific to each linear distance are available upon request.

*juvenile skull

TABLE 4 Variance Components (in Percentages) for X and Y Axis Coordinates Explained by Digitizing, Scanning, and Individual (Variability between Skulls or Repeatability) for CT Slice Images

Landmarks	*X AXIS*			*Y AXIS*		
	Among Skulls	*Scans Within Skulls*	*Digitization Within Scans*	*Among Skulls*	*Scans Within Skulls*	*Digitizations Within Scans*
IDS	85.33	13.86	0.81	99.92	0.07	0.01
RPTYHAM	98.96	0.66	0.37	99.96	0.02	0.02
LPTYHAM	99.01	0.80	0.20	99.93	0.05	0.02
RPALFOR	98.73	1.22	0.06	99.89	0.10	0.01
LPALFOR	98.96	0.98	0.05	99.88	0.10	0.02
PNS	98.32	0.97	0.71	99.89	0.08	0.04
RINPTYFI	99.57	0.00	0.43	99.95	0.03	0.02
LINPTYFI	99.27	0.68	0.06	99.94	0.03	0.03
BAS	99.25	0.37	0.37	99.90	0.08	0.01
RCONDISX	99.37	0.48	0.16	99.92	0.05	0.03
LCONDISX	99.39	0.15	0.46	99.86	0.11	0.03
RSUPTYFI	99.35	0.18	0.46	99.81	0.13	0.06
LSUPTYFI	99.62	0.18	0.21	99.92	0.04	0.04
VSJ	99.01	0.70	0.28	99.91	0.05	0.03
RHYPOGL	99.42	0.38	0.19	99.85	0.10	0.02
LHYPOGL	99.17	0.34	0.49	99.93	0.04	0.04
RFOROVAL	98.93	1.03	0.14	99.90	0.06	0.04
LFOROVAL	99.94	0.96	0.50	99.91	0.05	0.04
RCAROTID	99.18	0.53	0.29	99.91	0.05	0.04
LCAROTID	99.38	0.31	0.31	99.94	0.04	0.02
ROCCONDF	94.79	5.07	0.15	99.84	0.10	0.06
LOCCONDF	94.78	4.99	0.22	99.92	0.00	0.08
OPISTHIO	98.90	0.41	0.69	99.94	0.03	0.04
RJUGNOTC	99.82	0.12	0.06	99.74	0.16	0.10
LJUGNOTC	99.27	0.56	0.18	99.52	0.43	0.05
RIMSORBF	98.96	0.83	0.21	99.73	0.25	0.02
LIMSORBF	99.48	0.37	0.15	99.71	0.23	0.05
RPETISXN	99.39	0.49	0.12	99.91	0.06	0.04
LPETISXN	98.62	1.08	0.30	99.88	0.05	0.06
RINTACOU	99.44	0.42	0.14	99.78	0.16	0.06
LINTACOU	98.82	0.67	0.52	99.88	0.08	0.04
RANTCLIN	98.60	0.56	0.84	99.86	0.06	0.08
LANTCLIN	98.94	0.63	0.44	99.78	0.17	0.05
RSLSORBF	99.56	0.24	0.20	99.69	0.15	0.16
LSLSORBF	98.68	0.99	0.32	99.74	0.19	0.04

TABLE 5 Summary of Z-coordinate Error Organized by Scans for Each Skull

Skull	*Scan*	*Error Along Z Axis: Landmark and Value of Error*
66	1	no error
	2	no error
103	1	LJUGNOTC (1.5 mm)
		LPTYHAM (0.75 mm)
	2	RSUPTYFI (1.5 mm)
		LSUPTYFI (1.5 mm)
		LHYPOGL (1.5 mm)
		OPISTHIO (0.75 mm)
133	1	LSUPTYFI (0.75 mm)
		RFOROVAL (1.5 mm)
	2	IDS (3mm)**
		LPTYFI (1.5 mm)
		LSUPTYFI (0.75 mm)
		LFOROVAL (0.75 mm)
111	1	RFOROVAL (1.5 mm)
	2	no error
141	1	no error
	2	no error
147	1	no error
	2	PNS (0.75 mm)
151	1	no error
	2	OPISTHIO (0.75 mm)
29	1	no error
	2	RPTYHAM (1.5 mm)
		LPTYHAM (1.5 mm)
40	1	RSUPTYFI (1.5 mm)
		LSUPTYFI (1.5 mm)
		RJUGNOTC (1.5 mm)
		LJUGNOTC (1.5 mm)
	2	no error
5*	1	LSUPTYFI (1.5 mm)
		LJUGNOTC (1.5 mm)
	2	IDS (1.5 mm)

*juvenile skull
**reasons for this large error are discussed in text
Slice thickness = 1.5 mm

the linear distance or some division of the error between the two landmarks that define the linear distance.

In terms of the specifics of craniofacial bony anatomy, our study points to several landmarks that are particularly prone to validation error. A tabulation of the raw validation results (available upon request) shows that all landmarks were endpoints of at least one linear distance that showed a validation error of greater than 2.5 mm. Landmarks that showed this kind of error more frequently than others include: intradentale superior, vomer-sphenoid junction, hypoglossal canal, carotid canal, occipital condylar foramen, superior orbital fissure—superolateral point. Although the landmarks mentioned above can be found with acceptable precision on a scan, the coordinate location recorded may be inaccurate. Our results suggest that when recording the center of a rather large foramen or canal, much of the error may lie in estimating the center of the aperture. Foramina might be more accurately recorded by choosing a locus on the perimeter rather than the center, or by using imaging software to calculate the center of the foramen from pixel density information.

Discussion

Our study has measured the precision and repeatability of three-dimensional landmark data collected from CT images of 1.5-mm slice thickness and validated those data. Landmark data can be used as an accurate representation of the three-dimensional geometry of biologic points on forms that have undergone CT scanning. Since the identification of exact loci is prerequisite to the definition of any region of interest (e.g., lines, curves, areas, volumes, surfaces), our analysis suggests that more complex maps of biologic geometries collected from CT scans may also be of acceptable precision. Landmarks, or other types of data that can be defined irrespective of a particular coordinate system are coordinate-system invariant (Lele, 1993). Coordinate-system invariant landmarks can be repeatedly located regardless of whether the form is spun on an axis, moved from left to right, or flipped upside down. An example of a landmark that is not coordinate-system invariant is glabella, defined as the most anterior projecting point of the frontal bone. The location of this landmark will shift depending upon the orientation (coordinate system) of the skull. Adding the coordinate system to the definition of a landmark adds another level of potential error to landmark identification. Curves, outlines, and surfaces are sometimes defined on

TABLE 6 Validation of 3D Coordinate Data Collected from CT Slice Images (Measures in Millimeters)

Skull	*Scan*	*Average Absolute Difference**	*SD†*
66	1	1.90	1.91
66	2	1.91	1.88
103	1	1.43	1.26
103	2	1.32	1.14
133	1	1.47	1.72
133	2	1.40	1.54
111	1	1.72	1.93
111	2	1.78	1.92
141	1	1.37	1.33
141	2	1.39	1.36
147	1	2.37	3.15
147	2	2.28	3.13
151	1	1.13	1.04
151	2	1.20	1.16
29	1	1.53	1.52
29	2	1.50	1.48
40	1	1.29	1.12
40	2	1.35	1.13
5	1	1.64	1.83
5	2	1.66	1.86

*between linear distances measured using 3Space digitizer and those measured on CT images
†SD of absolute difference between linear distances measured using 3Space digitizer and those measured on CT images

the basis of a particular orientation and for that reason are not coordinate-system invariant. They should be identified and used with this caveat in mind.

We remind the reader that the results of this study reflect an idealized experimental design that allowed us to determine and minimize our error. First, we scanned dry skulls which were void of soft tissue that can obscure visualization of bony anatomy. Second, we had the luxury of choosing the specimens for scanning and refining our data collection techniques by repeated episodes of data collection. The idealized experimental situation was used primarily to offer a means of evaluation of these data by direct comparison with an already validated instrument, but also allowed determination of the kind and degree of error that can be expected from CT data. Due to our research design, the precision results probably represent the *minimal* error expected from CT data. Data collected from patient scans may contain a higher degree of error.

Although our study showed acceptable precision in locating landmarks along the Z axis, resolution along the Z axis (1.5-mm slice thickness) is nearly four times less than the resolution along the X and Y axes (1 pixel = 0.39 mm). Resolution along the Z axis and our use of table position as the Z coordinate may be partly responsible for our validation results. Since table position reflects the center of a slice, the Z coordinate of any landmark within that slice was projected onto the plane of the center of the slice. The measure of an object being imaged loses the equivalent of a full slice along the axis parallel to the movement of the gantry (in our study, the Z axis). The distance of a line perpendicular to the XY plane with endpoints at the extreme bottom of the first slice and at the extreme top of the last slice will be off by a measure that is equal to slice thickness (Fig. 5). This configuration may contribute to our validation errors. Reconstruction routines that allow the user to page through a slice at intervals less than the slice thickness may improve the accuracy of the Z coordinate in landmark placement.

Our validation results indicate that data collected from CT slice images are internally consistent, but they should not be used in combination with data collected directly from individuals (i.e., anthropometric measures taken with calipers). Since the CT data are sufficiently precise and accurate, studies using CT data exclusively are justified. Data from CT slice images are appropriate for tracking growth of a single child (longitudinal study), for a group of children (mixed longitudinal study), or for the comparison of measures between individuals or samples as long as all data are from CT scans. Measures taken from CT slice images should not be compared with, or used in combination with measures taken directly from individuals or cadaver specimens without further validation of the CT data. Such measures for which clinical standards exist (e.g., bi-occipital breadth, head circumference) should not be used to evaluate patient measurements collected from CT slice images.

Locating landmarks on a three-dimensional CT reconstruction may prove to be less labor intensive than finding them on slice images. No new data are added to a reconstruction, so there is no reason to expect improved accuracy and precision (Ohman and Richtsmeier, 1994). However, since a three-dimensional reconstruction more accurately reflects the biologic object and no special training in axial anatomy is required to identify and locate landmarks, it is possible that the use of three-dimensional reconstructions in place of slice images could result in error reduction. Reconstruction programs often allow the user to reslice the three-dimensional reconstruction along any axis. This could improve precision resulting in CT data that fit more closely to data collected from already validated instruments. Finally, landmarks located on facial and neurocranial surfaces are not easily visualized on axial slice images (e.g., the intersection of the zygomatic and maxillary bones at the inferior orbital rim; zygomaticofacial foramen; bregma, asterion), and may be more easily seen on a three-dimensional reconstruction. This would increase the number of landmarks available for study.

We have shown that three-dimensional landmark data collected from CT slice images are internally consistent and precise. CT images enable visualization of structures normally

FIGURE 5 Graphic representation of the potential loss in accuracy of determining the Z coordinate of a point. Actual anatomic location of a point on the Z axis can only be determined with an error of ± one half of the slice thickness.

seen only in the operating room or dissection laboratory and provide a means for preoperative planning and postoperative evaluation. The potential of CT data for quantitative assessment of patient and population parameters has not yet been fully realized. The challenge lies in developing new tools for data collection and analysis in three dimensions (e.g., Rohlf, 1990; Richtsmeier et al., 1992; Lele, 1993; Rohlf and Marcus, 1993) and creative designs for use of these data in solving problems defined by craniofacial biology.

Acknowledgments. We thank Dr. Nick Bryan, Department of Neuroradiology, The Johns Hopkins University for access to the GE scanner used in this study and Dr. Subhash Lele and Dr. Jim Cheverud provided helpful comments on the analytic design of this paper.

References

Boas F. The horizontal plane of the skull and the general problem of the comparison of variable forms. Science 1905;21:862–863.

Bookstein FL, Strauss RE, Humphries JM, Chernoff B, Elder RL, Smith GR. A comment upon the use of Fourier methods in systematics. Syst Zool 1982;31:92–95.

Broca P. Instruction craniologiques et craniométriques. Mém. de la Soc. d'Anthrop. de Paris, 2nd series, 1875;Vol. 2:1–203.

Christiansen EL, Thompson JR, Kipp S. Intra- and interobserver variability and accuracy in the determination of linear and angular measurements in computed tomography: an in vivo and in situ study of human mandibles. Acta Odontol Scand 1986;44:221.

Corner BD, Lele SL, Richtsmeier JT. Measuring precision of three-dimensional landmark data. Quantitative Anthrop 1992;3:347–359.

Corner BD, Richtsmeier JT. Cranial growth in the squirrel monkey (*Saimiri sciureus*): a quantitative analysis using three-dimensional coordinate data. Am J Phys Anthropol 1992a;87:67–82.

Corner BD, Richtsmeier JT. Experiments in nature: premature unicoronal cranial synostosis in mantled howler monkeys (*Alouatta palliata*). Cleft Palate Craniofac J 1992b;29:143–151.

Dufresne C, Richtsmeier JT. The interaction of craniofacial dysmorphology, growth and the prediction of surgical outcome. J Craniofac Sur. 1995; *In press.*

Fernbach SK, Feinstein KA. The deformed petrous bone: a new plain film sign of premature lambdoid synostosis. Am J Roentgenol 1992;156: 1215–1217.

Fok H, Jones BM, Gault DG, Andar U, Hayward R. Relationship between intracranial pressure and intracranial volume in craniosynostosis. Br J Plast Surg 1992;45:394–397.

Gault DT, Renier D, Marchac D, Jones BM. Intracranial pressure and intracranial volume in children with craniosynostosis. Plast Reconstr Surg 1992;90:377–381.

Genitori L, Cavalheiro S, Lena G, Dollo C, Choux M. Skull base in trigonocephaly. Pediatr Neurosurg 1991–1992;17:175–181.

Goodall C. Procrustes methods in the statistical analysis of shape. J R Stat Soc Series B, 1991;53:285–339.

Gower JC. Generalized procrustes analysis. Psychometrika 1975;40:33–50.

Hall BK, ed. Homology: the hierarchical basis of comparative biology. San Diego: Academic Press, 1994.

Hildebolt CF, Vannier MW. Three-dimensional measurement accuracy of skull surface landmarks. Am J Phys Anthropol 1988;76:497–503.

Kohn LAP, Cheverud JM. Calibration, validation, and evaluation of scanning systems: anthopometric imaging system repeatability. In: Vannier MW, Yates RE, Whitestone JJ, eds. Electronic imaging of the human body. Proceedings of a working group. Dayton, OH: CSERIAC, 1992:114–123.

Leboucq N, Montoya P, Martinez Y, Castan Ph. Value of 3D imaging for the study of craniofacial malformations in children. J Neuroradiol 1991;18:225–239.

Leboucq N, Montoya P, Martinez Y, Castan Ph. Les craniostenoses lambdoides. J Neuroradiol 1993;20:24–33.

Lele S. Some comments on coordinate-free and scale-invariant methods in morphometrics. Am J Phys Anthropol 1991;85:407–418.

Lele S. Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. Math Geol 1993;25:573–602.

Lele S, Richtsmeier JT. Euclidean distance matrix analysis: a coordinate free approach for comparing biological shapes using landmark data. Am J Phys Anthropol 1991;86:415–427.

Lele S, Richtsmeier JT. On comparing biological shapes: detection of influential landmarks. Am J Phys Anthropol 1992;87:49–66.

Martin R. Lehrbuch der Anthropologie in systematishcher Darstellung. 2nd ed., Vol. 2. Kraniologie, Osteologie. Jena: Gustav Fischer, 1928.

Matteson SR, Bechtold W, Phillips C, Staab EV. A method for three-dimensional reformation for quantitative cephalometric analysis. J Oral Maxillofac Surg 1989;47:1053.

Ohman J, Richtsmeier JT. Perspectives on craniofacial growth. Clin Plas Surg 1994;21:489–499.

Papillaut G. Entente internationale pour l'unification des mesures craniométriques et cephalométriques. L'Anthropologie 1906;17:559–572.

Posnick JC, Bite U, Nakano P, Davis J, Armstrong D. Indirect intracranial volume measurements using CT scans: clinical applications for craniosynostosis. Plast Reconstr Surg 1992;89:34–45.

Read DW, Lestrel P. Comments on the uses of homologous-point measures in systematics: a reply to Bookstein et al. Syst Zool 1986;35: 241–253.

Richardson ML, Van Vu M, Vincent LM, Saneorzan BJ, Benirschke SK. CT measurement of the calcaneal varus angle in the normal and fractured foot. J Comput Assist Tomogr 1993;16:261–264.

Richtsmeier JT. Beyond morphing: visualization to predict a child's skull growth. Advanced Imaging 1993; 8:24–27.

Richtsmeier JT, Lele S. Quantitative analysis of growth: models and theoretical considerations. Biol Rev 1993;68:381–411.

Richtsmeier JT, Walker A. Morphometric analysis of facial growth in *Homo erectus*. In: Leakey RE, Walker A, eds. The Nariokotome *Homo erectus* skeleton. Cambridge: Harvard University Press, 1993:391–410.

Richtsmeier JT, Cheverud JM, Lele S. Advances in anthropological morphometrics. Ann Rev Anthropol 1992;21:283–305.

Richtsmeier JT, Cheverud JM, Corner BD, Danahey SM, Lele S. Sexual dimorphism of ontogeny in the crab eating macaque (*Macaca fascicularis*). J Hum Evol 1993a;25:1–30.

Richtsmeier JT, Corner BD, Grausz HM, Cheverud JM, Danahey S. The role of postnatal growth pattern in the production of facial morphology. Syst Biol 1993b;42:307–330.

Richtsmeier JT, Sack GH, Grausz HM, Cork LC. Cleft palate with autosomal recessive transmission in Brittany spaniels. Cleft Palate Craniofac J 1994;31:364–371.

Rohlf FJ. Morphometrics. Ann Rev Ecol Syst 1990;21:299–316.

Rohlf FJ, Bookstein FL. Proceedings of the Michigan Morphometrics Workshop. The University of Michigan Museum of Zoology, Ann Arbor. Special Publication No. 2, 1990.

Rohlf FJ, Marcus LF. A revolution in morphometrics. TREE 1993;8:129–132.

Rohlf FJ, Slice D. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool 1990;39:40–59.

SAS Institute. SAS/STAT User's guide. Version 6, 4th ed. Vol. 2. Cary, NC: SAS Institute Inc., 1989.

Siegel AF, Benson RH. A robust comparison of biological shapes. Biometrics 1982;38:341–350.

Speculand B, Butcher GW, Stephens CD. Three-dimensional measurement: the accuracy and precision of the Reflex microscope. Br J Oral Maxillofac Surg 1988;26:276–283.

Vannier MW, Marsh JL, Warren JO. Three-dimensional CT reconstruction images for craniofacial surgical planning and evaluation. Radiology 1984;150:179–184.

Waitzman AA, Posnick JC, Armstrong DC, Pron GE. Craniofacial skeletal measurements based on computed tomography: Part I. Accuracy and reproducibility. Cleft Palate Craniofac J 1992a;29:112–117.

Waitzman AA, Posnick JC, Armstrong DC, Pron GE. Craniofacial skeletal measurements based on computed tomography: Part II. Normal values and growth trends. Cleft Palate Craniofac J 1992b;29:118–128.

Wong L, Dufresne CR, Richtsmeier JT, Manson PM. The effect of rigid fixation on the growing craniofacial skeleton: periosteal elevation, coronal suture plating, and frontal bone plating. Plast Reconstr Surg 1991; 88:395–403.