Chapter 12

Is It Possible to Escape Racial Typology in Forensic Identification?

John Albanese and Shelley R. Saunders

Summary

This chapter provides a review of metric and morphological methods for determining ancestry from skeletal forensic cases, as well as a comparative look at emerging genetic "origins"-determination methods. The authors address two major issues with respect to these methods. Are the methods consistent with observable patterns of human biological variation and with the apportioning of variation in skeletal reference samples used to represent population groups? Do the methods have any utility for positive identification of unknowns? In addition, the authors provide examples of the patterns of variation in cranial measurements, infracranial measurements, and morphological characters as observed in skeletal reference samples to illustrate some of the limitations of the underlying assumptions of "race"-determination methods.

The reality of human variation is not consistent with how forensic anthropologists have used (and continue to use) human variation to identify unknown individuals, and the substitution of various terms without a critical reanalysis of the underlying assumptions has not remedied the situation. False or misleading information is far worse than a lack of information. The relatively high risk of false information may outweigh the value of determining "race" may possibly have for the positive identification of an unknown individual.

Key Words: Race; personal identification; population affinities; ethnicity; discriminant function; morphological characters; anthroposcopic traits; nonmetric traits.

From *Forensic Anthropology and Medicine:*
Complementary Sciences From Recovery to Cause of Death
Edited by: A. Schmitt, E. Cunha, and J. Pinheiro © Humana Press Inc., Totowa, NJ

1. Introduction

The determination of ancestry population affinities or ethnicity (in the past, referred to as "race") is the most controversial question that a forensic anthropologist must face when assisting in identifying unknown individuals. In some parts of the world (for example, the United States, South Africa, etc.), there is a history of the use of racial classification as part of personal identification, and forensic anthropologists continue to be called on to address this question when positively identifying an individual. The recent trend in the forensic literature has been to use the term "ancestry" instead of "race," with no change in the underlying concepts, so that determining continental origin has been substituted for color terminology. Regardless of the terminology used, the underlying assumption *in forensic applications* is the same: using morphological, metric, or a combination of data, it is possible to assign an unknown individual into one of a limited number of continental or racial groups (usually two to six groups).

In this chapter, the authors provide a review of some of the metric and morphological methods for determining ancestry, as well as a comparative look at emerging genetic "origins"-determination methods. Two major issues with respect to these methods are addressed: are the methods consistent with observable patterns of human biological variation and with the apportioning of variation in reference samples used to represent population groups? Do the methods have any utility for positive identification of unknowns? In addition, two examples are provided, one cranial and one infracranial, to illustrate some of the limitations of the underlying assumptions of "race"-determination methods.

2. Theories and Methods for Allocating Unknowns: An Historical Perspective

Some of the earliest applications of skeletal biological methods to forensic cases date back to the 1930s in the United States. However, the racial approach to research during earlier periods had an enormous influence on physical anthropology throughout the 20th century and into the 21st century *(1–5)*. Comparative morphological and metric investigations of human variation related to race or continental origin date back to the mid-19th century in North America, with work by Samuel Morton and in Europe with research by Paul Broca *(6)*. Many of those investigations focused on the identification of racial traits, usually in the cranium, which were erroneously used to assess mental ability, rank various groups, support nationalist views, and justify social and economic inequality *(6)*. Although those approaches were theoretically and

methodologically flawed (*see* ref. *6* for a comprehensive review), they had an enormous influence on how physical anthropologists framed investigations of human variation throughout the 20th century.

These early approaches are more closely analogous to recent studies of population distance to *discriminate* between populations rather than *allocate* unknown individuals. Research directly related to allocating unknowns began when large documented "multiracial" collections became available in the United States *(7)*. Anatomists T. Wingate Todd and Robert J. Terry began amassing respective collections at medical institutions in Cleveland, OH and St. Louis, MO in the first decades of the 20th century *(7,8)*. These collections were formed at a time when race, as it was socially constructed in the first half of the 20th century, was considered as biologically meaningful as age or sex when investigating skeletal variation. The 19th century racial approach for investigating human variation is evident in how the collections were put together, what documentary data were collected and curated with the skeletal material, and in the research of the collectors (*see* refs. *9–12*). The racial designations in the Hamann-Todd Collection and the Terry Collection predate the adoption of modern evolutionary theory by physical anthropologists (*see* ref. *13*).*

By the latter half of the 20th century, at least three distinct race concepts emerged in the biological and social sciences: social race, bureaucratic race, and biological race (*see* refs. *14–17*) for different perspectives in a forensic context). Social and bureaucratic race are *socially constructed* concepts for grouping humans that are self-defined by individuals or groups, imposed by certain socioeconomic levels of society on others, or both. The biological race concept is theoretically based on phenotypic and genotypic variation. When asked to determine race or ancestry, forensic anthropologists are asked to determine social race or bureaucratic race based on morphological and/or metric variation. Whereas social race and bureaucratic race are real concepts that have social and economic effects on peoples' lives (an extreme example is Apartheid), the overwhelming evidence from many different studies—but particularly in the last 35 yr with the advent of protein and DNA analysis—clearly show that the race concept is not a valid biological concept, and that racial groups are not coarse but useful categories for investigating human variation *(4,18–26)*. The conclusions are consistent and clear *(21,22,25,27–29)*:

* Racial terms are presented here in quotation marks to highlight that these designations in reference collections are not based on phenotype and genotype.

1. Intrarace variation is *much* greater than interrace variation.
2. Only 6–13% of genetic and morphometric variation is attributable to race.
3. There is *no concordance* of human genetic and morphometric variation with racial categories, continental origin, or skin pigmentation.

Another trend in forensic literature in the last decade of the 20th century was to use the term "ancestry" instead of race. The terms "European," "African," and "(East) Asian" have replaced "Caucasoid/White," "Negroid/Black," and "Mongoloid/Yellow/Red." An example of this trend can be found in *A Lab Manual and Workbook for Physical Anthropology (30,31)*. This is a widely used introductory lab manual that has gone through several editions, where many students of physical anthropology get their first introduction to a "scientific" approach to racial classification of unknown individuals in forensic contexts. In the first edition *(30)*, France and Horn (p. 30) provide a summary of seven cranial morphologic and metric characters that can be used to classify an unknown individual as "Negroid," "Caucasoid," or "Mongoloid." In the fourth edition *(31)* of the lab manual, (p. 123, fig. 4.21) the same seven cranial morphological and metric characters are used to classify the unknown as being of "African," "European," or "(East) Asian" ancestry. The underlying assumptions are the same, but the terminology has changed.

3. Forensic Identification of Population Affinities

Generally, the investigation of human variation has followed two major methodological approaches, morphological and metric.* There has been some discussion about which is easier to apply with less training and which approach results in higher allocation accuracies *(33,34)*. The choice of method is almost entirely dependent on what variation is being observed. Some skeletal variation can be easily measured, and some variation can only be adequately assessed with a morphological approach (presence/absence or pronouncement of characters). Under the next two subheadings, metric and morphological methods for race determination are discussed, respectively.

3.1. Discriminant Function Approaches

In the early 1960s, Giles and Elliot *(35)* revolutionized "race" determination with the publication of their discriminant functions using samples of

* Three-dimensional digital approaches that combine both metric and morphological information are being developed for investigating patterns of human variation (*see* ref. *32* for a forensically relevant example). However, methods that are widely applicable in forensic cases are not currently available.

"Blacks" and "Whites" from the Hamann-Todd and Terry collections, and an aboriginal sample from Indian Knoll (*see* ref. *34* for details on applying these methods). Most metric methods developed since the mid-1960s, including recent computer based methods, are based on Giles and Elliot's approach involving discriminant functions (for example, *see* refs. *36–40*). Generally, cranial methods are considered the most reliable, pelvis and long bone combinations are less reliable, and other skeletal elements are considered the least reliable *(17,41)*.

Subsequently, Birkby *(42)* described several problems with Giles and Elliot's functions and with the entire approach for allocating individuals to a limited number of racial groups. He tested the Giles and Elliot method with aboriginal archaeological samples from across North America and found that they performed poorly. Only 52% of the crania where classified correctly because of two theoretical problems with metric race methods *(42)*. First, an unknown individual is forced into one of three categories regardless of whether that individual fits into any of those categories. Second, a category such as "American Indian" is not a single homogenous category. The allocation accuracies for most of the test samples were between 20 and 78%. The notable exception was the relatively high allocation accuracy (92%) for the test sample from Indian Knoll, the same population used by Giles and Elliot to develop their equations. Birkby's results suggest that it may be possible to determine ancestry and allocate an unknown individual to a specific biocultural group defined by geographic and temporal parameters, in this case, the Indian Knoll population. However, the method performed very poorly when it was applied to samples outside the reference sample used to develop the original method because racial categories consist of many heterogeneous populations that are not fixed through time.

In addition to Birkby's study, Giles and Elliot's method was tested using various identified and archaeological samples *(34,44)*. In all the studies, the methods performed poorly on the respective American Indian samples (accuracies ranged from 14 to 30%), and confirmed Birkby's conclusion that the Indian Knoll sample cannot be considered a proxy for the pattern of variation in numerous populations that are included in the group American Indian *(34,43,44)*.

Two of the tests of Giles and Elliot's method included forensic cases, and they resulted in allocation accuracies of 71.4 and 76.4%, lower than Giles and Elliot's original published accuracy *(34,44)*. Based on a review of the test results, allocation accuracy approaching the level described by Giles and Elliot could be achieved if Amerindian samples were left out of the test and if the Black–White sectioning point for males was modified *(34,44)*. For analytical purposes, knowing the allocation accuracy for each race and sex subsample

in the test is important. However, on a practical level, when applying the method to *one* unknown individual, what matters is whether the method can be applied with confidence in that one case. When dealing with a true unknown, for example, a forensic anthropologist in North America has no way of knowing if a given analysis is a case where the Black–White situation applies, whether the individual may be Amerindian or whether Giles and Elliot's sectioning point or the modified sectioning point should be used. The independent tests suggest that Giles and Elliot's method can be expected to give erroneous information in at least one out every four cases with no indication of when the method is providing incorrect information.

Other methods for determining race are available *(45)*. In many cases, the methods are based on relatively small sample sizes with uneven subsamples (few females and/or one group overrepresented), have not been tested with independent samples or have ignored Birkby's conclusions regarding the lack of homogeneity within a racial group and the problems associated with a forced allocation into only one of three categories. In cases when methods have been comprehensively tested, the resulting allocation accuracies have been low *(40,46–49)*. For example, in a comprehensive test of femur and tibia methods *(50)* using a large forensic sample from the Forensic Anthropology Databank (FDB), the allocation accuracy for "Whites" was worse than randomly guessing, and the high accuracy for "Blacks" was misleading because the functions simply classified almost everyone as "Black" *(46)*. As with methods that use only leg bones, methods that used the pelvis and leg bones *(37)*, had a tendency to classify most individuals as "Black," and a number of "Whites" were classified incorrectly *(46)*.

The general poor performance of both metric sex- and "race"-determination methods when tested with large independent samples has been attributed to problems with the representativeness of reference collections, specifically, the widely used Terry and Hamann-Todd collections *(46)*. The major problems with these collections have been noted for decades *(51,52)*. More recently, secular change has been described as an additional problem with the collections and a reason for the poor performance of forensic identification methods developed from these collections *(46)*. The FDB was established to address the poor performance of various "race"- and sex-determination methods and to address the problems with the older reference collections by providing an alternative source of data for the development of forensic methods *(46)*.

The FDB consists of data collected in forensic cases and submitted by various anthropologists, as well as a sample of individuals from the Terry Collection and a small number of individuals from the Hamann-Todd Col-

lection who where born in the 20th century *(40,46)*. Although an electronic database is not a substitute for a skeletal collection, the FDB has enormous research potential *(46)*. An example of this potential is the computer application known as FORDISC, which can be used to determine "race" and/or sex. FORDISC has several features that make it more useful than all other previous discriminant function approaches. First, unique discriminant functions are calculated based on what measurements can be collected from an unknown individual. Second, posterior and typicality probabilities are calculated in addition to the discriminant function score. The posterior probability is a measure of group membership, assuming that the unknown individual is in fact one of the options selected. The typicality probability is a measure of whether the unknown individual could belong to any of the groups selected in the analysis. This statistic addresses one of the major problems with all discriminant function approaches. Although the discriminant function score may force a placement into one of the selected groups, a typicality score of 0.05 or lower indicates that the unknown is not typical of any of the selected groups *(40)*. Aside from the interface, there are several substantive differences in the second version of the program. With FORDISC 2.0, infracranial measurements can be used, and there is the option of using Howells' *(53,54)* data instead of the FDB as the reference sample for the calculation of discriminant functions *(40,46,55)*.*

In contrast to the good results in an early test of FORDISC 2.0 *(55)*, when the method has been comprehensively tested with large independent samples, allocation accuracies when determining race are low (less than 60%) or do not follow any pattern of classification regardless of whether Howells' data or the FDB is selected as a reference sample *(48,49,56–58)*. Well-documented archaeological samples (i.e., continental origin was known) were used in several of these tests and may account for the low accuracies when the FDB data were used as the reference sample. This is a liability that was anticipated and clearly articulated by the developers of FORDISC *(40)*. However, in practice, when a forensic anthropologist is presented with an unknown individual there is no *a priori* way of knowing if the method should not be applied. Despite the confidence placed in FORDISC, there is little evidence that FORDISC performs at the 90% accuracy that has been reported *(55,59)* or performs any better than the earlier nonelectronic methods that rely on discriminant functions.

---

* FORDISC 3.0 has been recently released, but it could not be reviewed in time to be included here.

Because of the widespread confidence in FORDISC *(55,59)*, research on "race" and ancestry determination has largely been relegated to areas of the skeleton that are not included in the FDB and thus not available for analysis with FORDISC (for examples, *see* refs. *17* and *60–65*). In most cases, these methods provide reasonable allocation accuracies (generally more than 80%) for the sample used to develop the method. However, in cases where independent samples are used to test the methods *(47–49,64,65)*, allocation accuracies decrease to levels that undermine the applicability of the methods in actual cases. In one example *(65)*, the allocation accuracy of 75% for the test sample is misleading. The results follow a pattern for the femur previously described *(46)*, where every bone is classified as "Black." In another case, there are considerable differences in allocation accuracy by sex *(64)*. An allocation accuracy of 100% for females in the original sample dropped to 57% on a test sample of forensic cases. In tests of multiple methods, allocation accuracies were low, and there was no consistency between different methods when they were applied to the same individual *(47–49)*. For example, one unknown was classified as "Black" with the Giles and Elliot *(35)* method, "White" with the Gill *(17)* method, "Japanese" with FORDISC using the FDB data, and "from the Philippines" with FORDISC using the Howells' data *(48)*.

Discriminant function approaches have not resulted in highly reliable methods for race determination for various reasons. These problems are not restricted to the discriminant function technique, but they are exacerbated by this metric approach. Some of the contradictory results and low allocation accuracies result from limitations of the reference sample used to calculate a given discriminant function *(55)*. The Terry and Hamann-Todd collections and *all* reference collections, including the FDB, present some problems that are derived from how the collections were amassed *(66)*, but these are not insurmountable problems for developing forensically relevant methods *(67)*.

The poor performance of "race"-determination methods tested with the FDB data have been attributed to secular changes *(46)*. But there is some evidence that secular change observed in these skeletal collections may be a result of sampling error when data from separate collections are combined *(24,66)*. When trying to reproduce the results of Meadows, Jantz, and Jantz *(68)*, it was found that the increases in femur length occurred when the source of the data (i.e., collection) changed *(24,66)*. The differences in femur length between collections that are coincidently separate in time have been attributed to secular changes. Regardless of how this variation in reference collections is interpreted, research conducted in the first half of the 20th century, using living individuals and a sample from the Hamann-Todd Collection, demonstrated the nonconcordance between race and morphology when secular

change was not a confounding factor in reference collections *(69)*. The different patterns of cranial and infracranial morphology attributed to, for example, "White" males born in the 20th century vs "White" males born in the 19th century *(46)*, clearly demonstrates that skeletal variation is not fixed or genetically based for the group "White males." If the changes are secular changes, 100 yr (or roughly five generations) is too little time for such significant changes in morphology to be because of genetic variation. If the variation that has been attributed to secular changes is in fact because of sampling error derived from the methodology for sampling and limitation in the reference collections *(24)*, the different pattern of variation for "White" males born in the 20th century vs "White" males born in the 19th century clearly shows that the "White" group is not a valid category for apportioning human variation.

Although racial categories may not be a biological reality, it may be possible to use a statistical association in the reference sample to allocate unknowns *(15,40,70,71)*. The independent test results from various samples over the last 40 yr presented previously clearly indicate that regardless of how robust the statistical association is in the reference sample, this association does not result in a method that can be confidently applied to cases outside the reference sample used to develop the method. There are two reasons for this. First, the parameters used to define the groups do not correspond with the real patterns of human variation. Based on phenotypic variation, an unknown individual is forced into a group that is defined by *variable* social criteria and not phenotypic or genotypic parameters. Second, variation because of age, sex, cause of death, living conditions, and so forth, is incorrectly apportioned to a race (*see* the pelvis example under Subheading 4.2.).

3.2. Morphological Characters

There is also a long tradition in physical anthropology of using morphological characters of the skull to assign individuals to population groups. Historically, this attention derives from anatomists who had a broader interest in human physical variation as early as the 17th century *(72)*. In the 1920s, the anthropologist Ernest Hooton developed a recording form for such characters during his tenure at the Peabody Museum of Harvard University. Hooton's recording form is believed to have had a major influence on subsequent students and researchers in North America studying skeletal samples or forensic cases *(73)*.

There are two main kinds of morphological characters of the skeleton relevant to this discussion, anthroposcopic traits and nonmetric traits. *Anthroposcopic traits* are features of shape observable in all skeletons, such as a particular form of the palate or the position and height of the bridge of the

nose, whereas *nonmetric traits* are minor skeletal and dental variants that may or may not be present. When discovered, nonmetric traits appear to be curious anomalies, and are assumed to cluster in large or small population groups. Several hundred nonmetric traits have been reported for the skull and infracranial skeleton *(72,74,75)*. Table 1 offers a sample listing and descriptions of some anthroposcopic and nonmetric traits. Unfortunately, the forensic anthropology literature uses either term interchangeably *(76)* or inclusively (e.g., ref. *77* refers to all traits as "anthroposcopic," whereas ref. *78* refers to all traits as "nonmetric"). Whereas the word *anthroposcopic*, literally defined, means "to see human," the distinction between these two terms is necessary because there is a long and separate tradition of research in biology and physical anthropology on nonmetric traits as population descriptors.

Many early workers would classify certain nonmetric traits as characteristic of specific groups and name them accordingly so that we have features, such as the "os inca" or "os japonicum." Ossenberg *(74)* considered the question of whether a battery of nonmetric traits might separate world population groupings when she used 24 cranial traits to compare samples of Native American Indians, Eskimos, African Americans, and African Blacks. The calculated distance statistics were found to be higher between the major groups than within them. However, this was tested further when Wijsman and Neves *(79)* examined whether the frequencies of nonmetric traits would mirror the genetic distances between Brazilian Blacks, Whites, and mulattos, and a model of genetic population admixture. They found significant deviations in the pattern of nonmetric trait distances from a linear model of genetic distance, and in support of this observation, many nonmetric traits have been found to have low estimates of heritable variation *(80,81)*. More recently though, Hanihara and colleagues *(82)* reported a comprehensive study of the frequencies of 20 nonmetric cranial traits in several thousand individuals from many populations from around the world. Using multivariate statistical analyses to calculate distances, they found the variation to be, at least in part, because of geographical factors rather than environmental factors, and similar to distances calculated from genetic or craniometric data. Hence, there is some evidence that traits cluster in regional world populations, but that variation forms a minimal portion of the total worldwide variation, most of which is within local populations *(29)*. One also cannot expect trait frequencies to reflect directly genetic allele frequencies because they are phenotypic features far from the genome with a different model of inheritance.

Early work with laboratory animals developed the quasicontinuous model of inheritance for the genetic control of these minor skeletal variants *(83)*. A good illustration of this model is Grüneberg's thorough study of the absence

Table 1
Sampling of Anthroposcopic vs Nonmetric Traits

Anthroposcopic traits[a]	Definition	Comments
1. Sutures	Scored as "simple," "medium," or "complex" on the basis of the sutures tracing a path deviating from a hypothetical straight line	It is not clear how to judge the difference between "simple," "medium," and "complex". Byers *(77)* shows diagrams for simple and complex patterns only.
2. Nasal opening	The opening is triangular, flared widely at the base, or flared centrally and at the base as well.	Rhine *(76)* and Byers *(77)* provide diagrams, but the boundaries between triangular and flared may still be difficult to judge.
3. Nasal form	Horizontal contour across the nasal root are either (a) low and rounded; (b) low to moderate in height, with relatively straight sides and angled in the midline; or (c) high, somewhat pinched in, with a break in contour at or near the naso-maxillary suture	Described by Brues *(73)*, this trait is illustrated with photographs but still creates difficulties with judgment.
4. Nasal depression	The deepest point of curvature of the nasal bones just inferior to nasion is deeply depressed, slight depressed, or straight.	What constitutes a "slight depression"?
5. Nasal sill	Located where the vertical maxillae may create a sharp ridge separating the nasal cavity from the maxillae. If the ridge is high, the score is "deep;" if shallow, it is scored as "shallow;" and if a sharp ridge is lacking, it is "blurred." A smooth curve leading from the maxillae into the nasal aperture without interruption is "guttered."	The greatest difficulties appear to lie with judging "shallow" and "blurred" sills (*see* ref. *77*, Fig. 7.5). Note that all nasal features are likely to be highly correlated with one another.

(continued)

Table 1 (Continued)

Anthroposcopic traits[a]		
6. Alveolar prognathism	Scored as "large," "medium," or "none," depending on the amount of alveolar projection.	The degree of prognathism may be difficult to judge, but *see* Fig. 7.3 in Byers *(77)*.
7. Canine fossa	A depression in the maxilla at the root of the canine.	What constitutes a "minimal" expression?
8. Shape of the chin	The chin is "bilobate" (with a central sulcus), "blunt" (smoothly rounded), or "pointed," as viewed from above.	How does one decide when a case appears equivocal?
Nonmetric traits[b]		
1. Inca bone	Defined as a suture running from asterion to asterion dividing the squamous portion of the occipital approximately in half. Sutures cutting off smaller portions of the occipital are not scored as Inca bones.	The occipital squama inferior to the highest nuchal line is ossified in cartilage; the superior portion is ossified in membrane. Union of the two parts is said to occur in the third intrauterine month. If the parts fail to unite the upper portion is known as the os Inca. Wormian bones are common in the lamboid suture, and naïve workers have confused other variants—the os apices, the lambdic bone, or the lamboid wormians—for the os Inca, the rarest variant.
2. Os Japonicum	Defined by a horizontal suture running from the zygomaticotemporal suture anteriorly to the zygomaxillary suture isolating and inferior section of the zygomatic bone.	This trait is probably easiest to identify of this list, although the observer has to be careful of postmortem alterations to the region. Difficulties may arise with identification of "barely discernable" traces of the suture.

3. Shoveling of the incisors	In the upper central and lateral incisors, the lateral margins fold sharply backward, so that the tooth resembles a miniature scoop or shovel.[c]	Recording criteria vary. The most comprehensive are those of Dahlberg *(86a)*, who provided model casts of many dental traits.
4. Carabelli's cusp	Accessory cusps located on the mesiolingual cusps of maxillary molars. Most prominent on the first molar.	Readers of the many studies of this trait will recognize that there are several gradations of expression, including a pit.
5. Mandibular torus	A torus on the inside of the body of the mandible. Seen as a small "lump," either unilaterial or bilateral.	There is strong evidence that the appearance of this trait is affected by biomechanical factors (Ossenberg *[86b]*).
6. Supraorbital foramina	The supraorbital nerves, which supply the frontal region, may be encased in one of more foramina along the superior medial margin of the orbital border, or there may be a "notch" or the area may be smooth.	Ossenberg *(86c)*
7. Mylohyoid bridge	Accessory bridge of bone covering the mylohyoid groove on the medial surface of the body of the mandible.	Ossenberg *(86c)*
8. Paracondylar process		Ossenberg *(86c)*

[a] Definitions of anthroposcopic traits are taken from Gill and Rhine *(45)*. Only a sampling of cranial–skeletal and dental traits was selected for this table even though there are many infracranial traits.

[b] Traits were selected based on frequent references to them in forensic anthropology sources. However, the last three traits—supraorbital foramina, mylohyoid bridge, and paracondylar process—have been identified as signifiant in separating groups.

[c] This definition is taken from St. Hoyme and Iscan *(14)*.

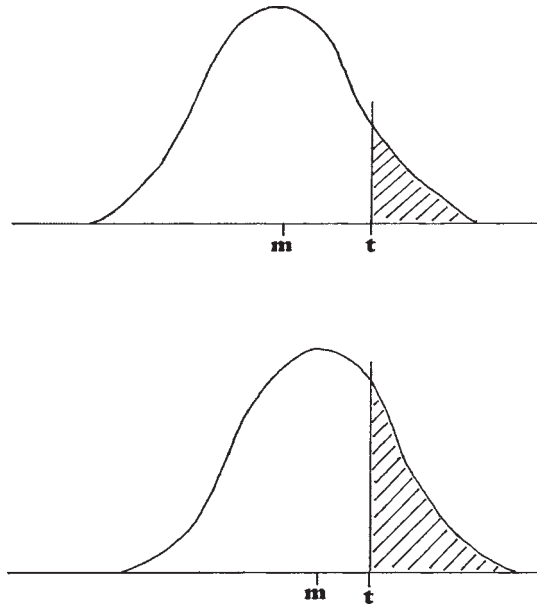

Fig. 1. Quasicontinuous model of inheritance taken from ref. *75*, pp. 97. Printed with permission.

of third molars, a trait also found in humans. He observed that the absence of the tooth is a discontinuous character arising from an *underlying continuous distribution* (Fig. 1), the size of the tooth rudiment. Tooth germ size is determined by the individual's genome and influenced by the genetic constitution of the mother, the maternal environment, and prenatal and postnatal environmental factors. Usually, the genes involved are multiple genes with additive effects. Tooth absence occurs if germ size falls below a critical level, shortly after birth in the case of mice. Thus, the expressions of size variations are affected by generalized and localized factors; whatever influences size will indirectly affect the presence of third molars. (For reviews of the problems and potential of nonmetric traits in population studies, *see* refs. *75* and *84–86*.)

Whereas some writers have admonished researchers for scoring nonmetric traits as discrete (present or absent) because they will vary in expression, Grüneberg's early model had already established that underlying continuity of liability was the correct way of interpreting them. Presence or absence recording usually improves the precision of observations (consistency in recording), whereas consistency in observation is probably the greatest difficulty with anthroposcopic traits (*see* comments in Table 1 and Fig. 2). This

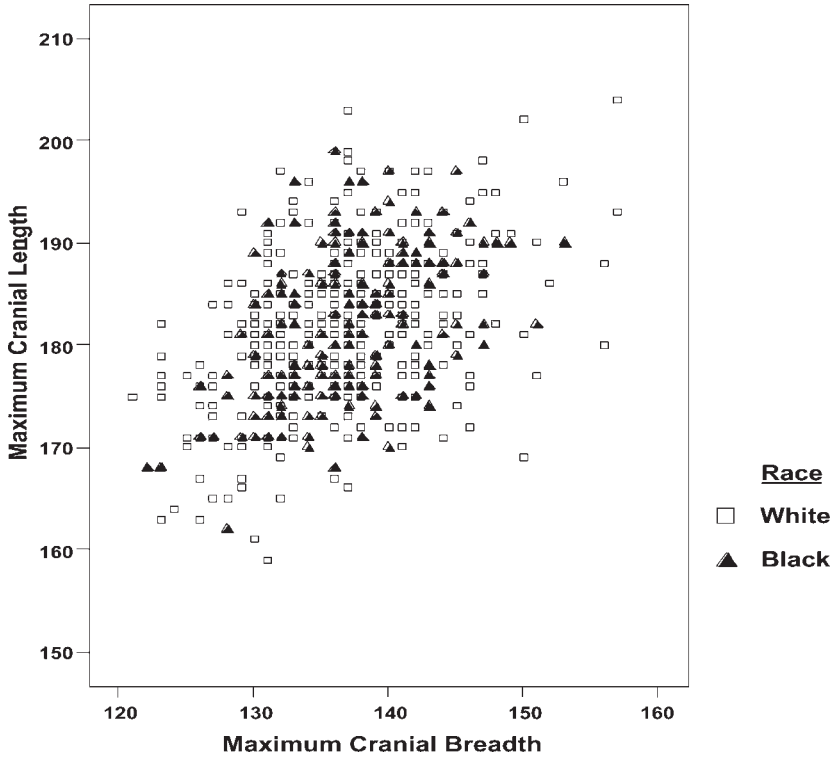

Fig. 2. Scatter plot of cranial length and width measurement by "race" of males and females from the Terry Collection and the Coimbra Collection (n = 526, *see* Table 2 for further details on sample composition). Note how the range of variation of the "Black" sample is entirely within the range of variation of the "White" sample.

can be said despite the main argument for using morphological traits to identify ancestry in forensic casework, which is ease of observation and recording *(76–78)*. Visual assessments require no expensive or delicate equipment and can be completed rapidly but could be useless if the collected data are faulty and imprecise. In addition, most texts will tell the student that considerable experience with recognizing traits and skill in forensic anthropology are necessary before employing traits to judge a forensic case. This contradicts the claim that they are easy to employ and warns us that the method may, in fact, be quite difficult.

What many seem to forget or neglect to mention is that the genetic backgrounds of trait causation can vary from individual to individual and from

group to group. Common features may cluster in members of a large family. Major gene effects can modify skeletal or dental development and produce traits that are produced by other genetic factors in other individuals and populations. The discovery that different mutations can produce the same phenotypic effect was recognized long ago in the field of genetics but seems to be ignored in forensic anthropology.

Most of the forensic anthropology literature on the subject of ancestry informs the reader that the goal is to assign individuals to one of three major groups: White (or Caucasoid), Black, or Asian (including North American Indian) *(76,77)*. This is reflective of the American literature, which is where most of this information is published and where ancestry determination seems to be a significant goal of forensic casework. In fact, the American publications also refer to Hispanic as a "neorace" *(76)*, and these persons are defined as of mixed European and Native American heritage. Rhine reported on a test of a list of 45 mixed, anthroposcopic, and nonmetric features that were observed on a sample of 87 documented skulls with known backgrounds. Rather than report success rates of race assignment against documented race, he reported the frequencies of traits in the different groups (defined based on written documentation), including a listing of traits found in 30% or more and 50% or more of each sample, along with a notation of expectations. He recognized that the classification of the group samples could be problematic, stating, "We are not dealing with unmixed populations," and "not only is there a great deal of systematic populational variability (racial variability), there is a considerable amount of idiosyncratic variability as well." In fact, of the 45 traits in this study, 37 were found to be 30% or more frequent in more than one group. Six of the remaining eight traits were simply too rare and not found in this sample.* However, in conclusion, Rhine pointed out that even though there is a continuum of variation for morphological characters, making them hard to assess, they are of value in forensic cases where one cannot be confident of measurements because of fragmentation or postmortem alteration to the remains. On the other hand, some readers would interpret these results as disappointing, suggesting that the exercise of ancestry assessment from anthroposcopic and nonmetric traits should be rejected.

Even though recent forensic anthropology texts caution that there is no such thing as a pure ethnic group, race, or ancestral group, and that there is considerable overlap of traits that characterize different groups (as shown in previous paragraph) so that "the attribution of ancestral group is one of the

* A large portion of, but not all, ambiguity occurred within the Hispanic sample.

most difficult assessments made for skeletal remains" *(77)*, many in the field still claim that the anthropologist *must* communicate this information to law enforcement personnel, the general public, and students. It is worthwhile reconsidering this basic claim.

Byers' *(77)* recent text in forensic anthropology states that, when possible, forensic anthropologists should give an assessment of ancestry from skeletal remains with the categories of "White," "Black," "Asian," "Native American," and "Hispanic." This is an American text referring to categories used by many law enforcement agencies in the United States. In comparison, in Canada, the situation does not appear to be so straightforward. Canada is a country of immigrants, and the 1996 and 2001 censuses report people originating from Europe, the Middle East, Western Asia, Southern Asia, Eastern and Southeast Asia, Africa, the Pacific, and the Caribbean. In addition, the proportion of persons reporting multiple continental origins is 36%!

Whereas it seems that, particularly in the United States, the imperative to identify ancestry is tied to issues of racism, in Canada, missing persons lists use a two-variable category relating to ancestry, White or nonWhite *(87)*. Ascribing an unknown to either of these two categories will undoubtedly assist in narrowing the possible matches for identification, but it is not the only important variable. More research on establishing careful estimates of other biological parameters, such as age at death and stature, can do much to improve the success of individual identification. The authors illustrate with an example. In the spring of 2001, one of the authors (S. Saunders) was called to a rural road outside of Hamilton, Ontario, by police investigators. The partially skeletonized remains of an individual had been discovered under melting snow. Foul play was evident from the presence of perimortem trauma to the skull.

After assisting with the recovery, both authors evaluated the remains with anthropological methods. A suggestion was made that the individual might be of Southern Asian ancestry based on prominent presence of alveolar prognathism, convexity to the nasal profile, concavity beneath the border of the nasal spine, and moderate shoveling of the maxillary incisors. In the meantime, police investigators were researching the sources of some clothing and jewelry items found at the scene. In addition, they were attempting the rehydration of fingerprints from some preserved skin. Ultimately, the woman was identified by the recovered fingerprints matched to a criminal record. The suggestion of ancestry had been of some help in narrowing the investigation, but it was the combination of recovered information from a variety of investigated sources that led to the solution of the murder. In fact, the investigators

considered the anthropologists' estimation of age at death of the victim to be of equal or greater significance in contributing to identification.

4. Interpreting the Sources and Patterns of Variation in Reference Collections

Despite the more recent prominence of the FDB, arguably, the most important collection for the development of "race" and ancestry-determination methods has been, and continues to be, the Terry Collection. The Terry Collection has been continuously available for research for more than 60 yr; it is an important component of the FDB and FORDISC *(40,46)*, and it has been a major source of data for the "race"-determination methods that have been widely used for the last five decades *(35–37,39–41,88–90)*.

Using data from the Terry Collection and the Coimbra Collection (a cemetery-derived identified collection from Portugal), one cranial example and one infracranial example are presented here to illustrate some of the potential problems with identifying and interpreting sources of variation in reference collections. The first example illustrates the lack of concordance between cranial variation and racial categories or continental origin. Using the pubic bone, the second example illustrates that statistical significance, without historical and biocultural context, may lead to the apportionment of variation to the wrong source. Both examples illustrate the theoretical and methodological limitations of determining social or bureaucratic race from skeletal remains, and how expected patterns of variation have been described in scholarly literature and popular discourse even when the data did not support the perceptions *(91)*. This second issue is analogous to Walker's *(92)* observations on sex determination, where results can be driven by the expectations of researchers rather than actual observable patterns of variation.

4.1. Example 1: Patterns of Variation in the Cranial Index

For more than 150 yr, the cranial index and the cephalic index were used as tools for investigating human variation and to classify individuals into racial categories *(6)*.* The *cranial index* is defined as cranial breadth divided by cranial length multiplied by 100. The cranial index is calculated with data collected from skeletal material, and the *cephalic index* is the equivalent col-

* By the early 20th century, Boas's *(93)* research showed that cranial shape as approximated by the cephalic index was influenced by environmental factors and was not fixed. Two separate reanalyses of Boas's original data have reignited the debate over the plasticity of cranial shape *(94–97)*.

Table 2
Mean Cranial Indices and Sample Sizes by Unit of Analysis (Collection, "Race," Sex)

Unit of analysis	Mean	*n*	Standard deviation	Standard error	Minimum	Maximum
Co males	73.4	116	3.03	0.282	66.8	83.3
Co females	74.4	118	2.77	0.255	67.6	82.5
Te "Black" males	74.6	84	3.41	0.372	67.9	83.0
Te "Black" females	76.0	91	2.63	0.276	68.8	82.4
Te "White" females	77.5	68	2.84	0.344	71.8	84.2
Te "White" males	77.5	49	4.06	0.580	68.9	88.8

Co, Coimbra Collection; Te, Terry Collection.

Note how the Terry Collection "Blacks" are intermediary between the Coimbra Collection sample and the Terry Collection "Whites" and the range of Terry Collection "Blacks" falls within the range of the Coimbra Collection sample.

lected on living subjects *(98)*. These measurements have been used to calculate cranial index scores, and these scores are often converted into categories of cranial shape that range from long crania to hyperround crania: dolicocranic, up to 75; mesocranic, 75–79.9; brachycranic, 80–84.9; and hyperbrachycranic, 85 or greater *(98)*.

Data for this example were collected from the Terry Collection *(8)* and the Coimbra Collection *(99,100)*. The sample was selected to include and account for a wide range of variation associated with age at death and year of birth *(67)*. Details regarding sample size are available in Table 2. Figure 2 is scatter plot of maximum cranial length by maximum cranial breadth from both collections combined into racial categories. There is no pattern in the distribution of variation by "race" and the range of variation of the "Black" sample is entirely within the range of variation of the "White" sample.* In contrast, Fig. 3 is a scatter plot of the same data but coded by sex instead of race. As expected when considering sexual dimorphism in *Homo sapiens*, there is a clear clustering by sex and overlap in the ranges of both sexes. In other words, there is a clear pattern of variation in cranial morphology by sex but not by "race."

When looking at the cranial index instead of its component parts, racial categories still do not explain the variation in the samples. Figure 4 is a plot

* The data follow the same pattern, complete overlap between races, when it is graphed for each sex separately for the Terry Collection alone (not shown here). Sex differences are not obscuring "race" differences.

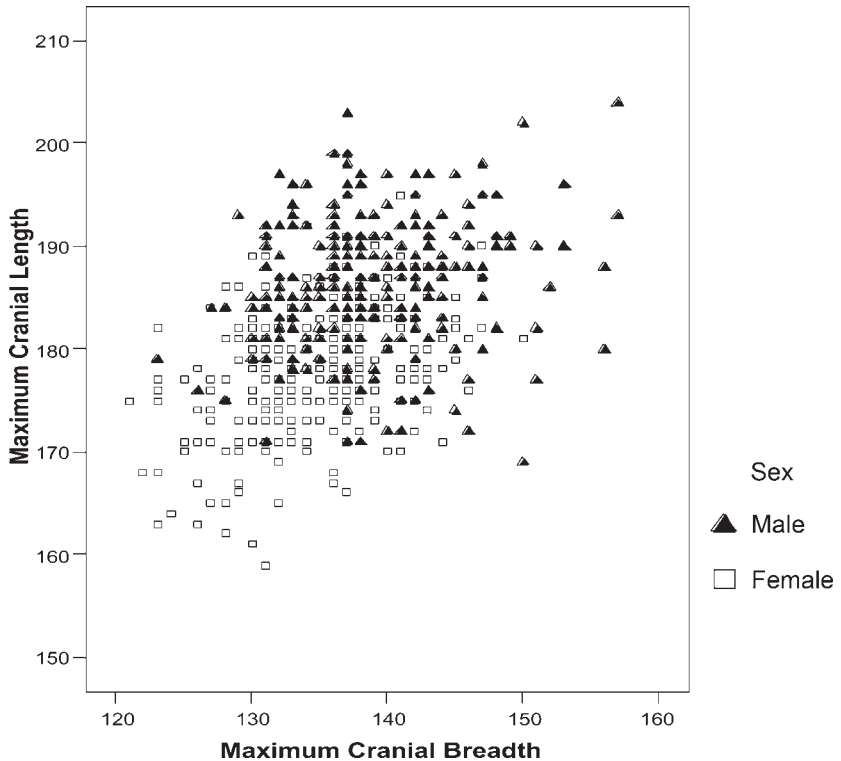

Fig. 3. Scatter plot of cranial length and width measurement of males and females from the Terry Collection and the Coimbra Collection (*n* = 526, *see* Table 2 for further details on sample composition). Identical data from Fig. 2 are displayed but coded by sex. Note how there is a clear cluster of data by sex.

of the 95% confidence intervals of the mean of the cranial index by unit of analysis (samples divided into collection–"race"–sex groups). The mean cranial index does not follow cited racial patterns. "Blacks" are usually described as dolicocranic *(17,31,36,41,90,98,101)*. "Whites" are alternatively described as dolicocranic *(101)*, mesocranic *(17,31)*, both dolicocranic and mesocranic *(90)*, brachycranic *(36)*, and as spanning the mesocranic and brachycranic categories *(41,98)*. The results from the current analysis show that the mean for "Black" males (74.6) is only marginally dolicocranic, and the mean for "Black" females (76.0) is mesocranic. The means for the European-born Coimbra Collection females (74.4) and males (73.4) are in the dolicocranic range, whereas the means for the Terry Collection "White" females (77.5) and males (77.5) are in the mesocranic range. Based on the mean cranial index,

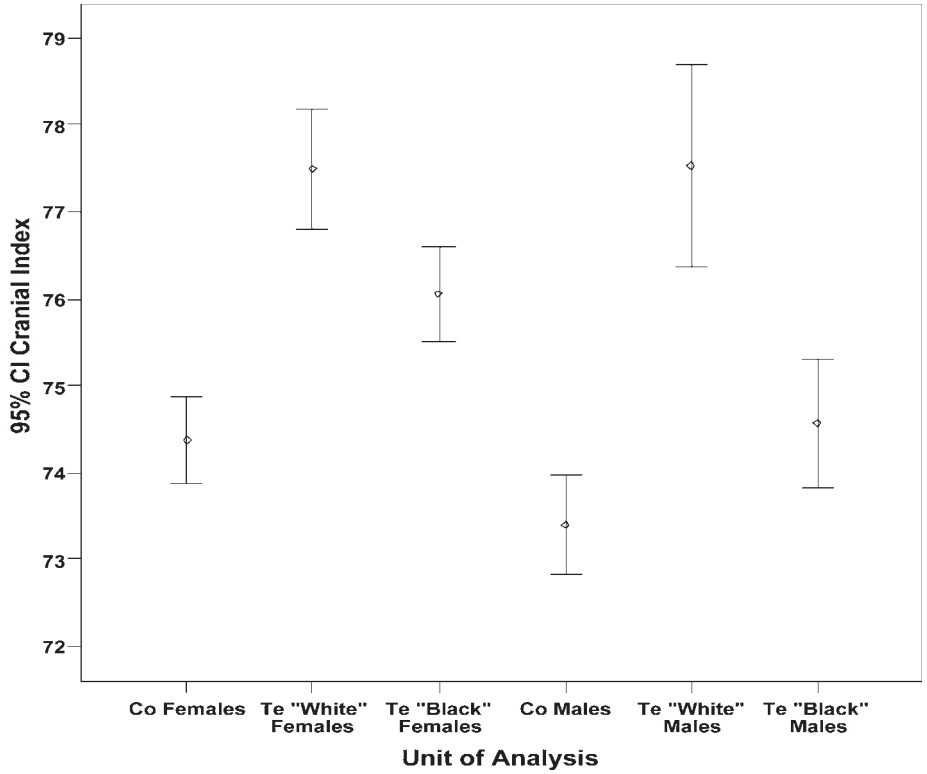

Fig. 4. Plot of 95% confidence interval of mean of cranial index by unit of analysis (collection–"race"–sex). Mean cranial shape as approximated by the cranial index does not follow often cited racial patterns (n = 526). Co, Coimbra Collection; Te, Terry Collection. Note: For each sex, the Terry Collection "Blacks" follow a pattern that falls between the Terry Collection "Whites" and the Coimbra Collection.

"Whites" from the Terry Collection are more similar to the "Blacks" from the Terry Collection than they are to the European-born Coimbra Collection sample. When the range of the cranial index is considered for each of the units of analysis, the entire range of variation in the Terry Collection "Black" sample falls within the range of variation in the Coimbra Collection sample. The similarities between the Coimbra Collection and the "Blacks" from the Terry Collection are not unexpected because these two samples are derived from the most disadvantaged segments of their respective communities *(66,102)*. Using a one-way analysis of variance with the Tukey honest

significant test *post hoc*,* the means of the Terry Collection "Whites" and the Coimbra Collection are significantly different ($F = 24.981$, $p < 0.0001$) from each other, whereas the Terry Collection "Blacks" are intermediary between the two "White" samples. Because of the sampling methodology (which controlled for age at birth and age at death), the significant differences between Terry Collection "Whites" and Coimbra Collection individuals are likely not because of age factors or secular changes *(67)*. This pattern of results, nonconcordance between cranial morphology and skin pigmentation, is not unique to the Terry Collection. Todd and Tracy *(11)* studied various facial traits, cranial traits, and the cranial index from samples of American "Blacks," an archaeological sample from Africa, and an archaeological sample from Europe. They found there was considerable overlap in variation between the three samples, and African and American "Blacks" did not cluster into one group just as American and European "Whites" did not cluster in the current example.

Although the components of the cranial index are rarely used alone for race determination, these measurements are the foundation for several prominent approaches *(35,40)*, and the cranial index categories are widely described as racial characters *(17,31,36,41,90,98,101)*. These differences between Terry Collection "Whites" and the European-born Coimbra Collection sample illustrate that morphometric variation in the cranial index or its components is not concordant with racial categories or continental origin. This example is consistent with Relethford's *(21,25,29)* conclusions regarding the lack of association between skin color and human variation, which is theoretically incompatible with a more widely held view *(15,70,71)* that it is possible determine "race," continental origin, or skin color with a reasonably high accuracy outside of the reference sample used to develop the method.

4.2. Example 2: Variation in Pelvic Dimensions and the Misinterpretation of Mortality Bias as Racial Variation in the Terry Collection

Data for this second example were collected only from the Terry Collection *(8)*, and as in the example above, the sample was selected to control

* Tukey's honest significant test was selected because it is neither too conservative (as with the Scheffe or Bonferroni tests) nor too liberal (as with the least significant test) in assessing significant differences, and Tukey's honest significant test is both a multiple comparison test (pairwise comparisons are made between means to identify significant differences) and a range test (similar means are grouped into homogeneous subsets).

Table 3
Mean Age at Death for the Entire Terry Collection and for the Subsamples Used in the Current Analysis

	Entire Terry collection[a]			Current sample		
Unit	*n*	Mean[b]	SD	*n*	Mean[c]	SD
Black females	366	51.75	19.05	80	36.35	9.15
White females	306	65.39	14.21	50	45.98	11.34
Black males	531	47.44	15.88	56	45.66	12.45
White males	453	59.30	13.37	37	50.46	12.96
Total	1656	229				

[a] Includes individuals 18 yr of age and older whose age is certain and who where classified as "White" or "Negro/Black" on original morgue documents.

[b] All means in this column are significantly different from other means in the same column at the $p < 0.0001$ level.

[c] The mean age for "White" females is significantly higher than "Black" females ($t = 5.317$, $p < 0.0001$). There are no significant differences in the mean age of "Black" males and "White" males ($t = 1.790$, $p = 0.077$).

for age-at-death and year-of-birth effects *(67)*. In Table 3, the mean age at death for this sample is compared with the Terry Collection as a whole for each unit of analysis. The summary statistics for the entire Terry Collection are based on individuals 18 yr of age and older whose age is certain, and who where classified as "White" or "Negro/Black" on original morgue documents *(8)*. There is a clear age bias in the Terry Collection that is confounded with racial designations, year of birth (YOB), and procedures for adding to the collection *(8,103)*. For the entire collection, the mean age at death for each unit of analysis is significantly different from the mean for every other unit of analysis ($p < 0.0001$). The methodology used to select the sample for this study has reduced some of the effects of age at death and YOB: the mean age for "White" females is still significantly higher than "Black" females ($t = 5.317$, $p < 0.0001$), but the difference in mean age is lower; there are no significant differences in the mean age of "Black" males and "White" males ($t = 1.790$, $p = 0.077$).

After the skull, the pelvis has been considered a good source of information for determining ancestry or "race" *(36,37,39,41,88,89)*. For this example, an alternative measurement of the pubic bone known as superior pubis ramus length was collected *(67)*. A significant cubic relationship was found between age and the superior pubis ramus length ($r^2 = 0.18$, $F = 8.61$, $p < 0.0001$), but

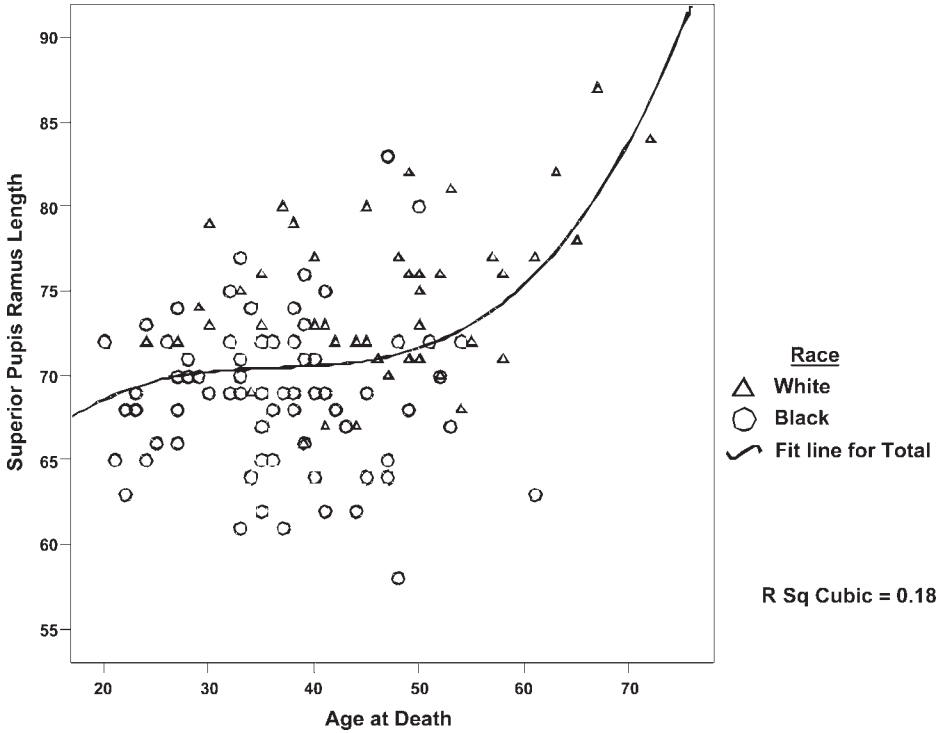

Fig. 5. Scatter plot of superior pubis ramus length by age at death for a subsample of females from the Terry Collection. *See* Table 3 for details regarding sample size. The line represents the cubic relationship between the variables for all the females. Note how the "White" females have consistently larger pelvic dimensions, but they are also consistently older.

only in females, and this relationship is graphically illustrated in Fig. 5.* This difference in pattern by sex is expected for biological and sampling reasons. The association between age at death and pelvic dimensions in females in archaeological and reference collection samples has been investigated, and various explanations have been suggested *(104,105)*. In this example, the significant association between age at death and pelvic dimensions is likely because of a mortality bias. Death resulting from complications from childbirth is not listed as the cause of the death for any females in the Terry Collec-

* An analysis of other pelvic measurements indicates the same pattern in the most sexually dimorphic elements of the hip bone including the iliac breadth.

tion. Rather, the correlation of pelvic dimensions with age is likely a nonspecific health indicator for females. Females with larger pelves likely had better living conditions during the period of their growth and development and also lived longer. The attribution of the variation in the pelvis to race is because a disproportionate number of younger females with smaller pelves were socially described as "Black" when they were included in the collection, and a disproportionate number of older females with larger pelvises were socially described as "White" when they were included in the collection.

Different selection pressures on the pelvises of males, who obviously will never bear children, result in a different pelvic morphology and different patterns of variation under various environmental conditions *(106–108)*. Additionally, the sample selection methodology described above was more successful in minimizing the age bias in the male sample (*see* Table 3). This current analysis actually understates the age–YOB–"race" bias in the Terry Collection. The sample selection methodology for this example reduced but did not eliminate the effects of age at death and YOB.

Pelvic "race"-determination methods that use samples drawn from the Terry Collection *(36,37,39,88,89)* allocate unknown individuals on the basis of age-related variation in the pelvis, which has incorrectly been attributed to "race." These race-determination methods work best to allocate females from the Terry Collection *(36,39,88,89)*, the samples where age-at-death and YOB differences are greatest in that collection because of historical accidents in how the collection was assembled. With one method, when age was recognized as an issue there was an attempt to statistically eliminate the effects of the "aging process" from the method *(89)*. Whereas the reductions in allocation accuracy were considerable for the adjusted functions, particularly for females who were most affected by age at death, the methodology for controlling for effects of age did not necessarily control for the true effect of age *(109)*, and the authors assume rather than demonstrate that the rest of the variation is because of "racial, and thus genetic, differences" *(89)*. Statistically significant association without any biocultural context for the variation has resulted in the apportionment of variation to the wrong source, and the resulting methods cannot be confidently applied to real forensic cases.

5. Genetic Identification of Population Affiliation: Relevance to Forensic Anthropology

DNA fingerprinting (a technique for identifying individual organisms based on the uniqueness of their genetic pattern) is a method of identifying perpetrators and victims of crime and is now widely accepted as scientifically

valid and acceptable as evidence in court. The emphasis has always been on individual identification or the matching of the sample of DNA to a specific person. More recently, however, as a result of the development of large public and private databases of genetic information and demand from the public, genetic ancestry testing (or, allocation of an individual to a specific population group) has become a growth field in the United States *(110)* and is having an influence on forensic investigations (*see* ref. *111*, and example in the following paragraphs).

There are large numbers of North American peoples wishing to trace their genealogical roots, and this has fostered the appearance of a number of private companies offering to trace personal genetic histories (PGH) by comparing individual samples to genetic data on human genetic polymorphisms from a variety of human populations. As of August 2004, there were 11 sources listed on the Internet offering fee-for-service tests of genetic ancestry *(110)*. Currently, there are two methods of tracing PGH: lineage-based tests, which amplify mitochondrial DNA (mtDNA) and the nonrecombining Y chromosome, and biogeographical ancestry (BGA) or autosomal marker-based tests, which purport to use genetic markers on the autosomal chromosomes informative of ancestry (ancestry-informative markers) to place people within biologically and geographically defined populations. Most existing tests are lineage-based, taking advantage of the fact that mtDNA and Y chromosome DNA do not recombine at fertilization (the genetic material comes either from the mother or the father), are more likely to accumulate marker mutations within lineages because of the smaller number of ancestors (a smaller effective population size, and have higher mutation rates, contributing to substantial variability. The results of the tests currently offered are designed to determine whether an individual has paternal or maternal lineages that originate from Native American, European, African, or Asian populations. The less common BGA tests aim to estimate a person's ancestry in terms of the proportional representation of ancestry-informative markers from a selection of reference databases treated as representing ancestral populations. Determination of ancestry is based on statistical tests of probability of ancestry by the maximum likelihood approach (a statistical concept used to quantify the probability that a certain hypothesis or model is correct given a set of data).

BGA tests have been applied to a recent forensic case. In the spring of 2003, the murders of five women in Louisiana had been linked through sample analyses of forensic DNA samples by short-tandem-repeat marker panels (also called microsatellite markers, these are short DNA sequences, typically from one to four nucleotides long, that are tandemly repeated several times) to implicate a single perpetrator of the crimes. However, there were no hits when

the genetic sequence was compared with the national combined DNA index system (CODIS) database of convicted felons. The police had restricted their investigations to White men, screening more than 600 individuals. Then they sought the assistance of a company offering PGH screening. The results of the BGA tests indicated that the perpetrator of the murders was mainly of West African descent. This ultimately led to the arrest of a suspect whose short tandem repeat profiles matched those of the perpetrator.

Considering that geneticists attempting ancestry determinations are working from basic DNA code, it would seem that this approach would have great advantages over anthropological assessment of bones to identify a victim or unidentified decedent. The DNA of body cells is not altered during an individual's lifetime (except for cases of homeoplasy or somatic mutations within mtDNA) or subject to environmental influences. Once sequenced, DNA results should be uncontested; observations of sequences are not subjective or variable as with skeletal traits. In addition, the statistical calculations applied to estimating ancestry from genetic data are far more exacting and sophisticated than any methods currently used for nonmetric or morphological traits (metric methods are comparable, limitations lie with the reference samples as discussed under Heading 4). However, an examination of the scientific literature on genetic ancestry determination reveals a number of limitations to the approach and provides an object lesson to forensic anthropologists wishing to assign ancestry affiliation from skeletal remains. Many of the limitations described for genetic data are applicable to, or instructive for, forensic anthropology.

5.1. Limitations of Genetic Methods

Those in the field of PGH lament the fact that there needs to be an increase in the number of markers used for analysis because the more markers used, the higher the probability of estimating affiliations correctly. Theoretically, this applies to forensic anthropology too. A nonmetric or metric description of only the nasal area will provide considerably less information than a thorough examination of the entire skull. However, there are statistical problems with dealing with correlated data when many forensic traits are used.

More important than markers is the need in the genetics field for improvement, increase in size, and the sharing of genetic databases *(110)*. A few years ago, a European researcher identified the fact that there were many errors in a large public database of mtDNA sequences *(112–114)*. More recently, a similar claim has been made by others *(115)*. Not only do many genetic databases contain recording errors, but also, few have even considered the problems of quality of the background data on individuals whose DNA sequences are included in the databases. The gatekeepers of genetic databases often give no

details about the number and geographical spread of samples included, so it is difficult to even assess the quality of databases. The problems are just as prevalent within genetics as have been described for forensic anthropology.

For lineage-based tests, the maternal and paternal lineages sought do not represent the entire genetic make-up. For example, an individual's mtDNA comes from his or her mother, who received it from her mother, and so on. At the great-grandparental generation, only one of eight individuals of the great grandparents (mother's mother's mother) is being sampled. In forensic anthropology, all of the phenotypic data represent recombined genetic data from all ancestors, and the nature of the morphology observed becomes the basic problem. Forensic anthropology texts state that remains that exhibit ambiguous (or mixed) ancestral groupings should be assigned to the group that is considered the minority (example in United States: a skeleton that exhibits both White and Black "features" should be assigned as "Black") because this is how they would have been classified in life. Now the forensic anthropologist (and the geneticist) becomes mired in social definitions of race. How many cases are there of mixed heritage individuals who functioned in their own "chosen" racial group or even changed designation several times in their lifetimes?

These limitations illustrate that PGH estimation is far from being an exact science, as some of the practitioners admit *(110)*. They also show that highlighting genetic differences among people might unfortunately reinforce the stereotypic features of these identities, a risk to forensic anthropology as well because judging the unknown to come from a specific group can limit the investigation as well. Nevertheless, the desire on the part of many to link genetic phenomena to ancestry or race cannot be ignored. Many want PGH estimation to justify their socially mediated constructions of population differences. In the medical field, a resurgence of interest in race relates to questions of risk for various disease conditions and the risks of blood transfusion reactions for those of different population origins *(111)*. Surely, a clearer understanding of the complexities of biological population diversity can only illuminate the debates that swirl around these issues.

6. Conclusions

The reality of human variation is not consistent with how forensic anthropologists have used, and many continue to use, human variation to identify unknown individuals, and the substitution of various terms without a critical reanalysis of the underlying assumptions has not remedied the situation. With no biological basis for racial categories, how can forensic anthropologists determine social race or bureaucratic race? Several authors have suggested

that such a contradiction is not an impediment to determining "race" and that high allocation accuracies are possible with various "race"-determination methods *(15,70,71)*. The claim of 90% accuracy that has been reported for "race"-determination methods *(15,17,59)* is unsubstantiated. Despite the relatively high-allocation accuracies (often more than 80%, but rarely more than 90%) and the strength of the statistical significance that are noted when various methods are first described, the comprehensive independent tests of "race"-determination methods consistently result in low-allocation accuracies. Some forensic anthropologists have argued that race determination is a forensic necessity, and forensic anthropologists would be either derelict in their professional responsibilities or ill equipped to positively identify an unknown individual if the assessment of "race" is not investigated *(15,17,31,70,71)*. False or misleading information is far worse than a lack of information. The relatively high risk of false information outweighs the value that determining "race" may possibly have for the positive identification of an unknown individual.

Cranial and infracranial variation in different groups living under various biocultural conditions through time and space is a reality, but this variation does not neatly cluster into two to five racial categories or by continent. A racial approach for identifying unknown individuals in a forensic context will be typological because:

1. It ignores the heterogeneous patterns of phenotypic variation in the highly plastic species *H. sapiens*.
2. It runs contrary to the genetic evidence that there is a great deal of genetic homogeneity in *H. sapiens*.
3. It ignores the fact that both phenotypic and genotypic variations are continuous.
4. It tries to categorize continuous phenotypic and genotypic variation into a few *socially* constructed categories.

New methods or updates of older methods, new collections, and new terminology (use of ancestry without a reevaluation of the underlying concepts) will not solve the problems associated with "race"-determination methods if these four issues are not considered.

The greater problem is that racial designations are part of various folk taxonomies that are related to social and economic issues and inequality rather than any phenotypic or genotypic reality. As Brace *(70)* notes when referring to the various waves of migration to North America of African, Asian, and European peoples, "the social barriers between these three *artificially distinct* human constituents of the Western Hemisphere have ensured the perpetuation of discrete identity of those components despite an increase in the

blurring around the edges, and this is what constitutes the 'reality' that is the 'something there' for the forensic anthropologist to discover" (p. 174, emphasis added). Socially, race is relevant and law enforcement authorities continue to ask about race because it plays a prominent role in personal identification and racial issues are prominent in the justice systems in various jurisdictions. In Canada, a country of great population diversity, law enforcement personnel are as aware of the problems of inferring geographic and population background as are the anthropologists. The simple argument that "investigators require it" is not sufficient to justify the claim for a noncritical application of anthropological methods. The authors think that the law enforcement field can be receptive to critical explorations of the complex interrelationships between sociopolitical processes and the scientific knowledge affecting our understanding of human physical diversity because the police are having to deal with such explorations within their own ranks.

References

1. Brace, C. L. The roots of the race concept in American physical anthropology. In: Spencer, F., ed., A History of American Physical Anthropology, 1930–1980. Academic, New York, NY, pp. 11–29, 1982.
2. Blakey, M. L. Skull doctors: intrinsic social and political bias in the history of American physical anthropology with special reference to the work of Aleš Hrdlièka. Crit. Anthropol. 7:7–35, 1987.
3. Blakey, M. L. Beyond European enlightenment: towards a critical and humanistic human biology. In: Goodman, A. H., Leatherman, T. L., eds., Building a New Biocultural Synthesis: Political-Economic Perspectives on Human Biology. University of Michigan Press, Ann Arbor, MI, pp. 379–406, 1998.
4. Armelagos, G. J., Goodman, A. H. Race, racism, and anthropology. In: Goodman, A. H., Leatherman, T. L., eds., Building a New Biocultural Synthesis: Political-Economic Perspectives on Human Biology. University of Michigan Press, Ann Arbor, MI, pp. 359–378, 1998.
5. Armelagos, G. J., Van Gerven, D. P. A century of skeletal biology and paleopathology: contrasts, contradictions, and conflicts. Am. Anthropol. 105:53–64, 2003.
6. Gould, S. J. Mismeasure of Man. Norton, New York, NY, 1996.
7. Thompson, D. D. Forensic anthropology. In: Spencer, F., ed., A History of American Physical Anthropology, 1930–1980. Academic, New York, NY, 357–369, 1982.
8. Hunt, D. R., Albanese, J. History and demographic composition of the Robert J. Terry anatomical collection. Am. J. Phys. Anthropol. 127:406–417, 2004.
9. Todd, T. W. Age changes in the pubic bone: I. The male White pubis. Am. J. Phys. Anthropol. 3:285–334, 1920.
10. Todd, T. W. Age changes in the pubic bone: II, the pubis of the male Negro-White hybrid; III the pubis of the White female; IV the pubis of the female Negro-White hybrid. Am. J. Phys. Anthropol. 4:1–70, 1921.

11. Todd, T. W., Tracy, B. Racial features in the American Negro. Am. J. Phys. Anthropol. 15:53–110, 1930.
12. Terry, R. J. The clavicle of the American Negro. Am. J. Phys. Anthropol. 16:351–379, 1932.
13. Washburn, S. The new physical anthropology. Trans. NY Acad. Sci. 13:298–304, 1951.
14. St. Hoyme, L. E, Işcan, MY. Determination of sex and race: accuracy and assumptions. In: Işcan, M. Y., Kennedy, K. A., eds., Reconstruction of Life from the Skeleton. Wiley-Liss, New York, NY, pp. 53–93, 1989.
15. Sauer, N. J. Forensic anthropology and the concept of race: if races don't exist, why are forensic anthropologists so good at identifying them? Soc. Sci. Med. 34:107–111, 1992.
16. Goodman, A. H. Bred in the bone? Sciences 37:20–25, 1997.
17. Gill, G. W. Craniofacial criteria in the skeletal attribution of race. In: Reichs, K., ed., Forensic Osteology. Charles C. Thomas, Springfield, IL, pp. 293–317, 1998.
18. Keita, S. O. Y., Kittles, R. A. The persistence of racial thinking and the myth of racial divergence. Am. Anthropol. 99:534–542, 1997.
19. Cartmill, M. The status of the race concept in physical anthropology. Am. Anthropol. 100:651–660, 1998.
20. Templeton, A. R. Human races: a genetic and evolutionary perspective. Am. Anthropol. 100:632–650, 1998.
21. Relethford, J. H. Global analysis of regional differences in craniometric diversity and population substructure. Hum. Biol. 73:629–636, 2001.
22. Brown, R. A, Armelagos, G. J. Apportionment of racial diversity: a review. Evol. Anthropol. 10:34–40, 2001.
23. Lieberman, L. How "Caucasoids" got such big crania and why they shrank—from Morton to Rushton. Curr. Anthropol. 42:69–95, 2001.
24. Albanese, J. The use of skeletal data for the study of secular change: methodological implication of combining data from different sources. Am. J. Phys. Anthropol. 34:36, 2002.
25. Relethford, J. H. Apportionment of global human genetic diversity based on craniometrics and skin color. Am. J. Phys. Anthropol. 118:393–398, 2002.
26. Molnar, S. Human Variation: Races, Types, and Ethnic Groups, 5th Ed. Prentice Hall, Upper Saddle River, NJ, 2002.
27. Livingstone, F. On the non-existence of human races. Curr. Anthropol. 3:279–281, 1962.
28. Lewontin, R. The apportionment of human diversity. Evol. Biol. 6:381–398, 1972.
29. Relethford, J. H. Craniometric variation among modern human populations. Am. J. Phys. Anthropol. 95:53–62, 1994.
30. France, D. L., Horn, A. D. Lab Manual and Workbook for Physical Anthropology. West, St. Paul, MN, 1988.
31. Francem D, L. Lab Manual and Workbook for Physical Anthropology, 4th Ed. Wadsworth Thomson Learning, Belmont, CA, 2001.
32. Ross, A. H., Slice, D. E., Ubelaker, D. H., Falsetti, A. B. Population affinities of 19th century Cuban crania: implications for identification criteria in South Florida Cuban Americans. J. Forensic Sci. 49:11–16, 2004.

33. Hooton, E. The Indians of Pecos pueblo: a study of their skeletal remains. Yale University Press, Andover, MA, 1930.
34. Snow, C. C., Hartman, S., Giles, E., Young, F. A. Sex and race determination of crania by calipers and computer: a test of the Giles and Elliot discriminant functions in 52 forensic science cases. J. Forensic Sci. 24:448–460, 1979.
35. Giles, E., Elliot, O. Race identification from cranial measurements. J. Forensic Sci. 7:147–157, 1962.
36. Işcan, M. Y. Race determination from the pelvis. OSSA 8:95–100, 1981.
37. DiBennardo, R., Taylor, J. V. Multiple discriminant function analysis of sex and race in the postcranial skeleton. Am. J. Phys. Anthropol. 61:305–314, 1983.
38. Işcan, M. Y., Cotton, T. S. The effect of age on the determination of race from the pelvis. J. Hum. Evol. 14:275–282, 1985.
39. Schulter-Ellis, F. P., Hayek, L. A. Predicting race and sex with an acetabulum/pubis index. Coll. Antropol. 8:155–162, 1984.
40. Ousley, S., Jantz, R. L. FORDISC 2.0: Computerized Discriminant Functions. University of Tennessee, Knoxville, TN, 1996.
41. Krogman, W., Işcan, M. Y. The Human Skeleton in Forensic Medicine. Charles C. Thomas, Springfield, IL, 1986.
42. Birkby, W. H. An evaluation of race and sex identification from cranial measurements. Am. J. Phys. Anthropol. 24:21–27, 1966.
43. Fisher, T. D., Gill, G. W. Application of the Giles and Elliot discriminant function formulae to a cranial sample of Northwest Plains Indians. In: Gill, G. W., Rhine, S., eds., Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, pp. 59–64, 1990.
44. Ayers, H. G., Jantz, R. L., Moore-Jansen, P. H. Giles and Elliot race discriminant functions revisited: a test using recent forensic cases. In: Gill, G. W., Rhine, S., eds., Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, pp. 65–72, 1990.
45. Gill, G. W., Rhine, S. Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, 1990.
46. Ousley, S., Jantz, R. L. The forensic data bank: documenting skeletal trends in the United States. In: Reichs, K., ed., Forensic Osteology. Charles C. Thomas, Springfield, IL, pp. 441–458, 1998.
47. Billinger, M. S. Geography, Genetics and Generalizations: the Abandonment of "Race" in the Anthropological Study of Human Biological Variation. Masters Thesis, Carleton University, Ottawa, Ontario, 2000.
48. Ginter, J. Dealing with Unknowns in a Non-Population: the Skeletal Analysis of the Odd Fellows Series. Masters Thesis. University of Western Ontario, London, Ontario, 2001.
49. Ginter, J. Issues of skeletal identification: ancestry determination of the Odd Fellows series. Abstract. Can. Assoc. Forensic Sci. J. 35:41, 2002.
50. Işcan, M. Y., Cotton, T. S. Osteometric assessment of racial affinity from multiple sites in the postcranial skeleton. In: Gill, G. W., Rhine, S., eds., Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, pp. 83–91, 1990.

51. Terry, R. J. On measuring and photographing the cadaver. Am. J. Phys. Anthropol. 26:433–447, 1940.
52. Giles, E. Sex determination by discriminant function analysis of the mandible. Am. J. Phys. Anthropol. 22:129–135, 1964.
53. Howells, W. W. Cranial Variation in Man. Peabody Museum of Archaeology and Ethnology Papers. Harvard University Press, Cambridge, MA, 1973.
54. Howells, W. W. Skull Shapes and the Map. Harvard University Press, Cambridge, MA, 1989.
55. Ubelaker, D. H. Book review: Personal computer forensic discriminant functions. Int. J. Osteoarchaeol. 8:128–123, 1998.
56. Fukuzawa, S., Maish, A. Racial identification of Ontario Iroquoian crania using FORDISC 2.0. Can. Assoc. Forensic Sci. J. 30:167–168, 1997.
57. Belcher, R., Williams, F., Armelagos, G. J. Misidentification of Meroitic Nubians using FORDISC 2.0. Am. J. Phys. Anthropol. 34:S42, 2002.
58. Leathers, A., Edwards, J., Armelagos, G. J. Assessment of classification of crania using Fordisc 2.0: Nubian X-group test. Am. J. Phys. Anthropol. 34:S99–S100, 2002.
59. Snow, C. C. Murder most foul. Sciences 35:16–21, 1995.
60. Marino, E. A. A pilot study using the first cervical vertebra as an indicator of race. J. Forensic Sci. 42:1114–1118, 1997.
61. Smith, SL. Attribution of foot bones to sex and population groups. J. Forensic Sci. 42:186–195, 1997.
62. Byers, S. N., Churchill, S. E., Curran, B. Identification of Euro-Americans, Afro-Americans, and Amerindians from palatal dimensions. J. Forensic Sci. 42:3–9, 1997.
63. Duray, S. M., Morter, H. B., Smith, F. J. Morphological variation in cervical spinous processes: potential applications in the forensic identification of race from the skeleton. J. Forensic Sci. 44:937–944, 1999.
64. Holliday, T. W., Falsetti, A. B. A new method for discriminating African-American from European-American skeletons using postcranial osteometrics reflective of body shape. J. Forensic Sci. 44:926–930, 1999.
65. Trudell, M. B. Anterior femoral curvature revisited: race assessment from the femur. J. Forensic Sci. 44:700–707, 1999.
66. Albanese, J. Identified Skeletal Reference Collections and the Study of Human Variation. McMaster University, Hamilton, Ontario, 2003.
67. Albanese, J. A metric method for sex determination using the hipbone and the femur. J. Forensic Sci. 48:263–273, 2003.
68. Meadows Jantz, L., Jantz, R. L. Secular change in long bone length and proportion in the United States, 1800–1970. Am. J. Phys. Anthropol. 110:57–67, 1999.
69. Cobb, W. M. Race and runners. J. Health Phys. Ed. 7:1–9, 1936.
70. Brace, C. L. Region does not mean "race"—reality versus convention in forensic anthropology. J. Forensic Sci. 40:171–175, 1995.
71. Kennedy, K. A. But professor, why teach race identification if races don't exist? J. Forensic Sci. 40:797–800, 1995.

72. Hauser, G., De Stefano, G. F. Epigenetic Variants of the Human Skull. Schweizerbartsche, Stuttgart, 1989.
73. Brues, A. M. The once and future diagnosis of race. In: Gill. G. W., Rhine, S., eds., Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, pp. 1–8, 1990.
74. Ossenberg, N. S. Within and between race distances in population studies based on discrete traits of the human skull. Am. J. Phys. Anthropol. 45:701–716, 1976.
75. Saunders, S. R. Nonmetric skeletal variation. In: Işcan, M. Y., Kennedy, K. A., eds., Reconstruction of Life from the Skeleton. Liss, New York, NY, pp. 95–108, 1989.
76. Rhine, S. Nonmmetric skull racing. In: Gill, G. W., Rhine, S., eds., Skeletal Attribution of Race. Maxwell Museum of Anthropology, Albuquerque, NM, pp. 9–20, 1990.
77. Byers, S. N. Introduction to Forensic Anthropology. Allyn and Bacon, New York, NY, 2002.
78. Gill, G. W. Challenge on the frontier: discerning American Indians from Whites osteologically. J. Forensic Sci. 40:783–788, 1995.
79. Wijsman, E. M., Neves, W. A. The use of nonmetric variation in estimating human population admixture: A test case with Brazilian Blacks, Whites and mulattos. Am. J. Phys. Anthropol. 70:395–405, 1986.
80. Sjovold, T. Non-metrical divergence between skeletal populations. OSSA 1:S1–S133, 1977.
81. McGrath, J., Cheverud, J., Buikstra, J. Genetic correlations between sides and heritability of asymmetry for nonmetric traits in rhesus macaques on Cayo Santiago. Am. J. Phys. Anthropol. 64:401–411, 1984.
82. Hanihara, T., Ishida, H., Dodo, Y. Characterization of biological diversity through analysis of discrete cranial traits**.** Am. J. Phys. Anthropol. 121:241–251, 2003.
83. Grüneberg, H. Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. J. Genet. 51:95–114, 1952.
84. Rösing, F. W. Discreta of the human skeleton: a critical review. J. Hum. Evol. 13:319–323, 1984.
85. Scott, G. R., Turner, C. G. The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent Human Populations. Cambridge University Press, Cambridge, 1997.
86. Tyrrell, A. Skeletal non-metric traits and the assessment of inter- and intra-population diversity: past problems and future potential. In: Cox, M., Mays, S., eds., Human Osteology. Greenwich Medical Media, London, pp. 289–306, 2000.

86a. Dahlberg, A. A. 1951. The dentition of the American Indian. In Laughlin, W. Ed. (ed.): Papers on the Physical Anthropology of the American Indians. Viking Fund, New York: pp. 138–176.

86b. Ossenberg, N. S. 1969. Discontinuous Morphological Variation of the Human Cranium. Ph.D. thesis, University of Toronto.

8bc. Ossenberg, N. S. 1981. An argument for the use of total side frequencies of bilateral nonmetric skeletal traits in population distance analysis — the regression of symmetry on incidence. Am. J. Physical Anthropol. 54:471–479.

87. Public Safety and Emergency Preparedness Canada. http://www.psepc-sppcc.gc.ca/index_e.asp. Last accessed Mar. 7, 2006.
88. Işcan, M. Y. Assessment of race from the pelvis. Am. J. Phys. Anthropol. 62:205–208, 1983.
89. Işcan, M. Y., Cotton, T. S. The effects of age on the determination of race. Coll. Antropol. 8:131–138, 1984.
90. Rogers, SL. Personal Identification from Human Remains. Charles C. Thomas, Springfield, IL, 1987.
91. Jackson, F. L. C. Anthropological measurement: the mismeasure of African Americans. Ann Am. Acad. Pol. Soc. Sci. 568:154–171, 2000.
92. Walker, P. L. Problems of preservation and sexism in sexing: some lessons from historical collections and palaeodemographers. In: Saunders, S. R., Herring, A., eds., Grave Reflections: Portraying the Past Through Cemetery Studies. Canadian Scholars Press, Toronto, pp. 31–47, 1995.
93. Boas, F. Changes in bodily form of descendants of immigrants. Am. Anthropol. 14:530–562, 1912.
94. Sparks, C. S., Jantz, R. L. A reassessment of human cranial plasticity: Boas revisited. Proc. Natl. Acad. Sci. USA. 99:14,636–14,639, 2002.
95. Gravlee, C. C., Bernard, H. R., Leonard, W. R. Boas's changes in bodily form: the immigrant study, cranial plasticity, and Boas's physical anthropology. Am. Anthropol. 105:326–332, 2003.
96. Gravlee, C. C., Bernard, H. R., Leonard, W. R. Heredity, environment, and cranial form: a reanalysis of Boas's immigrant data. Am. Anthropol. 105:125–138, 2003.
97. Sparks, C. S., Jantz, R. L. Changing times, changing faces: Franz Boas's immigrant study in modern perspective. Am. Anthropol. 105:333–337, 2003.
98. Olivier, G. Practical Anthropology. Charles C. Thomas, Springfield, IL, 1969.
99. Rocha, A. Les collections ostéologiques humaines identifiées du Musée Anthropologigue de l'Université de Coimbra [in Portuguese]. Antropol. Port. 13:7–38, 1995.
100. Cunha, E. Testing identification records: evidence from the Coimbra Identified Skeletal Collections (nineteenth and twentieth centuries). In: Saunders, S. R., Herring, A, eds., Grave Reflections: Portraying the Past Through Cemetery Studies. Canadian Scholars Press, Toronto, pp. 179–198, 1995.
101. Skinner, M., Lazenby, R. A. Found! Human Remains. Archaeology Press, Simon Fraser University, Burnaby, British Columbia, 1983.
102. Santos, A. L. A Skeletal Picture of Tuberculosis. Ph.D. Thesis. University of Coimbra, Coimbra, Portugal, 2000.
103. Trotter, M. Robert James Terry, 1871–1966. Am. J. Phys. Anthropol. 56:503–508, 1981.
104. Tague, R. G. Maternal mortality or prolonged growth: age at death and pelvic size in three prehistoric Amerindian populations. Am. J. Phys. Anthropol. 95:27–40, 1994.
105. Fuller, K. Adult females and pubic bone growth. Am. J. Phys. Anthropol. 106:323–328, 1998.

106. Stini, W. A. Adaptive strategies of human populations under nutritional stress. In: Watts, E. S., Johnston, F. E., Lasker, G. W., eds., Population Adaptation. Mouton, The Hague, pp. 19–41, 1975.
107. Stini, W. A. Sexual dimorphism and nutrient reserves. In: Hall, R. L, ed., Sexual Dimorphism in *Homo sapiens*. Praeger, New York, NY, pp. 391–419, 1982.
108. Stinson, S. Sex differences in environmental sensitivity during growth and development. Yearb. Phys. Anthropol. 28:123–147, 1985.
109. Norušis, M. J. SPSS 12.0 Statistical Procedures Companion. Prentice Hall, Upper Saddle River, NJ, 2003.
110. Shriver, M., Kittles, R. Genetic ancestry and the search for personalized genetic histories. Nat. Rev. Genet 5:611–618, 2004.
111. Duster, T. Buried alive: the concept of race in science. In: Goodman, A. H., Heath, D., Lindee, M. S., eds., Genetic Nature/Culture. University of California Press, Berkeley, CA, pp. 258–277, 2003.
112. Rohl, A., Brinkmann, B., Forster, L., Forster, P. An annotated mtDNA database. Int. J. Legal Med. 115:29–39, 2001.
113. Forster, P. To err is human. Ann. Hum. Genet. 67(Part 1):2–4, 2003.
114. Budowle, B., Allard, M. W., Wilson, M. R., Chakraborty, R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu. Rev. Genomics Hum. Genet. 4:119–141, 2003.
115. Yao, Y. G., Bravi, C. M., Bandelt, H. J. A call for mtDNA data quality control in forensic science. Forensic Sci. Int. 141:1–6, 2004.