J Forensic Sci, November 2008, Vol. 53, No. 6
doi: 10.1111/j.1556-4029.2008.00852.x
Available online at: www.blackwell-synergy.com

Carl N. Stephan,[1] *Ph.D. and Ellie K. Simpson,*[2] *Ph.D.*

Facial Soft Tissue Depths in Craniofacial Identification (Part I): An Analytical Review of the Published Adult Data*

ABSTRACT: With the ever increasing production of average soft tissue depth studies, data are becoming increasingly complex, less standardized, and more unwieldy. So far, no overarching review has been attempted to determine: the validity of continued data collection; the usefulness of the existing data subcategorizations; or if a synthesis is possible to produce a manageable soft tissue depth library. While a principal components analysis would provide the best foundation for such an assessment, this type of investigation is not currently possible because of a lack of easily accessible raw data (first, many studies are narrow; second, raw data are infrequently published and/or stored and are not always shared by some authors). This paper provides an alternate means of investigation using an hierarchical approach to review and compare the effects of single variables on published mean values for adults whilst acknowledging measurement errors and within-group variation. The results revealed: (i) no clear secular trends at frequently investigated landmarks; (ii) wide variation in soft tissue depth measures between different measurement techniques irrespective of whether living persons or cadavers were considered; (iii) no clear clustering of non-Caucasoid data far from the Caucasoid means; and (iv) minor differences between males and females. Consequently, the data were pooled across studies using weighted means and standard deviations to cancel out random and opposing study-specific errors, and to produce a single soft tissue depth table with increased sample sizes (e.g., 6786 individuals at pogonion).

KEYWORDS: forensic science, soft tissue thickness, facial approximation, facial reproduction, facial reconstruction, superimposition

Facial soft tissue depths are determined by measuring at various points on the face, the distance from the skin surface to the most superficial surface of the underlying hard tissue. These measurements, therefore, describe how the face fits over the skull but they do so in a general way because a variety of organs are encompassed in any single measurement and mean values are generally calculated (1). Discriminative information is not provided about any single anatomic component of the face (e.g., fat or muscle) nor do these soft tissue depths give precise estimations of any individual's soft tissue thickness [but see Simpson and Henneberg (2) for attempts to increase the specificity of these data]. Despite this, soft tissue depth measurements hold a significant role in both facial approximation and craniofacial superimposition methods because they provide a basis for quantification and thus, repeatability.

To date, the research on soft tissue depths has been extensive: more than 60 studies exist ranging in publication dates from 1883 to 2007. Most examine a variety of subsamples and overall, more than 6700 sets of data have been collected from adults, representing more than 103,100 individual soft tissue depth measurements on human faces.

A variety of measurement techniques have been used to measure the soft tissue depths [including needle puncture, clinical calipers, radiographs (2D and 3D), ultrasound (A-mode and B-mode), computed tomography (CT), and magnetic resonance imaging (MRI)] and sample sizes have been broad, ranging from 1 to 297 individuals. However, group sizes generally tend to include fewer than 40 individuals for averaging (3). Numerous age groups have been investigated, including individuals aged as young as 1 year (4) to as old as 101 years (2). Furthermore, many population groups have been investigated, including: Australian Aborigines; nonaboriginal Australians; Black Americans; British; Chinese; Egyptians; Germans; Japanese; Swiss; and Zulus to name just a few (see Tables 1 and 2). It should be noted, however, that precise and explicit criteria for the assignment of individuals to any one of these population groups is rarely established or meticulously enforced (5–19). Rather, assignment of individuals is often based on subjective interpretations of a person's physical appearance or place of study conducted, in the absence of supporting genetic data or participant self-reports of ancestral background.

Despite the value of a large quantity of data from such a wide section of the human population, the breadth of soft tissue data nonetheless compromises its usefulness. Measurements have been extensively subcategorized, and even if concerning similar samples, they have been independently segregated when produced by different authors. A predicament, therefore, arises as which data set is the most appropriate to use for an individual coming from a commonly measured population group. For example, a large number of studies report different mean values for males of European extraction, aged 50 years (see e.g., 20–26). Furthermore, the data are often reported or measured in different ways which decreases standardization and complicates the selection of the data for forensic casework.

It is often stated that studies on living people should be given higher priority than those on cadavers (3,27), but no formal empirical evidence has ever been presented to justify this standpoint in specific regard to facial soft tissue depths. Simpson and Henneberg (2) provide some circumstantial differences (they studied fresh, embalmed, and embalm-cured cadavers) but then the magnitude of the mean differences in this study was only slightly larger than the

[1]Anatomy and Developmental Biology, School of Biomedical Sciences, The University of Queensland, Brisbane, Qld 4072, Australia.
[2]Forensic Science South Australia, Adelaide, SA 5000, Australia.

*Portions of this work have been presented at the 59th Annual Meeting of the American Academy of Forensic Sciences February 2007, in San Antonio, TX, U.S.A.

Received 28 Oct. 2007; and in revised form 16 Feb. 2008; accepted 3 Mar. 2008.

TABLE 1—*Soft tissue depth studies of females drawn from the literature and used to assess each variable examined in this study.*

Publication Year (Caucasoid Data)	Measurement Methods (Caucasoid Data)	Race	Sex
Needle puncture studies: (2,21,22,26,29,43,60,65,82,100,101), Rhine (56 cited in 57) radiographic studies: (4,39,40,47,51,75,80,81,100,102,103), Köstler (1940 cited in 25), Weining (54 cited in 25), Bankowski (53 cited in 25)	Needle puncture studies: (2,21,22,26,29,43,60,63,65,82,100), Rhine (56 cited in 57) Radiographic studies: (4,39,40,47,51,75,80,81,102,103), Köstler (1940 cited in 25), Weining (54 cited in 25), Bankowski (53 cited in 25) Ultrasound studies: (24,25,94) CT study: (79) MRI studies: (43,70)	All the data for Measurement Methods plus additional data from: (3,24,44,48,66,74), Rhine (55 cited in 57)	(2–4,21,22,24–26,29,39,40,43,44,47, 48,51,60,63,65,66,69,70,74,75, 80–82,94,100–106), Köstler (1940 cited in 25), Weining (54 cited in 25), Bankowski (53 cited in 25), Rhine (55 cited in 57), Rhine (56 cited in 57)

CT, computed tomography; MRI, magnetic resonance imaging.

TABLE 2—*Soft tissue depth studies of males drawn from the literature and used to assess each variable examined in this study.*

Publication Year (Caucasoid Data)	Measurement Methods (Caucasoid Data)	Race	Sex
Needle puncture studies: (2,20–23,26,29,43,60,65,82,100,101,107), Rhine (56 cited in 57) Radiographic studies: (4,39,40,47,51,75,80,81,100,102,103), Weiβer (1940 cited in 25), Weining (54 cited in 25), Bankowski (53 cited in 25)	Needle puncture studies: (2,20–23,26,29,43,60,65,82, 100,101,107), Rhine (56 cited in 57) Radiographic studies: (4,39,40,47,51,75,80,102, 103,108,109), Weiβer (1940 cited in 25), Weining (54 cited in 25), Bankowski (53 cited in 25) Ultrasound studies: (24,25,94) CT study: (79) MRI studies: (43,70)	All the data for measurement methods and additional data from the following: (3,24,44,46,66,68,74), Rhine (55 cited in 57)	(2–4,20–26,29,39,40,43–47,51, 60,63,65,66,68–70,72–75, 79–82,94,100–105,107–112), Fisher and Moorman (year unknown, cited in 52), Weiβer (1940 cited in 25), Bankowski (53 cited in 25), Weining (54 cited in 25), Rhine (55 cited in 57), Rhine (56 cited in 57)

CT, computed tomography; MRI, magnetic resonance imaging.

reported measurement errors (1–2 mm). Technologically advanced techniques (ultrasound/MRI) also tend to be favored over more ''primitive'' methods (i.e., needle puncture), but yet the research indicates that all measurement methods have similar measurement errors (2,28–33), thus suggesting that no method holds strong advantages over any other in terms of precision. Furthermore, the prior literature has heavily relied on the results of statistical significance tests for establishing differences between the sexes and so-called ''races'' (typically Negroid, Caucasoid, and Mongoloid) with little regard to the practicality of the size of the differences actually observed (32,33), an approach which is widely known to risk overemphasis of any statistically indicated differences.

This paper examines the published soft tissue depth data in an attempt to determine what data distinctions are justified, and if any synthesis or simplification is possible. While a multivariate analysis, such as principal component analysis, would provide the best basis for such an examination, this approach is (currently) impossible since comprehensive single studies have not been undertaken and because raw data of smaller studies are infrequently published, rarely stored, and are not always shared by some authors. An alternate approach of comparing published data means is, therefore, used to analyze trends and differences across studies in a univariate fashion. Regard was given to measurement errors and other forms of data uncertainty (e.g., choice of measurement method and error in physical positioning of soft tissue depths in forensic casework) in addition to the results of statistical significance tests. The conclusions drawn are, therefore, guided by empirical values but are not subject to mechanical decision-making hinged solely upon statistical significance tests. This is important because large measurement errors can explain the differences between groups if these group differences are no greater than data uncertainty levels and because small (statistically significant) differences may be negligible in practical terms.

Past investigations using needle puncture techniques on cadavers (2,29,34), ultrasound techniques on living persons (28,30,31), and MRI (see 32,33) indicate that measurement errors are generally 10% of the measurement value, but can be higher than 30% at some landmarks (see e.g., 2). At moderate soft tissue depths, the absolute error equates to 1–2 mm. However, large measurements (such as mid ramus or gonion) tend to have higher measurement errors than sites that yield smaller measurement values (e.g., rhinion or nasion; 2,29,31). Note here that total measurement error can only be assessed by carrying out the entire data collection procedure afresh on the same subjects. Studies that leave clues as to prior landmark placement (see e.g., 29) or do not replicate that entire measurement process (i.e., re-use original radiographs or imaging scans, see e.g., 32) are likely to underestimate the true error involved.

The typical error rate of 10% may seem relatively small, but it is twice the accepted threshold for scientific statistical significance tests and is comparable in magnitude to many of the differences typically found between different data sets (see e.g., almost any published study on soft tissue depths that carries out statistical significance tests between the sexes or ''races''). It is also important to note here that measurement error is only one contributor to the total uncertainty that surrounds soft tissue depth measurements. Other potential sources of uncertainty include: choice of soft tissue depth measurement method (30,35)

and errors associated with the physical placement and representation of the soft tissue depths on skulls in casework (for further discussion and examples see 36). Given these sources for data uncertainty, the total amount of error associated with the use of soft tissue depths can only be conservatively expected to exceed 2 mm.

General Materials and Methods

Data Organization

Literature searches were conducted for publications concerned with mean facial soft tissue depths using Medline (Silverplatter), Current Contents, and traditional methods (reference lists of other articles). Data from 66 studies were obtained. Two papers, one by Smith and Buschang (37) and the other by Helwin (38) essentially repeated findings of former studies and so were excluded (see 39,40). Also excluded were the studies by Altemus (41) and Burstone (42), which used unorthodox methods to measure horizontal distances between hard and soft tissue profiles rather than distances between defined hard and soft tissue landmarks. Thus, 62 of the 66 studies found from the literature were used in this analysis (see Tables 1 and 2). It is worth noting here that this sample includes several studies that are not common to the mainstream craniofacial identification literature including: an unpublished data set (Anderson and Henneberg, personal communication), data from two theses (26,43), some infrequently cited craniofacial identification papers (40,44–46), and a number of data sets drawn from other disciplines (see e.g., 47–51).

Upon collation, many inconsistencies were found between the original data and reproductions made by other authors. This suggests that a cautious approach needs to be taken when referring to "second hand" citations. Thus, every effort was made to obtain original manuscripts for this investigation. This was not possible in seven cases, however, we continue to include these "cited but unseen" data for completeness as they represent a small component of the total sample of this study in terms of sample size and because some are widely referenced in the literature. The studies concerned are: Fisher and Moorman (year unknown) cited in (52); Weiβer (1940) cited in (25); Bankowski (53) cited in (25); Köstler (1940) cited in (25); Weining (54) cited in (25); and Rhine (55,56) cited in (57).

By re-evaluating the raw data available for some studies (e.g., 20,22), previously unreported standard deviations were also calculated and used. In some cases, analyses of the raw data revealed minor calculation errors which led to alternate values being used in this paper. For example, our value for Welcker's (20) "chin point" is 8.6 mm in contrast to the 8.5 mm originally reported. Also, it was evident from Kollman and Büchly's (22) paper that 22 males and 7 females were measured in total; with sample sizes varying depending on which landmarks were considered. This contrasts with common reports that Kollman and Büchly examined 21 males and 4 females for every measurement point (see e.g., 52,58; O'Grady and Taylor cited in 27,59). The smaller sample size reported in the literature is because of the exclusion of individuals classified as "very lean," but the validity of this exclusion rule is dubious since it is inconsistently applied. That is, the smaller sample also includes another extreme individual, classified as "very well fed." Thus, all individuals measured by Kollman and Büchly (22) were included in this study.

Minor discrepancies were also noted between the data reported by Czekanowski (60) and Martin (52), and the raw data presented by Czekanowski in the appendix of the original paper. In the raw data, Czekanowski reports on 68 males and 52 females (n = 120) in contrast to 65 males and 54 females (n = 119) reported in the text (52,60). The averages we generated from these raw data differed slightly from the published values (by about 1 mm), but because of technical difficulties associated with translation, we gave the benefit of our doubt to Czekanowski and Martin, and used their previously cited values. This procedure precluded the use of standard deviations for these data, although they potentially could have been calculated.

The 62 studies outlined above were collated under two broad age groups: adults (equal to or greater than 18 years) and subadults (<18 years). Data were split at the 18-year point, despite the fact that adolescent biological growth continues in males into the third decade (61,62), because the data had predominantly been categorized in this fashion previously and thus did not lend themselves to any re-arrangement. Papers reporting on samples that spanned this division point were classified based on whichever age group dominated the sample. Data from one individual included in the German sample by Stadtmuller (63) were simply reported as being "young," and as we could not determine without doubt whether or not this individual was indeed an adult or sub-adult, the data for this person were excluded. These procedures resulted in 55 studies related to adults (see Tables 1 and 2; also note that seven of these studies also concerned subadults), and a further seven studies that considered subadults alone. Only the adult data are addressed in this paper; the subadult data are discussed in part two of this manuscript (see 64).

Although some studies subclassified individuals according to body fatness (22,55,65,66), we included all individuals in this study because: (i) not all studies used a body weight classification; (ii) studies that did use such a classification did not always use identical group definitions; and (iii) body weight cannot yet be determined accurately from skeletal remains so weight-categorized data are typically of limited value. Weighted means (and standard deviations) were, therefore, calculated across all body weight categories to obtain overall study values. As many studies have excluded subjects using fatness variables (see e.g., 2,3,67–71), the data used in this analysis are probably biased toward those individuals thought to be "normal" rather than a true random sample.

In order to compare data between different studies, identical but differently named landmarks were reclassified to a common anthropometric point. For example, "*Hőchster Punkt des Jochbogens*" (22,63; translated as the widest point on the zygomatic arch), mid-zygomatic arch (O'Grady and Taylor cited in 59; Rhine et al. cited in 57), and other similar, though slightly different descriptions (45,60,72–74) were reclassified as "zygion." Similarly, measurements from point B (47,75), chin lip fold (24,44), chin fissure (2), and labiomental groove (76), etc., were reclassified as "mentolabial sulcus."

Reclassifications were also made for other landmarks where the original terms were found to be inappropriate. For example, the landmark "alare" which is often used to imply the soft tissue depth taken from the junction of the alar wing with the face (see e.g., 2,25), but which actually defines as the most lateral point on the contour of each alar margin (77) was reclassified as the "alar curvature point"—defined as the most lateral point on the curved insertion line of each ala that indicates the facial insertion of the nasal wing base (77).

The lack of methodological standardization for many other landmarks also made comparisons difficult (see also 3,78). Even in similar anatomic regions several measurement sites were sometimes used. For example, on the infraorbital rim, three variations existed: (i) measurements taken in the mid-plane of the orbit (e.g., 76);

(ii) measurements taken directly under the pupil (e.g., 68,69) or centered on the eye (e.g., 24); and (iii) measurements taken at the lowest point of the orbital margin, i.e., orbitale (e.g., 25). Even identically named landmarks concerned different measurements between some studies. For example, soft tissue depths at "nasion" included six variations: (i) measurements taken directly anterior to nasion or bisecting the bone curvature (2,25,66,71,76,79); (ii) measurements taken between nasion and sellion (74,80); (iii) measurements taken from nasion to a point on the soft tissue surface directly between the eyes (24); (iv) measurements taken along a plane from nasion to basion (4); (v) measurements taken from nasion to the point of skin surface in the midline and on a plane joining the upper limits of the two upper eyelid folds (68); and (vi) other variations (see e.g., 81). Ambiguously defined, but identically named landmarks, such as "beneath chin" (see e.g., 56,70,73,79,82), probably represent a similar scenario as different investigators are unlikely to measure the depth in exactly the same way. It should also be noted here that some landmarks, though named to imply skeletal sites, have actually been defined based on soft tissue features alone. See, for example, glabella, nasion, supracanine, infracanine, and supra M2 in Manhein et al. (24). Since these studies irrespectively carry weight in the literature, we continued to include them here.

Given the likely and relatively large (1–2 mm) error for repeated measurements irrespective of which measurement technique is used (2,8–34) and the clear proximity of many of these landmarks despite minor but specific differences in their location, identically named and/or proximally located measures were amalgamated. For example, the measurements on the inferior border of the orbital rim were all designated to the "mid-infraorbital" landmark and all measurements of "nasion" were considered to be comparable. It should also be mentioned that "gnathion" and "menton" provided a special problem as these terms have been used interchangeably for both identical and different landmark sites (83), and in many instances descriptions/diagrams provided by authors' were not sufficient to clarify their intentions. Thus, we were forced to speculate in some cases as to which landmark had actually been used. Another option would have been to exclude these data points but then this is not the procedure typically employed in forensic casework where practitioner interpretations must be used for deciphering ambiguously reported, though published, measurement sites. Because of space restrictions specific data amalgamations made for all landmarks cannot be presented here, however, major amalgamations are evident by comparing the data values reported in the Figures to the studies listed in Tables 1 and 2. It should be noted that the direct

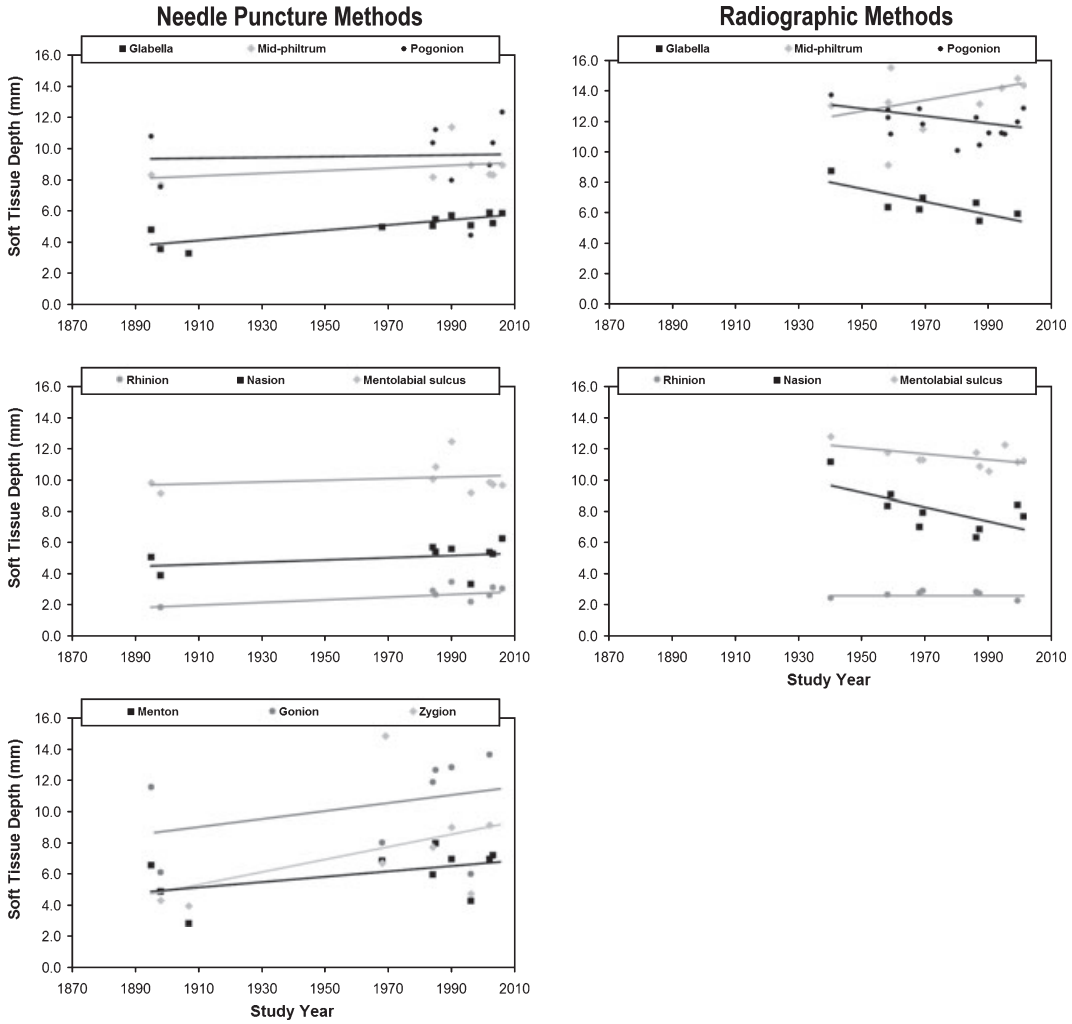


FIG. 1—*Trends of Caucasoid female soft tissue depths by publication year for needle puncture and radiographic methods. See Table 1 for studies included in this comparison.*

comparison of similarly placed, though technically different landmarks, is a procedure commonly employed in other soft tissue depth studies (for clear examples see 79).

Data were initially entered and grouped by the first author with all data entries and classifications reviewed by both authors on a second occasion. Where means and/or standard deviations were reported for right and left sides, the two values were averaged and the mean value used. The adult data were not specifically considered with respect to age because many studies used broad and varied age groups making direct comparisons impossible. It should be noted, however, that some changes have been found with increasing adult age (see for further discussion: 24,47).

Data Evaluation

The soft tissue depth variables were sequentially analyzed in a hierarchical order, starting with year of measurement, then sequentially by method of measurement, "race" and sex. Where strong trends or differences failed to exist (especially in comparison to the magnitude of known measurement errors), data were amalgamated to decrease complexity and increase sample size.

Specific Methods and Results

Study Year

As needle puncture methods on cadavers and lateral radiographs on living people date back to the 1880s, these studies were used to analyze the effects of study year. Only Caucasoid data were used for this analysis as measurements for this population group are the most extensive. To provide a more robust analysis, samples with less than five individuals were excluded. Despite having small sample sizes, the data of His (21) were retained because they are widely referenced. Tables 1 and 2 present the studies included in this part of the analysis.

Plots of commonly measured landmarks for the male and female Caucasoid data, measured using needle puncture methods, demonstrated a slight increase in all measurements over the past 50 to 100 years (see Figs. 1 and 2). Males generally increased by 1 mm across the entire time span while the increase for females was somewhat less. Zygion and gonion showed more substantial increases in comparison to other landmarks, however, variation at these landmarks was large, sometimes skewed by studies on selective samples [e.g., Sutton's study (65) which includes a large numbers of obese individuals and unsurprisingly gave very large

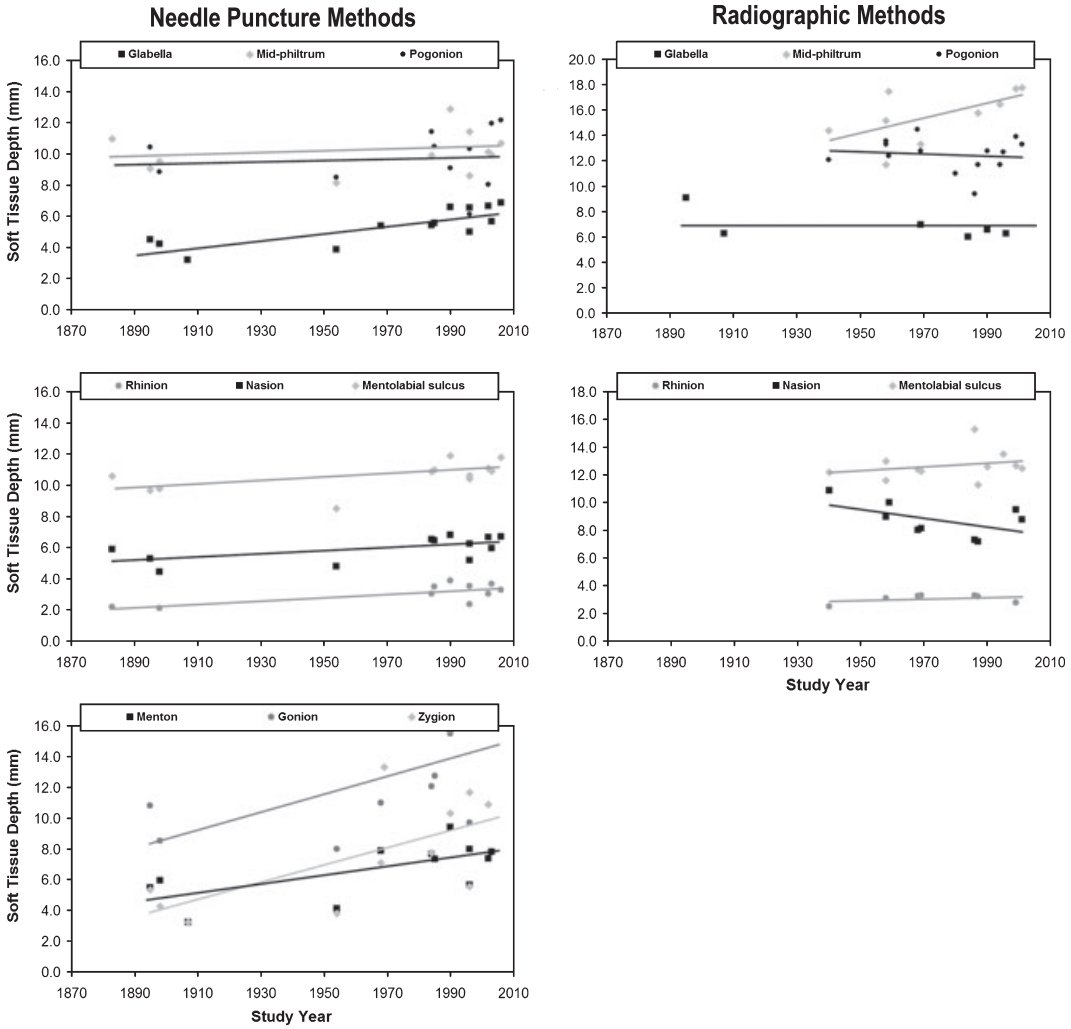


FIG. 2—*Trends of Caucasoid male soft tissue depths by publication year for needle puncture and radiographic methods. See Table 2 for studies included in this comparison.*

values], and perhaps biased by infrequent and small sampled studies prior to 1910 (see e.g., 20,22).

Males and females demonstrated identical trends by year for radiographic techniques, but these trends were not always concordant with those for needle puncture methods. For example, while both mid-philtrum and nasion showed slight increases for needle puncture techniques, radiographic methods showed a marked increase for the mid-philtrum landmark (> 2 mm), and nasion showed a marked decrease (2 mm). In addition, females displayed greater discontinuity between needle puncture and radiographic measurement, with glabella and pogonion displaying marked decreases with increasing study year for radiographic techniques. These findings are at odds with reliable changes in soft tissue depths by year since they should be in the same direction and of similar magnitude, irrespective of measurement technique used. As this is not the case, the method used to measure the soft tissue depths seems to have a greater influence on the data than any secular trend. Overall, the data that were consistent between needle puncture and radiographic techniques tended to show minimal increases with publication year. We therefore collapsed these data irrespective of publication date.

Method of Measurement

To compare data values by their method of measurement, the male and female Caucasoid data (pooled irrespective of publication year; see above) were analyzed the with respect to (i) needle puncture on cadavers, (ii) radiographs on cadavers, (iii) radiographs on living people, (iv) ultrasound on living people, and (v) MRI scans of living people. As data for each measurement method were generated from weighted means we did not exclude any studies on the basis of their sample size. The MRI Caucasoid data included a study by Sahni et al. (70) on South Asians from the Indian subcontinent as this population group is recognized to possess similar ancestry to Caucasoids (84). We also included a CT study on a mixed population because it comprised the only published data set using this measurement method on a comprehensive array of landmarks. The CT study examined a population of "Cape Colored" individuals who are thought to exhibit "a mixture of Caucasion, Negro, Khoi, and San features" (see 79, p. 52). Only landmarks that were commonly measured between studies were included in this analysis.

Plots of commonly measured landmarks sites for each measurement method revealed almost identical patterns between males and females (Fig. 3). Of particular note were the findings that the lateral radiographic method using living people consistently produced the highest data values for midline points. The only two bilateral points (gonion and zygion) measured using this method also consistently produced the lowest data values of any measurement technique. Radiographs on cadavers generally produced the second highest values, closely followed by both ultrasound on living subjects and needle puncture methods on cadavers. Ultrasound produced dramatically higher values at the supra-M2 landmark in both males and females in comparison to any other measurement method. At most landmarks, ultrasound and needle puncture methods produced very similar values. However, ultrasound tended to give slightly higher measurements overall and marked increases were observed at five landmarks: mid-infraorbital, gonion, supra-M2, infra-M2, and anterior-masseter border. Note here, however, that ultrasound methods on living persons did not yield measurements as high as radiographic techniques on cadavers. Needle puncture studies produced notably higher data values than ultrasound at mid-supraorbital and zygion landmarks, but otherwise were almost identical. CT tended to produce the lowest values of all methods, however, these data were derived from a single study (which was based on mixed individuals) so results must be viewed with some uncertainty. MRI tended to produce low values for all midline points, but for females it produced higher values for bilateral landmarks. The only landmarks that showed consistent tight clustering of values between different measurement methods were the mid-supraorbital point, glabella, and rhinion (see Fig. 3).

The wide range in mean soft tissue depth values, depending on which measurement technique was employed, highlights further the problems associated with defining facial soft tissue depths in humans. Not only does measurement error using a specific technique effect the confidence with which values can be regarded as being accurate, but the choice of the measurement method effects the magnitude, and hence the accuracy, of the values obtained as well.

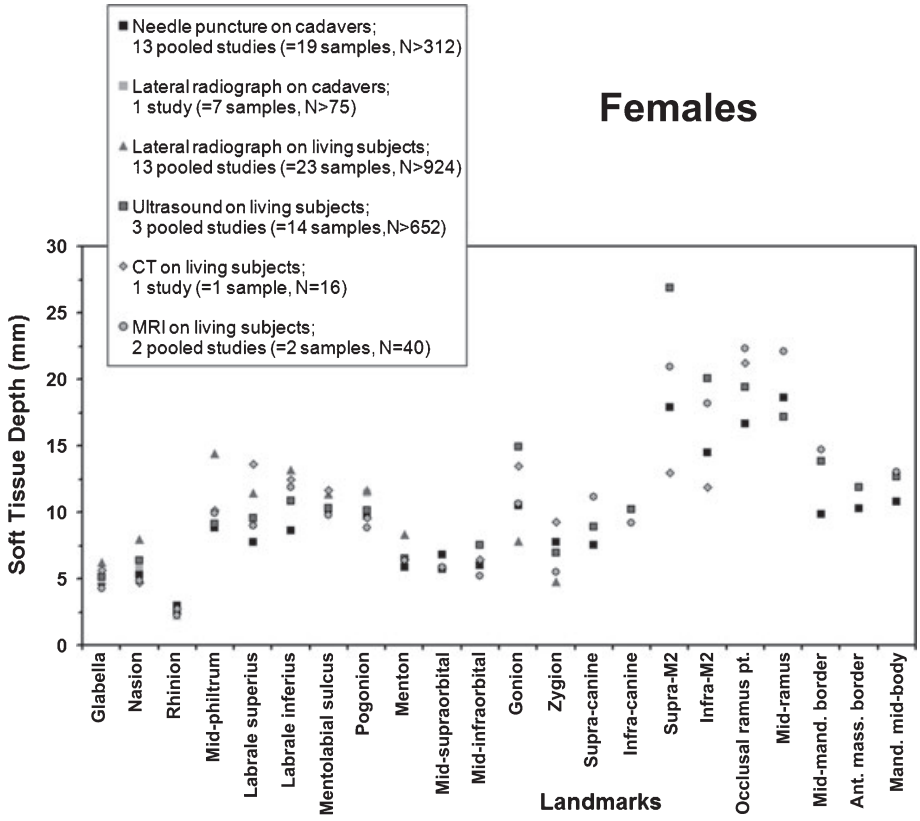
The large amount of variation evident between different measurement methods is perhaps not surprising given that other authors have found mean differences of 0.42 mm between soft tissue depth values when one head is measured on the *same* CT machine by the *same* investigator, using *means* calculated from 10 repeated measurements of landmarks *specifically chosen* to minimize measurement errors, but when different CT protocols are employed (35). There seems little alternative then to expect comparatively huge differences when totally different measurement methods are employed by independent investigators using more complete landmark sets which thereby also include more unreliable measurement points.

As each measurement method has inherent advantages and disadvantages that affect data quality in different manners (3; see also Table 3), there is no reliable way to tell which method best replicates the true soft tissue depths of humans. Consequently, the uncertainty associated with soft tissue depth measurements can only be considered to be underestimated by measurement errors for any given measurement technique. Given the spread in the data across measurement techniques, it seems that the total uncertainty associated with any soft tissue depth measurements is, at a conservative estimate, probably greater than 2 mm as first hypothesized.

The variation between soft tissue depth values measured using different techniques, the fact that each method holds its own advantages and disadvantages, and the fact that no method consistently produces midrange values makes it extremely risky to place absolute confidence in the values produced by any one method. One way forward is to pool all the data to triangulate upon real soft tissue depth values by averaging out the random errors and opposing systematic errors, so that the total cumulative error is minimized. While error neutralization cannot be guaranteed using this approach (especially when weighted means are used as errors in larger samples are given more weight) it offers the best chance to reliably approximate true soft tissue depth values. Thus the Caucasoid data were collapsed in this study across the measurement methods.

Race

To examine the affects of "race," only studies falling into the three broadly recognized racial classifications of Negroid, Mongoloid and Caucasoid were considered. It should, however, be noted here that the ability to define races within the human population is highly controversial, whether approached from a phenotypic (see 5,85,86) or a genetic standpoint (see 87–91). Since, such racial organization of data has been common place in the previous literature, and actively promoted (3,44,71,74,76,92–94), we continued these lines of investigation here despite empirical and objective tests of population group membership being rarely undertaken in soft tissue depth research.



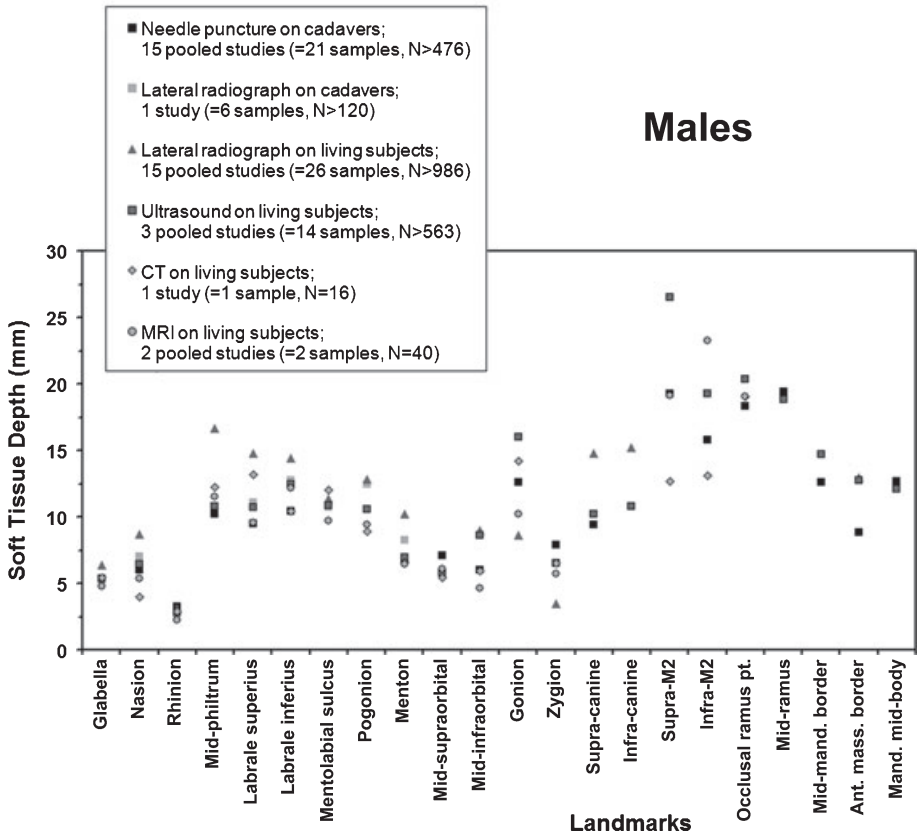


FIG. 3—*Trends in soft tissue depth data by measurement technique for adult "Caucasoid" males and females. See Tables 1 and 2 for studies included in these comparisons.*

TABLE 3—*Advantages and disadvantages of soft tissue depth measurement techniques.*[†]

Method	Advantages	Disadvantages
Needle Puncture	Inexpensive equipment required Used to measures decedents (participants do not move) Soft tissue directly measured Any site on the head can be measured No radiation output	Invasive Used to measures decedents (possible soft tissue changes with death) Participants usually in supine position (gravity effects on face) Landmark placement can be troublesome Contact method: Skin may ''dimple'' upon needle insertion and if rubber stoppers are used on needles to indicate the depth, their manipulation may cause error in soft tissue depth reading
Radiograph	Noninvasive (can measure living participants) Noncontact method Images usually taken of participants in upright position	Radiation output Soft tissue depth can only be measured in planes perpendicular to line of sight Imaging artifacts may be present, e.g., magnification affects Relatively expensive equipment is required
Ultrasound	Noninvasive (can measure living participants) Can measure participants in an upright position Can be used as a noncontact method (e.g., apply large amounts of ultrasound gel) Any site on the head can be measured Little radiation output	Often used as contact method (instrument application may cause soft tissue compression and inaccurate readings) When used as a noncontact method, participants heads have not been in upright positions (gravity effects on face) Imaging artifacts may be present, e.g., sound wave velocity may not *c*. 1540 m/sec depending on tissue structure(s) Expensive equipment is required
CT	Noninvasive (can measure living participants) Noncontact method Any site on the head can be measured	Radiation outputs are high Expensive equipment is required Imaging artifacts may be present Participants usually in supine position (gravity effects on face)
MRI	Noninvasive (can measure living participants) Noncontact method Any site on the head can be measured Little radiation output	Expensive equipment is required Imaging artifacts may be present Participants usually in supine position (gravity effects on face)

CT, computed tomography; MRI, magnetic resonance imaging.
[†]Portions of this table have been drawn from (100) with permissions from Left Coast Press.

Prior indications that measurement method strongly influences soft tissue depth values complicates the comparison of single studies derived using different measurement methods for race effects, however, some gereric value may be retained in these comparisons. Consequently, we proceeded in comparing means of non-Caucasoid studies (i.e., Negroid and Mongoloid) to the weighted means of the pooled Caucasoid data to examine their scatter. Only non-Caucasoid data with sample sizes greater than five were included in the analysis in an attempt to eliminate bias. Similarly, only Caucasoid studies with samples greater than five were used to calculate minimum and maximum values, while all Caucasoid data were used to generate the weighted total mean. Data from ''so-called'' mixed race samples were not included in this analysis.

The plots indicated that the data for any ''race'' group showed a broad degree of variation (Fig. 4). This is somewhat expected since the data includes studies using different methods, however, for more than half of the data points examined, clustering about the Caucasoid mean in both males and females was clearly observed (Fig. 4). For the other landmarks, few data points fell outside the Caucasoid range (Fig. 4). Consistently, small values were observed for Suzuki's (66) study on a Japanese sample, but no other Japanese studies were found to demonstrate values that were so small (see 44,46,48), rather the means of these other Japanese studies fell relatively close to the Caucasoid mean (see Fig. 4).

Overall, these findings suggest that ''race'' effects on soft tissue depth data are not strong since studies display broad but similar soft tissue depth ranges and central tendencies irrespective of ''race.'' Furthermore, any race differences that do exist are likely overpowered by differences between measurement methods. These observations make it unsurprising that other authors have found minimal effects when race-classified average soft tissue depth tables are incorrectly applied across ''race'' boundaries in facial approximation (95).

The lack of clear distinctions between disparate population groups is also clearly illustrated by directly comparing well-known studies to each other (see Table 4). These comparisons show that there is as much variation between different studies of the same population as there is between different studies of different population groups and in most cases these differences are also no larger than typical measurement errors (see Table 4). Given the current levels of uncertainty associated with the use of average soft tissue depths in forensic casework (i.e., intra-observer error, inter-observer error, measurement method choice, and the errors of physically placing soft tissue depths on skulls in craniofacial identification casework) it seems justifiable to regard the effects of ''race'' on soft tissue depths as minimal.

Sex

Comparison of the pooled data by sex revealed that males had slightly larger mean values than females at almost all landmarks except in the region of the cheeks (Fig. 5). Overall, sex differences were extremely small. The median difference was 0.4 mm, with a range of 0.0 mm to 1.9 mm, which again is small in comparison to the magnitude of total uncertainty discussed above. The differences between the sexes (where they exist) are highly statistically significant ($p < 0.001$) when the data for which standard deviations exist are compared using *t*-tests; but even so the practical meaning of this difference is limited given the much larger measurement errors that exist. The data were, therefore, collapsed across the sexes using weighted means to yield one set of soft tissue depth data for all adult individuals. The approximated reference landmarks for these pooled data are given in Table 5, while the data values are presented in Table 6. Only easily determinable and commonly measured landmarks have been included. Data values have been rounded to the nearest 0.5 mm so as not to overemphasize

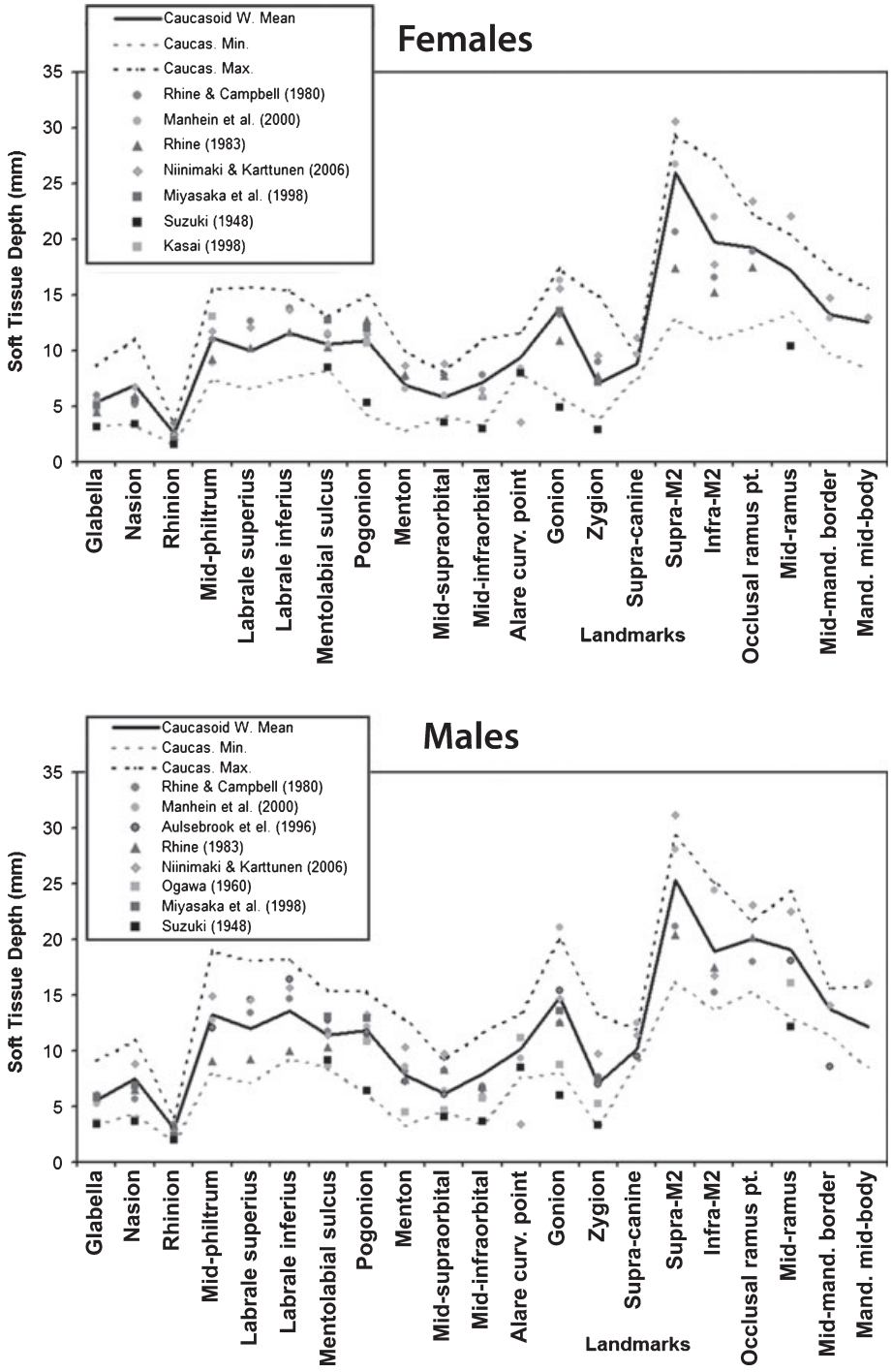


FIG. 4—*Comparison of non-Caucasoid soft tissue depth data to the Caucasoid weighted mean (and ranges) for males and females. Circles represent data on Black Americans (Negroid). The black outlined circle represents data on Zulus (Negroid). Triangles represent data on American Indians (Mongoloid). Diamonds represent data on Finns (Mongoloid). Squares represent data on Japanese (Mongoloid). Solid and dashed lines are used to more clearly illustrate Caucasoid data ranges and the mean; they should not be interpreted to indicate any relationships between the variables along the x-axis. See Tables 1 and 2 for studies included in these comparisons.*

accuracy as recommended by Aulsebrook et al. (68), however, we note that this is not an overly conservative measure given that uncertainty levels can be measured in whole millimeters (see e.g., 2,28–34). Despite this, we continue to use a 0.5-mm interval here as an improved precision indicator in contrast to the two decimal places frequently reported in the literature. Figure 6 illustrates the location of the hard tissue landmarks with directions of measurement commonly encountered.

A subset of data for studies that report standard deviations along with their means are also presented in Table 6 and are very similar to the total pooled means. As some of the data ranges calculated from these means and standard deviations are negative (see mid-infraorbital, gonion, and zygion landmarks), it is clear that their respective data distributions are not normal. This jeopardizes the value of the calculated means and supports claims made by other studies (29,96) that alternate central

TABLE 4—*Comparison of well-known race specific means.*

	Caucasoid Data			Black American	American Indian	Japanese
	De Greef et al. (94)	Helmer (25)	Simpson & Henneberg (2)	Rhine & Campbell (74)	Rhine (55 Cited in 57)	Miyasaka (44)
n	65	11	13	52	18	56
Nasion	7.0	7.3	6.7	5.6	6.5	6.9
Mentolabial sulcus	11.0	13.0	11.1	11.8	10.3	13.1
Pogonion	11.5	13.7	8.0	11.5	11.3	12.9
Menton	7.3	9.8	7.4	8.2	7.4	—
Zygion	7.5	5.5	10.9	7.7	7.5	7.5
Gonion	17.0	11.7	18.5	12.5	12.6	13.6

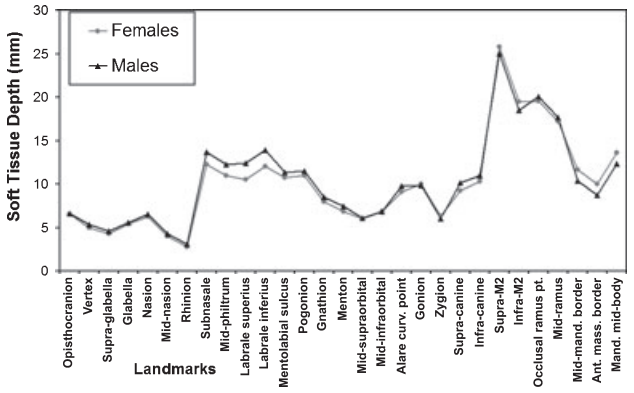


FIG. 5—*Sex comparison of the published soft tissue depth data. Solid lines are used to more clearly illustrate the data, however, they should not be interpreted to indicate any relationship between the variables along the x axis. See Tables 1 and 2 for studies included in these comparisons.*

tendency indicators (i.e., medians and modes) should be given greater attention.

Discussion

The findings reported here collectively suggest that the subcategorization of the soft tissue depth data according to variables such as publication year, method of measurement, race, and sex is of little practical benefit given the totality of data uncertainty that exists. Given the small magnitudes of difference between the groups and the relatively larger total sources of error, it seems reasonable to speculate that even when differences for each variable are simultaneously considered in a multivariate analysis, interpretations are unlikely to be vastly different. Some may argue that pooling data across variables has increased the variance of the data set which clouds any differences between the groups. We acknowledge that the procedures used here may increase the variance, but we also highlight the fact that even before data amalgamation, variance is high within the groups and differences between the groups small. It has previously been shown, for example, that sex distinction (the last variable to be assessed in this study after prior data amalgamations) is unjustified when a single study sample from a single population group, measured using one measurement method, is considered (96). The validity of the pooled data presented here is confirmed by their high similarity to the values reported by the many single soft tissue depth studies from which they were derived. Simultaneous consideration of all variables for their total additive influence should, however, be undertaken using multivariate methods in the near future.

A principal components analysis would be useful to provide more definitive clarifications, but unfortunately it is not presently possible because: (i) raw data are infrequently published; (ii) raw data published more than 5 years ago are frequently disposed of; and (iii) raw data from published studies are not readily shared by some authors. To avoid this situation in the future, and to enable larger data sets to be accumulated, an online raw data store has been established. Investigators can now voluntarily donate their raw data to this store and/or access the full database for their own research purposes (see http://www.craniofacialidentification.com). Submitted data are tagged according to whether or not they have formed part of a prior professionally peer-reviewed work, thus enabling investigators to gauge (to some degree) the reliability of the data they can access. The centralized collection of the raw data will also enable adult soft tissue depths to be more specifically analyzed with respect to age in the future.

While the pooled data means reported here hold many advantages in contrast to independent study means reported in the literature, they suffer from five main limitations: (i) they are biased toward people of "normal" weight; (ii) they are derived from means despite raw data being skewed; (iii) associated measurement errors are relatively large; (iv) data are not collected using a basic standard set of landmarks; and (v) variables widely known to affect skin fold measurements have not been studied or controlled in soft tissue depth research (i.e., the effects of air temperature, hydration status, stage of menstrual cycle, and pregnancy (97)). These limitations justify the collection of additional soft tissue depth values in the future under the provisions that: (i) individuals should not be excluded according to their body build; (ii) complete descriptive statistics should be reported not just arithmetic means; (iii) measurement errors should be recorded and minimized; (iv) a minimal set of standardized landmarks should be investigated; and (v) raw data are stored and made available for future larger-sampled analyses. By reviewing the published data, this study has identified 25 landmarks which are commonly measured or, at least, closely approximated by most authors (see Fig. 6 and Tables 5 and 6). As such these landmarks provide a good basis for a minimum set of standardized landmarks in future investigations.

Certainly, future efforts to elucidate the effects of air temperature, hydration status, stage of menstrual cycle, and pregnancy will be worth while along with studies that employ multiple methods of soft tissue measurement on common samples. Only two attempts have so far been made to address this latter aspect but sample sizes have been small and few measurement methods examined (see 30,35). Independent studies that investigate yet another population group without definitive (social or biological) criteria for group membership and/or do not give explicit reasoned justification for the research (especially when other similar population groups have already been studied) should be avoided.

TABLE 5—*Soft and hard tissue landmarks approximated by the pooled data.*[†]

Skeletal Landmarks Median Points	Definitions	Soft Tissue Landmarks Median Points	Definitions
Opisthocranion (op)	Midline ectocranial point at farthest chord length from glabella	Opisthocranion (op′)	Midline soft tissue point directly overlying hard tissue opisthocranion (op)
Vertex (v)	Highest midline ectocranial point	Vertex (v′)	Midline soft tissue point directly overlying hard tissue vertex (v)
Glabella (g)	Most anterior midline point on the frontal bone	Glabella (g′)	Most anterior midline soft tissue point overlying the glabella (g)
Nasion (n)	Midline point on the naso-frontal suture	Nasion (n′)	Midline soft tissue point directly overlying hard tissue nasion (n) and superior to sellion
Mid-nasal (mn)	Point on internasal suture midway between nasion and rhinion	Mid-nasal (mn′)	Midline soft tissue point directly overlying the hard tissue mid-nasal point (mn)
Rhinion (rhi)	Midline point at the inferior free end of the internasal suture	Rhinion (rhi′)	Midline soft tissue point directly above the hard tissue rhinion (rhi)
Subnasale (sn)	Midline point just below the anterior nasal spine	Subnasale (sn′)	Midline point of the angle at the comulella base where the septum and upper lip join
Mid-philtrum (mp)	Midline point midway between the base of the nasal spine and prosthion (see below) on the anterior edge of the maxillae	Mid-philtrum (mp′)	Midline point midway between soft tissue subnasale and the vermilion border of the upper lip in the groove of the philtrum
Labrale superius (ls)	Midline landmark at the most anterior edge of the superior alveolar ridge of maxillae (or prosthion)	Labrale superius (ls′)	Midline soft tissue point at the vermilion border of upper lip
Labrale inferius (li)	Midline point at the most anterior edge on the inferior alveolar ridge of the maxillae	Labrale inferius (li′)	Midline soft tissue point at the vermilion border of lower lip
Mentolabial sulcus (mls)	Deepest midline point in the groove superior to the mental eminence	Mentolabial sulcus (mls′)	Deepest soft tissue point at the midline of the groove just superior to the chin
Pogonion (pg)	Most anterior midline point on the mental eminence of the mandible	Pogonion (pg′)	Most anterior midline point on the eminence of the soft tissue chin
Gnathion (gn)	Midline point halfway between the most anterior (pg) and inferior (m) points on the bony chin	Gnathion (gn′)	Midline soft tissue point directly overlying the hard tissue gnathion (gn)
Menton (m)	Most inferior midline point at the mental symphysis of the mandible	Menton (m′)	Midline soft tissue point directly overlying the hard tissue menton (m)
Bilateral Points		Bilateral Points	
Mid-supraorbital (mso)	Point on the supraorbital rim at the mid-sagittal plane of the orbit	Mid-supraorbital (mso′)	Soft tissue point anteriorly overlying the hard tissue mid-supraorbital point (mso)
Mid-infraorbital (mio)	Point on the infraorbital rim at the mid-sagittal plane of the orbit	Mid-infraorbital (mio′)	Soft tissue point anteriorly overlying the hard tissue mid-infraorbital point (mio)
Alare curvature point (acp)	Point *c.* 3 mm lateral to the border of the nasal aperture	Alare curvature point (acp′)	Soft tissue point indicating the most lateral insertion of the alare base into the face
Gonion (go)	Point on the lateral aspect of the border of mandiblular angle where a tangent bisects the angle formed by the posterior ramus border and the inferior corpus border	Gonion (go′)	Soft tissue point directly overlying the hard tissue gonion (go)
Zygion (zy)	Most lateral extent of the lateral surface of the zygomatic arch	Zygion (zy′)	Soft tissue point directly overlying the hard tissue zygion (zy)
Supra canine (sC)	Point on superior alveolar ridge superior to the crown of the maxillary canine(s)	Supra canine (sC′)	Soft tissue point directly overlying the hard tissue supra canine point (sC)
Infra canine (iC)	Point on inferior alveolar ridge inferior to the crown of the mandibular canine(s)	Infra canine (iC′)	Soft tissue point directly overlying the hard tissue infra canine point (iC)
Supra M^2 (sM^2)	Point on superior alveolar ridge superior to the crown of the maxillary second molar(s)	Supra M^2 ($sM^{2\prime}$)	Soft tissue point directly overlying the hard tissue supra M^2 point (sM^2)
Infra M_2 (iM_2)	Point on inferior alveolar ridge inferior to the crown of the mandibular second molar(s)	Infra M_2 (iM_2')	Soft tissue point directly overlying the hard tissue infra M_2 point (iM_2)
Mid-ramus (mr)	Point at the center of the mandibular ramus	Mid-ramus (mr′)	Soft tissue point directly overlying the hard tissue mid-ramus point (mr)
Mid-mandibular border (mmb)	Point on the inferior border of the corpus of the mandible midway between pogonion and gonion	Mid-mandible border (mmb′)	Soft tissue point directly overlying the hard tissue mid-mandibular border point (mmb)

[†]Landmarks are positioned assuming that the skull is in the Frankfurt horizontal.

TABLE 6—*Pooled soft tissue depth data for adults (rounded to the nearest 0.5 mm).*[†]

Soft Tissue Depth Measurement	Total Weighted Mean	*n*	No. of Samples	Weighted Mean for *s* Studies	*s*	*n*	No. of Samples	Estimated Minimum (Mean − 3 z-scores)	Estimated Maximum (Mean + 3 z-scores)
Median points									
op-op′	6.5	1152	52	6.5	2.5	990	36	−0.5	13.5
v-v′	5.0	1055	43	5.0	1.0	785	29	1.5	8.5
g-g′	5.5	5791	163	5.5	1.0	4542	115	2.5	8.5
n-n′	6.5	6159	154	6.0	1.5	4417	103	1.0	11.0
mn-mn′	4.0	1272	67	4.0	1.0	919	38	0.5	8.0
rhi-rhi′	3.0	5511	146	3.0	1.0	4307	100	0.0	5.5
sn-sn′	13.0	1768	78	12.5	3.0	1170	43	3.0	22.5
mp-mp′	11.5	5508	116	11.0	2.5	3955	74	3.0	18.5
ls-ls′	11.5	5106	133	11.5	3.0	4216	97	3.0	20.0
li-li′	13.0	4886	110	13.0	2.5	4017	77	5.0	21.0
mls-mls′	11.0	5792	158	11.0	2.0	4497	106	5.5	16.5
pg-pg′	11.5	6786	168	11.0	2.5	4891	105	3.5	18.5
gn-gn′	8.5	545	18	8.5	3.0	381	10	−1.0	18.0
m-m′	7.0	4475	143	7.0	2.5	3795	104	0.0	14.0
Bilateral points									
mso-mso′	6.0	2225	78	6.0	1.5	1838	49	1.5	10.5
mio-mio′	7.0	2298	91	7.0	3.5	1910	61	−4.0	18.0
acp-acp′	9.5	1511	43	9.3	2.0	1361	31	2.5	16.0
go-go′	10.0	4168	113	10.0	6.0	3320	77	−8.0	27.5
zy-zy′	6.0	4390	103	6.0	1.0	3545	68	3.0	9.0
sC-sC′	9.5	3138	50	9.5	2.0	3113	48	3.5	15.5
iC-iC′	10.5	1184	27	10.5	2.0	1157	25	4.5	16.5
sM^2-$sM^{2\prime}$	25.5	1405	41	26.0	5.5	1212	33	10.0	42.0
iM_2-$iM_2{}'$	19.0	1344	39	19.5	4.5	1151	31	6.0	33.0
mr-mr′	17.5	2858	60	17.5	4.0	2637	37	6.0	28.5
mmb-mmb′	10.5	935	26	10.5	4.5	548	21	−2.5	24.0

[†]Landmark definitions are given in Table 5 and illustrated in Fig. 6. The ''number of samples'' represents the number of subcategorized groups previously used for averaging and, therefore, may include several samples from any one study. A subset of studies reporting standard deviations were used to generate the statistics for the ''Weighted mean for s studies.''

One of the most significant findings of this study was that measurements on decedents may be as large or larger than those on living individuals (or vice versa) depending on which method of measurement is used and which landmarks are studied. Consequently, this paper demonstrates that any systematic biases arising from the use of cadavers (if they exist) are not strong enough to overpower biases because of other variables like measurement technique. Therefore, generalized claims that cadaver-based soft tissue depth measurements are inferior in quality to living individuals is at this stage empirically unfounded. This finding does not mean, however, that the faces of cadavers are identical to those of living individuals. Rather the data suggest that soft tissue depth measurements are not very sensitive to the changes associated with death because the errors and biases inherent to present methods make any difference impossible to detect. Until uncertainty levels and measurement errors have been reduced in current soft tissue depth measurement methods, changes from the living to the dead state are likely to be better addressed by more targeted analysis of facial shape contours than by soft tissue depth measurements at sparsely located anatomic landmarks on the face.

The comparison of the data across multiple measurement methods in this study also provided some insights into the systematic errors (or biases) inherent to each method. For example, CT and MRI of living persons most commonly positioned in the supine position tended to produce low values in the midline but high values for bilateral points probably as a result of soft tissue movement under the effects of gravity. This conclusion is similar to that reached by other investigators (see e.g., 31,93,98), and it is worth noting that such trends were not observed by Kim et al. (35) in a repeatability study that compared CT and caliper measured cadaver heads. Notably, the cadaver heads in this study were sectioned from the rest of the body so they could be placed upright in the CT scanner (35).

Higher mid-plane values for radiographic methods in comparison to ultrasound methods are also another observation that has been reported by other researchers (see 30). Possible causes of these differences may be magnification differences between the views, and inability to account for three dimensional surface typology in radiographs thus generating mistaken landmark points (30).

In addition to the broader sweeping ramifications, the findings of this study also offer useful insights into specific soft tissue depth studies. Suzuki's data (66), for example, are consistently shown to fall outside the 99.7% confidence interval of other Japanese studies (see 44,46,48). In contrast to radiographic studies, needle puncture techniques are expected to produce lower values, but Suzuki's are so low that measurement methods alone seem unable to reconcile the differences observed. Suzuki's value for pogonion in females falls, for example, 3.9 mm below Kasai's (48) mean. This soft tissue depth difference is double that expected to be obtained due to the use of different measurement methods (see fig. 3) suggesting that Suzuki's data include other biases. Small sample size is one conspicuous feature of Suzuki's data (Kasai's sample is more than 42 times as large as Suzuki's and produces much larger values). These observations hold significant ramifications for ''race'' distinction based on soft tissue depths as Suzuki's data have been heavily relied upon in the literature as an exemplar of the differences that exist between different ''races'' especially in comparison to Caucasoids (see e.g., 3, 27, 74, 79).

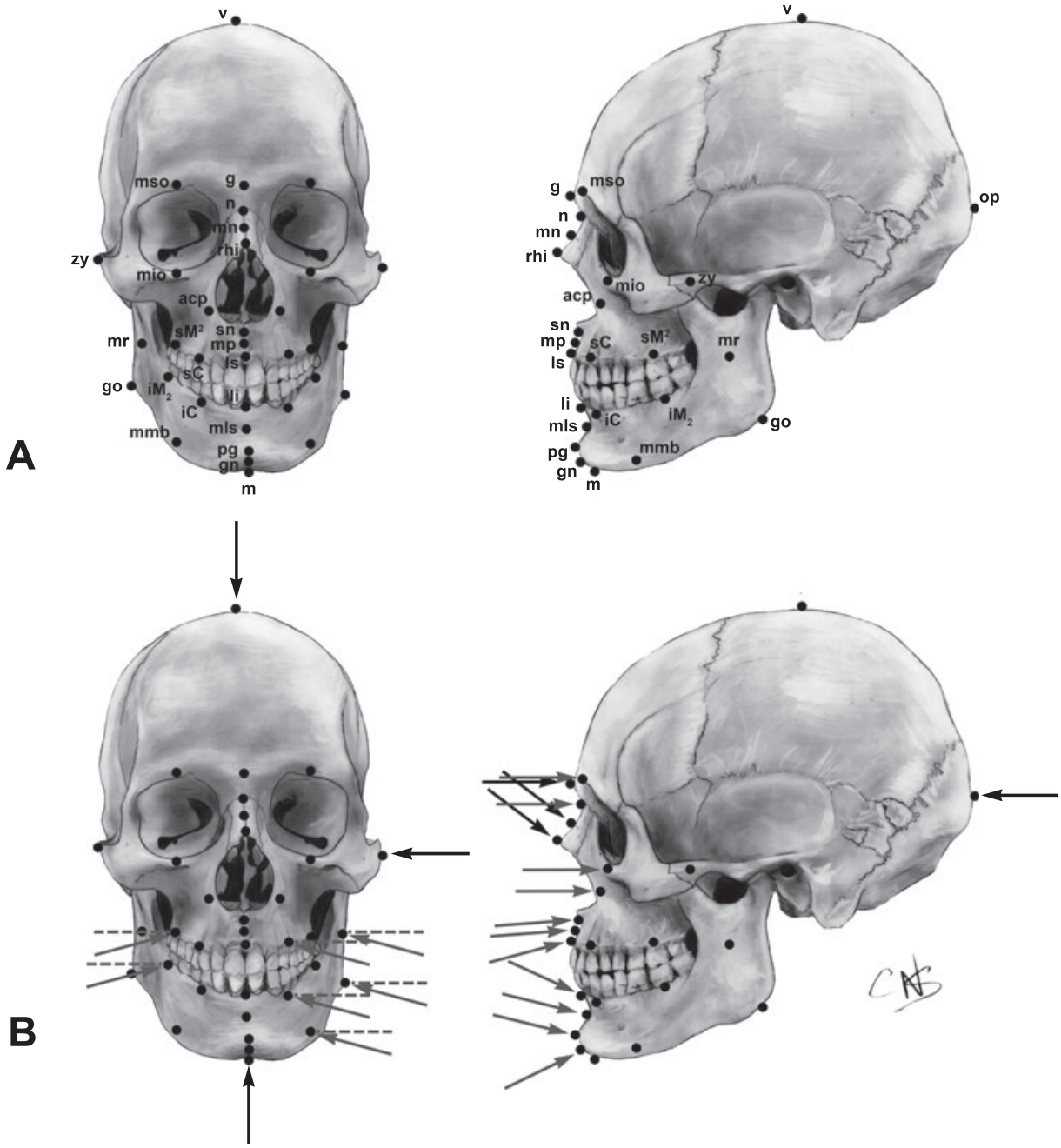


FIG. 6—*Soft tissue depth measurement sites approximated by the pooled data. Top row (a) illustrates the hard tissue landmarks (see Table 5 for definitions). Bottom row (b) illustrates the typical directions in which the soft tissue depth measurements at these landmarks are taken. Black arrows indicate measurements at angles that bisect the curvature of the bone surface (typically, though inaccurately, known as "perpendicular" measurements). Gray arrows indicate measurement directions that do not bisect the curvature of the bone surface, but are intentionally orientated in other directions (often horizontally or towards other soft tissue features, see Table 5). Arrows with accompanying dashed lines indicate horizontal orientation of the soft tissue depth measurement, but with anterior inclination.*

Streamlining and standardizing of facial soft tissue depth measurement methods according to optimal protocols will undoubtedly be beneficial in the future. Optimal methods concern those in which upright living individuals can be measured and measurement errors reduced. While this study showed that use of cadavers does not have a major effect on soft tissue depths (choice of measurement method is more important), minor changes probably still occur and need to be considered as measurement errors are decreased and interobserver repeatability improved. A promising way to achieve such improvements would be to limit the amount of variation that occurs as a result of compression of facial soft tissue because of subject position and use of equipment.

If all other factors are considered to be equal, high resolution upright MRI should be given precedence as it is a 3D, noncontact method that enables the skull and face to be visualized on living subjects with low radiation levels, and gives still-images that can be carefully analyzed and measured. However, given the expense, ultrasound appears to be the most suitable compromise (28). Ultrasound techniques should also be used on living upright

subjects and with high frequencies and resolution conducers. B-mode ultrasound should be employed as it allows the visualization of the soft tissue and bony profiles and, like MRI, enables onscreen measurement of still-frame images. "Standoff" ultrasonic gel platforms should be used to help decrease soft tissue compression and avoid inflated measurement errors. Such platforms can be easily constructed by applying a large amount of ultrasonic gel to the face above the landmark of interest and placing the conducer in the gel, but not touching the face. The ultrasonic gel has low acoustic impedance and attenuation values so it has little effect on sound wave transmission to and from the subject even if large amounts are applied. It is also readily deformable so any risk of soft tissue depression is dramatically reduced. This method also circumvents the problem of placing subjects into the prone position during data acquisition, as is required for other noncontact ultrasound techniques (see 28,99).

Conclusions

This study demonstrates that the multitude of facial soft tissue thickness data available in the literature for adults can be assimilated to provide a more statistically powerful, yet simplified data set. The low standard errors of the mean of the pooled data augur well for their high interpretative and practical weight as average soft tissue depth standards. The study also demonstrates that no method of soft tissue thickness measurement can be considered superior to any other, and that those studies using small samples should not be assigned heavy weight because of possible biases [Suzuki's data (66) exemplify such a scenario]. The landmarks used in this study for calculating the pooled data provide a solid basis for a minimum set of standardized landmarks that should be used in future soft tissue depth research. This investigation also demonstrates that new studies will maximize their value by seeking to overcome limitations inherent in the current data (i.e., problems associated with standardization, measurement error, and application error are of primary concern) and highlights the need for raw data to be stored so that in future, more comprehensive analyses can be conducted. A new online raw data store that can be readily accessed and contributed to by researchers (see http://www.craniofacialidentification.com) now makes this possible. An annually updated version of the pooled data values reported in this paper will also be available from this website in the future.

Acknowledgments

The authors thank Ronn Taylor for informal discussions that helped inspire this research; Roger Gent, Department of Radiography, The Women's and Children's Hospital, Adelaide, for discussion on optimizing facial soft tissue measurements in ultrasonography; Maciej Henneberg for communicating previously unpublished data (Henneberg and Anderson, 1996) and helping with weighted standard deviation formulae; Danielle Funke and Kari Anderson for technical assistance; Oliver Ningelgen, Hideki Saka, and Misako Brown for helping with translations; and to members of the Biological Anthropology and Comparative Anatomy Research Unit, The Department of Anatomical Sciences, The University of Adelaide, for helpful discussion. Special thanks are also extended to G. Hodson, L.S. Lieberman, P. Wright; M.A. Williamson, S.P. Nawrocki, T.A. Rathbun; M. Domaracki; and M. Henneberg who gave access to raw data to enable attempts at multivariate analysis.

References

1. Brues AM. Identification of skeletal remains. J Crim Law Criminol Police Sci 1958;48:551–6.
2. Simpson E, Henneberg M. Variation in soft-tissue thicknesses on the human face and their relation to craniometric dimensions. Am J Phys Anthropol 2002;118:121–33.
3. Niinimaki S, Karttunen A. Finnish facial tissue thickness study. In: Herva V-P, editor. People, material culture and environment in the north. Proceedings of the 22nd Nordic Archaeological Conference; Aug 18–23, 2004. Gummerus Kirjapaino Oy: University of Oulu, 2006;343–52.
4. Subtelny JD. A longitudinal study of soft tissue facial structures and their profile characteristics, defined in relation to underlying skeletal structures. Am J Orthod 1959;45:481–507.
5. Montagu A. Man's most dangerous myth: the fallacy of race. New York: Columbia University Press, 1942.
6. Brace CL. Region does not mean 'race'—reality versus convention in forensic anthropology. J Forensic Sci 1995;40:171–5.
7. Bhopal R, Donaldson L. White, European, Western, caucasion, or what? Inappropriate labeling in research on race, ethnicity and health. Am J Public Health 1998;88:1303–7.
8. Boehmer U, Kressin NR, Berlowitz DR, Christiansen CL, Kazis LE, Jones JA. Self-reported vs administrative race/ethnicity data and study results. Am J Public Health 2002;92:1471–3.
9. Editorial. Race, ethnicity, culture, and science. Br Med J 1994;309:286–7.
10. De Bono D. Describing race, ethnicity, and culture in medical research. Br Med J 1996;313:425.
11. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. J Am Med Assoc 2003;289:2709–16.
12. Stolley PD. Race in epidemiology. Int J Health Serv 1999;29:905–9.
13. Edwards GF. Discussion: race and demographic analysis. Soc Biol 1971;18:394–6.
14. Bogue G. How to get along without race in demographic analysis. Soc Biol 1971;18:387–93.
15. Osborne NG, Feit MD. The use of race in medical research. J Am Med Assoc 1992;267:275–9.
16. Fullilove MT. Comment: abandoning ''race'' as a variable in public health research—an idea whose time has come. Am J Public Health 1998;88:1297–8.
17. Editorial. Ethnicity, race, and culture: guidelines for research, audit, and publication. Br Med J 1996;312:1094.
18. Hahn RA, Stroup DF. Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. Public Health Rep 1994;109:7–15.
19. Rathore SS. Race, ethnic group, and clinical research. Br J Med 2003;327:763–4.
20. Welcker H. Schiller's Schadel und Todtenmaske, nebst Mittheilungen uber Schadel und Todtenmaske Kant's. Braunschweig: Viehweg F and Son, 1883.
21. His W. Anatomische Forschungen uber Johann Sebastian Bach's Gebeine und Antlitz nebst Bemerkungen uber dessen Bilder. Abh MathPhysikal KI Kgl Sachs Ges Wiss 1895;22:379–420.
22. Kollman J, Büchly W. Die Persistenz der Rassen und die Reconstruction der Physiognomie prahistorischer Schadel. Archiv für Anthropologie 1898;25:329–59.
23. Anderson W. The correlation between soft tissue thickness and bony proportions of the skull and how they relate to facial reconstruction. Adelaide: Department of Anatomical Sciences, The University of Adelaide, 1996.
24. Manhein MH, Listi GA, Barsley RE, Musselman R, Barrow NE, Ubelaker DH. *In vivo* facial tissue depth measurements for children and adults. J Forensic Sci 2000;45:48–60.
25. Helmer R. Schadelidentifizierung durch elekronicshe Bildmischung: Zugleich ein Beitrag zur Konstitutionsbiometrie und Dickenmessung der Gesichtsweichteile. Heidelberg: Krminalistik-Verlag, 1984.
26. Forrest AS. An investigation into the relationship between facial soft tissue thickness and age in australian caucasion cadavers [dissertation]. Brisbane: Department of Anatomy, The University of Queensland, 1985.
27. Wilkinson C. Forensic facial reconstruction. Cambridge: Cambridge University Press, 2004.
28. Smith SL, Throckmorton GS. A new technique for three-dimensional ultrasound scanning of facial tissues. J Forensic Sci 2004;49:1–7.
29. Domaracki M, Stephan CN. Facial soft tissue thicknesses in Australian adult cadavers. J Forensic Sci 2006;51:5–10.

30. Smith SL, Throckmorton GS. Comparability of radiographic and 3D-ultrasound measurements of facial midline tissue depths. J Forensic Sci 2006;51:244–7.
31. De Greef S, Claes P, Mollemans W, Loubele M, Vandermeulen D, Suetens P, et al. Semi-automated ultrasound facial soft tissue depth registration: method and validation. J Forensic Sci 2005;50:1282–8.
32. Vander Pluym J, Shan WW, Taher Z, Beaulieu C, Plewes C, Peterson AE, et al. Use of magnetic resonance imaging to measure facial soft tissue depth. Cleft Palate Craniofac J 2007;44:52–7.
33. Shaner DJ, Bamforth JS, Peterson AE, Beattie OB. Different techniques, different results—a comparison of photogrammetric and caliper-derived measurements. Am J Phys Anthropol 1998;106:547–52.
34. Suk V. Fallacies of anthropological identifications. Publications de la Facultae des sciences de l'Universitae Masaryk 1935;207:3–18.
35. Kim K-D, Ruprecht A, Wang G, Lee JB, Dawson DV, Vannier MW. Accuracy of facial soft tissue thickness measurements in personal computer-based multiplanar reconstructed computed tomographic images. Forensic Sci Int 2005;155:28–34.
36. Stephan CN, Henneberg M, Sampson W. Predicting nose projection and pronasale position in facial approximation: a test of published methods and proposal of new guidelines. Am J Phys Anthropol 2003;122:240–50.
37. Smith SL, Buschang PH. Midsagittal facial soft-tissue growth of French Canadian adolescents. Am J Hum Biol 2002;14:457–67.
38. Helwin H. Die identifizierung des paulssen-schadels. Biologishe Rundschau 1969;7:119–25.
39. Smith SL, Buschang PH. Midsagittal facial tissue thickness of children and adolescents from the Montreal growth study. J Forensic Sci 2001;46:1294–302.
40. Helwin H. Die profilanalyse, eine moglichkeit der identifizierung unbekannter schadel. Gegenbaurs Morphol Jahrb 1969;113:467–99.
41. Altemus LA. Comparative integumental relationships. Angle Orthod 1963;33:217–21.
42. Burstone CJ. Integumental contour and extension patterns. Angle Orthod 1959;29:93–104.
43. Blythe T. A re-assessment of the Rhine and Moore technique in forensic facial reconstruction. Manchester: Anatomical Sciences, The University of Manchester, 1996.
44. Miyasaka S. Progress in facial reconstruction technology. Forensic Sci Rev 1999;11:50–90.
45. Birkner F. Beitrage zur rassenanatomie der gesichtsweichteile. Corr Bl Anthrop Ges Jhg 1904;34:163–5.
46. Ogawa H. Anatomical study on the Japanese head by x-ray cephalometry. J Tokyo Dental Coll Soc [Shika Gakuho] 1960;60:17–34.
47. Michelow BJ, Guyuron B. The chin: skeletal and soft-tissue components. Plast Reconstr Surg 1995;95:473–8.
48. Kasai K. Soft tissue adaptability to hard tissues in facial profile. Am J Dentofac Orthop 1998;113:674–84.
49. Nanda RS, Ghosh J. Facial soft tissue harmony and growth in orthodontic treatment. Semin Orthod 1995;1:67–81.
50. Genecov JS, Sinclair PM, Dechow PC. Development of the nose and soft tissue profile. Angle Orthod 1990;60:191–8.
51. Formby WA, Nanda RS, Currier GF. Longitudinal changes in the adult facial profile. Am J Orthod Dentofacial Orthop 1994;105:464–76.
52. Martin R. Lehrbuch der anthropologie. Stuttgart: Gustav Fischer Verlag, 1957.
53. Bankowski IM. Die Bedeutung der Unterkieferform und-stellung fur die phtographische Schadelidentifizierung [dissertation]. Frankfurt: 1958.
54. Weining W. Rontgenologische Untersuchungen zur Bestimmung der WeichteildickenmaBe des Gesichts [dissertation]. Frankfurt: Johann Wolfgang Goethe-Universität, 1958.
55. Rhine S. Tissue thickness for southwestern Indians [dissertation]. New Mexico: Physical Anthropology Laboratories, University of New Mexico, 1983.
56. Rhine JS, Moore CE. Tables of facial tissue thickness of American caucasoids in forensic anthropology. Maxwell Museum Tech Ser 1984;1.
57. Taylor KT. Forensic art and illustration. Boca Raton: CRC Press, 2001.
58. Krogman WM, Iscan MY. The human skeleton in forensic medicine. Illinois: Charles C Thomas, 1986.
59. Taylor RG, Angel C. Facial reconstruction and approximation. In: Clement JG, Ranson DL, editors. Craniofacial identification in forensic medicine. New York: Oxford University Press, 1998;177–85.
60. Czekanowski J. Untersuchungen uber das Verhaltnis der Kopfmafse zu den Schadelmafsen. Archiv für Anthropologie 1907;6:42–89.
61. Hulanicka B, Kotlarz K. The final phase of growth in height. Ann Hum Biol 1983;10:429–34.
62. Roche AF, Davila GH. Late adolescent growth in stature. Pediatrics 1972;50:874–80.
63. Stadtmuller F. Zur Beurteilung der plastischen Rekonstruktionsmethode der Physiognomie auf dem Schadel. Z Morphol Anthropol 1922;22:337–72.
64. Stephan CN, Simpson EK. Facial soft tissue depths in craniofacial identification (part II): an analytical review of the published sub-adult data. J Forensic Sci DOI 10.1111/j.1556-4029.2008.00853.x.
65. Sutton PRN. Bizygomatic diameter: the thickness of the soft tissues over the zygions. Am J Phys Anthropol 1969;30:303–10.
66. Suzuki H. On the thickness of the soft parts of the Japanese face. J Anthropol Soc Nippon 1948;60:7–11.
67. Hodson G, Lieberman LS, Wright P. *In vivo* measurements of facial tissue thicknesses in American Caucasoid children. J Forensic Sci 1985;30:1100–12.
68. Aulsebrook WA, Becker PJ, Iscan MY. Facial soft-tissue thickness in the adult male Zulu. Forensic Sci Int 1996;79:83–102.
69. El-Mehallawi IH, Soliman EM. Ultrasonic assessment of facial soft tissue thickness in adult Egyptians. Forensic Sci Int 2001;117:99–107.
70. Sahni D, Jit I, Gupta M, Singh P, Suri S. Preliminary study on facial soft tissue thickness by magnetic resonance imaging in northwest Indians. Forensic Sci Commun 2002;4:1. Available at: http://www.fbi.gov/hq/lab/fsc/backissu/jan2002/sahni.htm. Accessed July 30, 2008.
71. Williamson MA, Nawrocki SP, Rathbun TA. Variation in midfacial tissue thickness of African-American children. J Forensic Sci 2002;47:25–31.
72. Eggeling Hv. Anatomische untersuchungen an den Kopfen con cier Hereros, einem Herero- und einem Hottentottenkind. In: Schultze, editor. Forschungsreise im westrichen und zentraien Sudafrika. Jena: Denkschriften, 1909;323–48.
73. Fischer E. Anatomische Untersuchungen an den Kopfweichteilen zweier Papua. Corr Bl Anthrop Ges Jhg 1905;36:118–22.
74. Rhine JS, Campbell HR. Thickness of facial tissues in American blacks. J Forensic Sci 1980;25:847–58.
75. Nanda RS, Meng H, Kapila S, Goorhuis J. Growth changes in the soft tissue facial profile. Angle Orthod 1990;60:177–90.
76. Wilkinson CM. *In vivo* facial tissue depth measurements for white British children. J Forensic Sci 2002;47:459–65.
77. Farkas LG. Examination. In: Farkas LG, editor. Anthropometry of the head and face. New York: Raven Press, 1994;3–56.
78. Brown RE, Kelliher TP, Tu PH, Turner WD, Taister MA, Miller KWP. A survey of tissue-depth landmarks for facial approximation. Forensic Sci Commun 2004;6:1. Available at: http://www.fbi.gov/hq/lab/fsc/backissu/jan2004/research/2004_01_research02.htm Accessed July 30, 2008.
79. Phillips VM, Smuts NA. Facial reconstruction: utilization of computerized tomography to measure facial tissue thickness in a mixed racial population. Forensic Sci Int 1996;83:51–9.
80. George RM. The lateral craniographic method of facial reconstruction. J Forensic Sci 1987;32:1305–30.
81. Garlie TN, Saunders SR. Midline facial tissue thicknesses of subadults from a longitudinal radiographic study. J Forensic Sci 1999;44:61–7.
82. O'Grady JF, Taylor RG, Clement JG. Facial tissue thickness: a study of cadavers in Melbourne. International Association of Forensic Science Scientific Symposium. Adelaide: 1990.
83. Krogman WM, Sassouni V. Syllabus in roentgenographic cephalometry. Philadelphia: College Offset, 1957.
84. Briggs CA. Anthropological assessment. In: Clement JG, Ranson DL, editors. Craniofacial identification in forensic medicine. London: Arnold, 1998;49–61.
85. Relethford JH. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 2002;118:393–8.
86. Rushton JP. Race, evolution and behavior. Port Huron: Charles Darwin Research Institute, 2000.
87. Cooper RS, Kaufman JS, Ward R. Race and genomics. N Engl J Med 2003;348:1166–70.
88. Roumualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, et al. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 2002;12:602–12.

89. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. Human population genetic structure and inference of group membership. Am J Hum Genet 2003;72:578–89.
90. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, et al. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 2003;348:1170–5.
91. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 2007;3:1–12.
92. Taylor JA, Brown KA. Superimposition techniques. In: Clement JG, Ranson DL, editors. Craniofacial identification in forensic medicine. London: Hodder Arnold, 1998;151–64.
93. Aulsebrook WA. Facial tissue thickness in facial reconstruction. In: Siegel JA, Saukko PJ, Knupfer GC, editors. Encyclopedia of forensic sciences. San Diego: Academic Press, 2000;779–88.
94. De Greef S, Claes P, Vandermeulen D, Mollemans W, Suetens P, Willems G. Large-scale in-vivo caucasian soft tissue thickness database for craniofacial reconstruction. Forensic Sci Int 2006;159S:S126–46.
95. Wilkinson CM, Neave RAH, Smith D. How important to facial reconstruction are the correct ethnic group tissue depths? In: Colonna M, Belviso M, Addante A, editors. Proceedings of the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification, September 11–14, 2002. Bari, Italy: 2002:111–21.
96. Stephan CN, Norris RM, Henneberg M. Does sexual dimorphism in facial soft tissue depths justify sex distinction in craniofacial identification? J Forensic Sci 2005;50:513–8.
97. Norton K, Olds T. Anthropometrica. Sydney: UNSW Press, 1996.
98. Hirsch S, Thelen A, Frey S, Landriere N, Hering P. Ultrafast holography for the facial analysis of position dependent soft tissue movement. Third International Conference on Reconstruction of Soft Facial Parts. Leuven, Belgium: 2006.
99. Smith SL, Buschang PH. A new method for measuring soft tissue thicknesses of the face using ultrasound. Am J Phys Anthropol. 2004;123(Suppl 38):184–5.
100. Leopold D. Identifikation durch schadeluntersuchung unter besonderer berucksichtigung der superprojektion. Leipzig: Karl Marx-Universitat, 1968.
101. Sutisno M. Human facial soft-tissue thickness and its value in forensic facial reconstruction [dissertation]. Sydney: Department of Pathology, Faculty of Medicine, The University of Sydney, 2003.
102. Sarnas K-V, Solow B. Early adult changes in the skeletal and soft-tissue profile. Eur J Orthod 1980;2:1–12.
103. Dumont ER. Mid-facial tissue depths of white children: an aid in facial feature reconstruction. J Forensic Sci 1986;31:1463–9.
104. Ligthelm-Bakker ASWMR, Prahl-Andersen B, Wattel E, Uljee IH. A new method for locating anterior skeletal landmarks from soft tissue measurements. J Biol Buccale 1991;19:283–90.
105. Lebedinskaya GV, Balueva TS, Veselovskaya EV. Principles of facial reconstruction. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull. New York: Wiley-Liss, 1993;183–98.
106. Phillips VM, Rosendorff S, Scholtz HJ. Identification of a suicide victim by facial reconstruction. J Forensic Sci 1996;14:34–8.
107. Stewart TD. Evaluation of evidence from the skeleton. In: Gradwohl RBH, editor. Legal medicine. St. Louis: CV Mosby, 1954;407–50.
108. Edelman H. Die profilanalyse: Eine studie an photographischen und rontgenographischen durchdringungsbildern. Z Morphol Anthropol 1938;37:166–88.
109. Welcker H. Das Profil des menschlichen Schadels mit RontgenStrahlen am Lebenden dargestellt. Korrespondenz-Blatt der Deutschen Gesellschaft fur Anthropologie Ethnologie und Urgeschichte 1896;27:38–9.
110. Burkitt AN, Lightoller GHS. Preliminary observations on the nose of the Australian aboriginal with a table of aboriginal head measurements. J Anat 1923;57:295–312.
111. Gerasimov MM. Vosstanovlenie lica po cerepu. Moskva: Izdat. Akademii Nauk SSSR, 1955.
112. Lebedinskaya GV, Veselovskaya EV. Ultrasonic measurements of the thickness of soft facial tissue among the Bashkirs. Ann Acad Sci Fenn A 1986;175:91–5.

Additional information and reprint requests:
Carl N. Stephan, Ph.D.
Joint Pow/MIA Accounting Command
Central Identification Laboratory
310 Worchester Ave.
Hickam AFB, HI, USA 96853-5530
E-mail: carl.stephan.AU@jpac.pacom.mil