IDENTIFIED SKELETAL REFERENCE COLLECTIONS AND THE STUDY OF HUMAN VARIATION

By

JOHN ALBANESE, H.B.A., M.SC.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

IDENTIFIED SKELETAL REFERENCE COLLECTIONS

Doctor of Philosophy (2003)
(Physical Anthropology)

McMaster University
Hamilton, Ontario

TITLE: Identified Skeletal Reference Collections and the Study of Human Variation

AUTHOR: John (Giovanni) Albanese, H.BA., M.Sc.

SUPERVISOR: Dr. S. R. Saunders

NUMBER OF PAGES: xv, 269

Abstract

A comprehensive model that builds on cemetery studies theory and new biocultural synthesis theory is presented for investigating human variation using reference collections. This model is used to investigate several hypotheses related to the use of skeletal reference collections and the race concept in skeletal biology. Are racial categories or equivalent terms useful for investigating human variation? Have biases in reference collections resulted in a misinterpretation of human variation? Is it possible to identify and control for some of the biases in reference collections and develop identification methods that are still useful in the 21st century?

Documentary, historical and skeletal data were collected from two different reference collections: the Terry and Coimbra Collections. Some data from the Forensic Anthropology Data Bank (FDB) were also used in conjunction with these two collections in analyses related to patterns of sexual dimorphism, sex determination methods, and assessment of secular changes using skeletal data.

A critical review of the methodology for the study of secular change in femur length illustrates that there are methodological costs of ignoring bias in reference collections. In a combined sample from several collections, what appear to be significant secular changes occur at the points of transition between data sources. Dividing samples into racial sub-samples does not control for so-called population differences and only further confounds the analysis of secular changes using reference collections.

An alternative methodology was followed for the development of a series of metric sex determination methods with allocation accuracies of 90-98.5%. Logistic regression was used to develop a series of models that use a new alternative measurement of the pubis along with traditional measurements of the hipbone and femur. Demographic and historical data were used to select a reference sample that included a wide range of human variation. Testing suggests that these metric models are widely applicable.

Various skeletal, demographic, and historical analyses indicated that racial categories or equivalent terms are not useful for investigating human variation using reference collections, and that some of the biases in reference collections have resulted in a mis-interpretation of human variation in past research. It is possible to identify and control for some of the biases in reference collections and develop identification methods that are still useful in the 21st century.

Acknowledgments

This research was supported by a doctoral fellowship from the Social Science and Humanities Research Council of Canada (grant # 752-99-1267), by an Ontario Graduate Scholarship and funds from the School of Graduate Studies, McMaster University. It was possible to present some preliminary results of this research at several conferences because of the travel assistance provided by the Graduate Students Association at McMaster University. I would like to thank the administration at the Museum of Anthropology at the University of Coimbra in Portugal for allowing access to the Coimbra Collection and the curators of the Forensic Anthropology Data Bank at the University of Tennessee, Knoxville for data.

There are a few people who I would like to thank. First, I would like to thank Dr. Shelley Saunders, my thesis supervisor, for assistance and support throughout the dissertation and the Ph. D. program. I would also like to thank the members of my thesis committee Dr. Ann Herring and Dr. Michael Spence. In Portugal, I would like to thank Dr. Eugenia Cunha, Dr. Ana Luísa Santos, and Ana Maria Silva. In Washington, I would I like to thank Dr. David Hunt for his friendship and assistance; and Troy Case who made the endless hours of data collection at the Smithsonian Institution even longer but bearable. For his help and friendship (and for a whirlwind-one-day-tour of Lisbon), I would like to thank Hugo Cardoso. Hugo is responsible for the illustration of the hipbone in Chapter 6. Thanks for everything Hugo. I would like to thank my Mom and Dad for their encouragement to attend graduate school and their help in so many ways. Most of all I want to thank Sandra. The technical support with the database management was a big help in many ways, but the technical help was secondary to the unwavering support through the good times and during those not-so-good times. Sandra, I could not have done it without you.

This thesis is dedicated to the memory of Alfredo Maggiacomo. We miss you.

Preface

NOTES REGARDING AUTHORSHIP AND COPYRIGHT

Chapter 3, "The History and Demographic Profile of the Robert J. Terry Anatomical Collection," was co-authored with Dr. David Hunt[1] and is the only paper in this thesis that was co-authored. The paper grew out of two separate research projects. Dr. Hunt was already working on a version of this paper when I arrived at the National Museum of Natural History (NMNH) to begin collecting data in January, 2000. He had sent me an early version of the paper (which he presented as a poster at the 68th Annual Meeting of the American Association of Physical Anthropologists held in Columbus, Ohio, April 26 to May 1, 1999) when I first contacted him regarding information about the Terry Collection. This early version of the paper provided some historical context for the collection and briefly listed some of the items of research interest that were associated with the collection (photos, slides, etc). Dr. Hunt's goal was to publish an article that prospective researchers could refer to as they developed research projects involving the Terry Collection.

When I arrived at the NMNH I had two goals. First, I wanted to reconstruct how the collection had been amassed because I was interested in determining how the process may have affected the demographic profile of the collection, the skeletal variation in the collection, and how previous researchers may have interpreted this variation. Second, I wanted to assess the quality and accuracy of the documentary records. When we realized how complementary our research was, we decided to co-author a paper on the history of the Terry Collection. We used the earlier, shorter version of Dr. Hunt's paper as a detailed outline for the paper that now appears in Chapter 4. After multiple re-writes, we both contributed to the data collection, analysis and writing of every section in the paper. We did not contribute equally to each section but the paper could not have been possible without both of our efforts in each and every section. For example, the section "The Spread of an Idea" has its roots in a section of Dr Hunt's original paper where he drew on his firsthand experience with the Huntington Collection and briefly described the importance of Dr. Huntington's influence on Drs. Terry, Todd and Hrdlička. I drew on several historical references to support his work and found additional information to demonstrate how this concept for a reference collection spread to other collectors. In another example, to test for the quality and accuracy of the reference data, I conducted a detailed review of the records of 356 individuals from the collection and cross-referenced the information with documents that Dr. Hunt had located at Washington University. Although I did all of the analysis for this part of the paper, it could not have been as thorough a review without the data that Dr. Hunt had already collected. This paper is in the final stages of preparation for submission for publication. Because of the differences in the reasons for our interests in writing the paper and some of the possible political implications of some of the research in this thesis, I want to stress that the research presented in the rest of the thesis is entirely mine and does not necessarily reflect Dr. Hunt's views or research interests. However, I cannot state enough times that I am very grateful for all his assistance with my research on the Terry Collection while I was in Washington and after my return to Toronto.

Various topics in Chapter 6 have been presented under the titles "A New Metric Method for Sex Determination Using the Pelvis and Femur," and "Measurement Error and Metric Methods for Sex Determination." The first paper was presented at the 70th Annual Meeting of the American Association of Physical Anthropologists held in Kansas City, Missouri, March 28-31, 2001. The abstract of this paper was published in 2001 in the *American Journal of Physical Anthropology* Supplement 32:31. The second paper was presented at the 29th Annual Meeting of the Canadian Association for Physical Anthropology in Winnipeg, Manitoba, October 25-27, 2001. The version of the paper that appears in this thesis has been published in *The Journal of Forensic Sciences* Mar. 2003 Vol 48(2):263-273, Paper ID JFS 2001378_482, copyright ASTM International, 100 Barr Harbor Drive, West Conshohocken, PA 19428, and may not be republished elsewhere without the expressed written permission of ASTM.

John Albanese
Toronto, 2003

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

For almost a century, there has been an ongoing debate in physical anthropology regarding the relevance of race when investigating human variation. This debate predates the emergence of physical anthropology as a separate discipline in North America and has persisted throughout the discipline's existence (Brace 1982, Armelagos and Goodman 1998, Caspari 2003, Gravlee et al 2003). Boas (1912, 1940) has been credited as an early mainstream anthropologist who challenged the biological basis for race and the utility of race in investigating human variation (Armelagos and Goodman 1998, Caspari 2003, Gravlee et al 2003). Although Boas' (1912) work on plasticity of the human cranium has been recently critiqued (Sparks and Jantz 2002), a large body of evidence has been published since Boas's early work that clearly demonstrates that genetic and phenotypic variation does not correspond with racial categories (Goodman 1995, Keita and Kittles 1997, Templeton 1998, Cartmill 1998, Armelagos and Goodman 1998, Lieberman 2001, Brown and Armelagos 2001, Relethford 2001, Molnar 2002, Relethford 2002). Racial categories account for only about 6-13% of genetic and craniometric variation. Still, racial categories or equivalents continue to be used extensively for investigations of human variation in the second half of the 20th century (Cartmill 1998) and into the beginning of the 21st century. While some researchers (for example, Nei and Roychouldhury 1982, Ryman et al 1983, Gill 1990, Hartmann et al, 1994) have continued to try to establish racial categories as biologically meaningful categories (Brown and Armelagos 2001), others continue to use racial categories or continental/geographic origin as a research tool for investigating human variation (for example, Ousley and Jantz 1998, Meadows Jantz and Jantz 1999, Hennessy and Stringer 2002).

Specifically for skeletal biological and forensic anthropology, racial categories have been used extensively for investigating human variation and for establishing parameters for the applicability of forensic identification methods using several key reference collections such as the Hamann-Todd and Terry Collections (for example, İşcan et al, 1985, 1987; İşcan and Miller-Shaivitz 1984, Holland 1991, Holman and Bennett 1991, Marino 1995, Ousley and Jantz 1996, Smith 1997). In this thesis, the term reference collection is used to describe a human skeletal collection that was amassed for general research and anatomical instruction. Examples include the Hamann-Todd Collection, the Terry Collection, the Cobb Collection, and the Grant Collection which were derived from cadavers that were used for anatomical instruction. These reference collections contrast with archaeological cemetery collections because of the manner in which they were amassed, the reasons why they were amassed, and the research questions that can be addressed and have been addressed using the collections. Not all reference collections have been derived from anatomical sources. The Forensic Anthropology Data Bank (Ousley and Jantz 1998) can be considered a virtual reference collection. Additionally, there are several examples of reference collections, such as the Coimbra

Identified Skeletal Collection (Santos 2000) and Luís Lopes Collection (Cardoso 2000), which are reference collections derived from cemeteries. Both collections are very different from archaeological cemetery collections, such as the St. Bride's Church Crypt Collection (Scheuer and Bowman 1995), that have at times been used as reference collections because some of the individuals in the collection are positively identified. The Coimbra Collection has been carefully constructed as a reference collection and the Lopes Collection continues to expand with a focus on the collection of sub-adults (Hugo Cardoso, personal communication).

The purpose of this thesis is to test several hypotheses related to the use of reference collections and the race concept in skeletal biology. First, are racial categories or equivalent terms useful for investigating human variation? Second, have biases introduced in the collection process resulted in a mis-interpretation of human variation, specifically the mis-apportionment of variation in the Terry Collection and the FDB to race? Third, is it possible to identify and control for some of the biases in the Terry Collection and other reference collections in order to maximize their research potential and develop identification methods that are still useful in the 21st century?

The theoretical framework for this research is drawn from two complementary bodies of theory: cemeteries studies theory (Saunders and Herring 1995a) and the theory described as the new biocultural synthesis (NBS) (Goodman and Leatherman 1998a). In many ways reference collections are similar to highly specialized archaeological cemetery collections, such as crypt samples, and cemetery studies theory has a great deal of explanatory power for assessing and controlling for potential sources of bias. The emphasis on the use of a combination of skeletal and documentary data has some clear benefits (Saunders and Herring 1995b, Saunders et al 1995) and is transferable to research involving reference collections (for example, Cunha 1995). There are five key interconnected issues that are part of the NBS (Goodman and Leatherman 1998a) and all can be relevant when conducting research involving reference collections. However, two issues are particularly useful when assessing sources of bias in reference collections: an emphasis on historical context and considering the ideology, knowledge, and perspectives of the researchers. The history of the American reference collections, that have been used to develop or test many identification methods, is very closely connected to the history of physical anthropology in the United States and specific individuals who directed the additions to those collections.

In past research, the application of the two bodies of theory has resulted in more rigorous research and more certain conclusions (see Saunders and Herring 1995a, Goodman and Leatherman 1998a). By applying a combination of these approaches for assessing the effects of bias on reference collections, a goal of this thesis is to have a similar constructive impact on the development of identification methods that can be applied with greater confidence to a wider range of forensic and archaeological cases.

Although this proposed comprehensive approach can likely be applied to a number of different research questions involving reference collections, this study focuses on skeletal sexual dimorphism, metric sex determination methods, and the assessment of secular change using skeletal data. Two different but complementary reference collections are used extensively in this research: the Terry Collection and the Coimbra Collection.

These collections were selected because they have both been extensively used to investigate a number of different research questions but on opposite sides of the Atlantic Ocean. Additionally, the source of skeletons for each collection is very different. Supplemental data from the Forensic Anthropology Data Bank (FDB) are also used in some of the analyses. The FDB is emerging as an important virtual reference collection in an era when new, large reference collections cannot be amassed. However, the FDB, as with all skeletal collections, is likely not free of bias. Following the methodology used for research involving cemeteries, a combination of documentary evidence, historical data and skeletal data are used in this research.

In Chapter 2, some of the relevant literature is reviewed. Topics that are reviewed include cemeteries studies theory (Saunders and Herring 1995a), the new biocultural synthesis theory (Goodman and Leatherman 1998a), the use of the race concept in physical anthropology, the history of the methodology for developing age determination methods, and the history of the methodology for developing metric sex determination methods. In the latter two reviews, the emphasis is on the effects of the race concept on methodology.

Chapter 3 is a broad, descriptive review of the history and demographic profile of the Terry Collection. There are two major complementary goals in Chapter 3. The first objective is to provide a detailed description of the collection in a single document that researchers can refer to when developing future research projects involving the Terry Collection. The second objective is to describe the process that resulted in the collection in its current form. Chapter 3 is an application of the combined cemetery studies and NBS theory to identify some of the biases in the collection, and their effects on the demographic structure and skeletal variation in the collection.

Chapter 3 is divided into 4 major sections. In the first section Dr. Terry's interest in collecting a large well documented collection is placed in the context of other anatomists who were collecting skeletons at the end of the 19^{th} century and the first half of the 20^{th} century, and Mildred Trotter's influence on the collection is also explored in some detail. In the second section, the documents and other research materials that are available for the collection are described and the basic demographic profile of the collection is presented. The research value of any reference collection is highly dependent on the quality of the documentary data. Therefore, section three is devoted to reviewing the quality of the documentary data available for any one individual in the collection and for the Terry Collection as a whole. In the last section of Chapter 3, the representativeness of the Terry Collection is assessed and some general guidelines are provided for maximizing the research potential of the collection in future research by minimizing the bias in reference samples.

Chapter 4 has three major sections. In the first section, the history and demography of the Coimbra Collection is briefly reviewed. In the second section a proposed comprehensive model is presented for assessing and controlling biases in samples selected from reference collections. The last section is a description of the application of this model for selecting samples used in this research from the Terry and Coimbra Collections to investigate sources of variation: age at death, year of birth, collection process, and historical and legal issues.

The preliminary analysis that eventually led to Chapter 5 was an attempt to reproduce various studies that identified secular change using skeletal data (Trotter and Gleser 1951, Angel 1976 and Meadows Jantz and Jantz 1999). Initially the goal was to identify, quantify, and possibly control for the effects of secular change on sexual dimorphism in the Terry Collection and the Coimbra Collection, and to determine if older reference collections are still useful for developing new metric methods for sex determination. In other words, the initial goal was to identify bias in the reference collections associated with secular change in stature. Instead, other sources of bias in the reference collections and the research methodology were identified.

Others have described the social costs and political implications of the continued use of racial groups in the study of human variation (Keita and Kittles 1997, Armelagos and Goodman 1998, Lieberman 2001). In Chapter 5, some of the methodological costs of ignoring the biases in the reference collections and research methodology are illustrated. In the original analysis (Albanese 2000) only data from the Terry and Coimbra Collections were used. In the paper in Chapter 5, the analysis was expanded and data from both collections were used along with data from the FDB, and more systematic, rigorous and robust statistical tests were applied. The pattern of results from both analyses are identical. Results may be seriously compromised when samples from different data sources (Terry Collection, Coimbra Collection, FDB, Huntington Collection, etc.) that do not represent the same population are combined for the study of secular change. Additionally, dividing a sample by "race" does not control for any of these differences between samples and defining a sample or population using racial categories can produce very misleading results. The last issue addressed in Chapter 5 is the effects of variations in sex ratios in a sample when assessing secular changes.

In Chapter 6, some of the theoretical issues discussed in Chapter 2 are applied in the development of a new general methodology for the development of sex determination methods and a new series of specific methods for sex determination. Considering many of these issues has resulted in a promising new metric method for sex determination. Whereas Chapter 5 is a detailed illustration of some of the shortcomings of not assessing the effects of biases at different levels, the approach presented in Chapter 6 is a concrete example of the benefits of a proposed approach for assessing levels of representativeness of reference collections. The high allocation accuracy of the new method can be attributed to a major departure from the established methodology for developing sex determination methods. First, using the wealth of historical and demographic data available for the various reference collections, it was possible to identify and avoid racial assumptions about human variation. Underlying sources of variation were sampled instead, including year of birth and age at death. By using this approach, it is possible to identify the source of the variation and maximize the representativeness of both the Coimbra Collection and the Terry Collection in carefully constructed reference samples that sample a wide range of human variation.

Second, logistic regression, and not discriminant function analysis, is used to develop the various models. With a focus on sex determination, the more robust logistic regression that is specifically designed for situations with a binary dependant variable (female/male) is a better choice for analysis (Saunders and Hoppa 1997). Third, data from the pelvis and the femur are considered together to determine sex. Fourth, a new measurement of the pubis (SPRL) was tested and used that has substantially less intra-observer measurement error than the traditional pubis measurement.

Throughout this thesis, the use of quotation marks around the terms race, White, Black, Negro, etc, may seem irregular and rather confused at first glance. Racial categorization is a social reality in many countries. At the same time, there is a great deal of evidence that the race concept does not apply to humans, and racial categories are not useful groups for investigating human biological variation. Thus, racial terms appear both in quotation marks or not depending on the context of use. Quotation marks are not used when referring to the social constructs of race. In the context of this type of discussion, racial concepts are a social reality. There is a great deal of self-reporting of race in numerous situations including official censuses and in these cases Black, White, etc, refers to a social categorization and has nothing to do with biological variation. Additionally, quotation marks are not used when describing the history of the race concept. The race concept is still a concept regardless of some of the reported shortcomings when applied to the study of human variation. Quotation marks are used around racial terms when referring to the designation of any one individual or group from any of the reference collections, or in cases where in the past these terms have been used to refer (directly or indirectly) to phenotypic or genotypic variation. The quotation marks are used to highlight that the racial category is not a meaningful description of variation for the individual or the group. For example, the use of the description "White" males for a sample from the Terry Collection illustrates that the term male has some genetic, biological and phenotypic significance, and “White” does not reveal anything about the pattern of skeletal variation in any given individual or the group as a whole.

CHAPTER 2

Literature Review and Historical Context

INTRODUCTION

The prospects for the collection of new, large skeletal collections in North America are not good. There have been some efforts at collecting identified skeletons at some institutions. For example, a collection is curated at the Department of Anthropology at the University of Tennessee, Knoxville that numbers about 200 individuals (Ousley and Jantz 1998). However, collection protocols such as those that lead to the establishment of the Terry or Hamann-Todd Collections that number thousands of documented individuals is almost completely out of reach in North America. There are two significant barriers to such a collection: a close affiliation with a source of skeletal remains (such as an anatomy department), and substantial multi-year funding for processing, cataloguing, and storage of the skeletal remains *and* the supporting documentary information. The skeletons are of limited value without substantial resources devoted to the accurate documentation for each individual and the collection as a whole.

This combination of factors means that a large skeletal collection is not possible. Even if funding could be secured, access to the source of a large number of skeletons that were dissected in a manner that would preserve most bones without damage would be difficult. All the reference collections that are now affiliated with anthropology departments at universities or anthropology divisions at various institutions in North America were collected in anatomy departments. These collections are no longer curated in anatomy departments because collecting and curating skeletons is not a priority and stopped being a research interest 40 years ago or more in most departments of anatomy in North America. Key individuals, Drs Terry, Hamann, Todd, Cobb, Grant, were able to collect skeletons because they were the chairs of their respective anatomy departments and could devote significant resources to collecting skeletons at a time when almost all physical anthropologists were anatomists. There are many examples of collections that follow this historical pattern including the Grant Collection at the University of Toronto, the Cobb Collection at Howard University (Rankin-Hill and Blakey 1994), University of Iowa-Stanford Collection (Robert Franciscus, personal communication), the Stanford-Meyer Osteopathology Collection at the San Diego Museum of Man, (Rose Tyson, personal communication) [1], the Huntington Collection at the Smithsonian Institution's

---

1 Dr. Arthur W. Meyer, Stanford University, followed a pattern of collecting that is similar to other collectors (Terry, Hamann, Todd, Grant, Huntington). A series of isolated bones with gross pathological conditions were mounted and curated for instructional purposes and a series of complete skeletons with some documentation (age at death, sex, etc) were also collected. In some cases, such as Meyer's collections, the pathology collection is now curated at one institution while the documented skeletons are at another. The Stanford-Meyer Osteopathology Collection was transferred to the San Diego Museum of Man in 1981 (Rose Tyson, personal communication; Rose Tyson is the curator of physical anthropology at the San Diego Museum of Man) and the University of Iowa-Stanford Collection was transferred to the Department of Anthropology at the University of Iowa in 1998 (Robert Franciscus, personal communication; Robert

National Museum of Natural History, the Terry Collection at the Smithsonian Institution's National Museum of Natural History, and the Hamann-Todd Collection at Cleveland Museum of Natural History.

As the prospects for new research materials diminished in the last three decades of the 20th century, the research value of the older anatomical collections was increasingly questioned. The major reference collections have been described as not representative of the populations in the U.S.A because of the source of the collection: cadavers for anatomical instruction (Giles 1964). More recently, others have suggested that these collections may no longer be useful for the development of forensic identification methods because of the source of the skeletons and because of the age of the collections (Ousley and Jantz 1998). The majority of individuals in all the major anatomical collections have birth years in the 19th century and the collections may not reflect some of the major secular changes in the U.S.A in the 20th century (Ousley and Jantz 1998). While there is a general consensus that these collections are biased, the nature of the bias and effects of the bias on the pattern of variation in the collections have not been investigated in any detail. Only Ericksen (1982), has directly attempted to address the issue of representativeness in the Terry Collection. Due to some limitations in sample size and other issues, firm statements about the representativeness of the Terry Collection were not possible (Ericksen 1982). Despite some of the problems with the reference collections, all the major anatomical collections continue to be used extensively to address a number of

---

Franciscus is an assistant professor in the Department of Anthropology, University of Iowa).

different research questions including the development of forensic identification methods.

New, large identified skeletal collections would be wonderful sources of data for many different types of research besides the development of identification methods; however, even if a collection could be amassed, it cannot be assumed that newer (real or virtual) collections would be free of bias. There are some benefits to investigating alternative theoretical approaches to research involving all reference collections including anatomical collections (Terry Collection), databases (FDB), and cemetery-derived reference collections (Coimbra Collection) where it may be possible to identify the bias, control the bias, and even exploit the bias to address specific research questions.

The two main sections of this chapter are devoted to reviewing some of the issues relevant to research involving reference collections, and to exploring some alternative approaches to research involving these collections. In the first section, two complementary bodies of theory (Saunders and Herring 1995a, Goodman and Leatherman 1998a) are reviewed and their relevance to research involving reference collections is described. The race concept has had a huge influence on physical anthropology in the 20th century (Washburn 1963, Brace 1982, Aremalgos and Goodman 1998, Caspari 2003). The race concept has had and continues to have an influence on research involving reference collections, and this concept also had an influence on the collection process and the demographic structure of the reference collections. Therefore, the second major section of this chapter includes a review of the current status of the race concept in physical anthropology (with an emphasis on skeletal biology); a reconstruction of the historical context in which the Terry and Hamann-Todd Collections were amassed; and two separate reviews of the influence of the race concept on the methodology for developing (1) age determination methods and (2) sex determination methods.

CONSIDERING OLD COLLECTIONS IN A NEW THEORETICAL FRAMEWORK

Cemetery Studies Theory and Methods

At the 62nd Annual Meeting of the American Association of Physical Anthropologists in 1993, a symposium was held entitled, *Representativeness: The Comparison of Skeletons and Records from Historic Cemetery Studies*. At the symposium, which was later published as an edited volume under the title *Grave Reflections* (Saunders and Herring 1995a), several core issues that are critical to skeletal research were addressed: "How well do mortality samples represent the populations from which they were drawn? Can we make accurate reconstructions of the people who lived in the past from samples of their dead?" (Saunders and Herring 1995b:1).

The excavation of church cemeteries, hospital graveyards, and battlefield burial grounds from the historic period in Europe and North America have provided opportunities to test and evaluate the reconstruction of past populations through the comparison of skeletal data with the available documentary data for those populations (Saunders and Herring 1995b). Various examples from Europe and North America published in the volume illustrate that representativeness of skeletal samples from cemeteries can be assessed at several levels: is the recovered and analyzable sample representative of the entire cemetery and does the cemetery represent greater levels of biological or social units such as a parish, local community, or region (Saunders and Herring 1995b)? The editors of the volume conclude that,

> the studies in this volume suggest that we can be cautiously optimistic about making statements concerning the health and biological characteristics of specific cemetery samples that skeletal collections come from if those collections are sufficiently large and in good condition. Alternatively, skeletal samples often reflect the uniqueness of the sample of individuals who were *interred in a site*, sometimes under very unusual circumstances. In many instances, it is unlikely that mortality samples will accurately represent broader population groups or communities simply because there are probably a multiplicity of communities that existed in the past whose composition fluctuated over time. This is certainly true of many historic period groups who lived when population size had become larger and more complex. Equally sobering is the distinct possibility that certain biological characteristics underwent secular change, which makes the deciphering of features, such as skeletal age changes, particularly difficult. (Saunders and Herring 1995b:2; emphasis added).

Although these authors are describing the reconstruction of past human populations, if the phrase "interred in a site" (italicized) is replaced with "included in a reference collection," they may well be describing many of the issues associated with reference collections including the Terry Collection, the Hamann-Todd Collection and the FDB. If identified reference collections are considered in the same theoretical framework as cemetery collections, then the bad news is clearly that all reference collections are biased. However, there are some clear benefits of considering reference collections in this theoretical framework,

> as this and other papers in this volume show, all skeletal samples will probably contain some kind of bias of composition. *Identifying these biases will help researchers*

> *overcome many obstacles to their analyses*, and the judicious use of complementary historical documents, when available, are invaluable to this process (Saunders et al 1995:110-111; emphasis added).

With very few exceptions (for example, Trotter and Gleser 1951, Angel 1976, Meadows Jantz and Jantz 1999) reference collections are usually not directly used to reconstruct the health and biology of the populations they were derived from; however, there are several aspects of this theoretical framework that are particularly applicable to research involving reference collections. First, it is possible to assess the nature and source of the bias but only if skeletal and documentary data are considered together. Second, biased samples do not necessarily produce biased results. There are different levels of representativeness.

In *Grave Reflections*, there are multiple examples that illustrate approaches that can be used to assess bias at various levels through the use of documentary and skeletal data. A major source of bias in any cemetery or reference collection is filtering of the sample at a number of different levels in space and time which influences the analyzable sample. This bias is clearly illustrated by Molleson using an area graph of documentary and skeletal data from Christ's Church, London cemetery collection (1995:202, Figure 2; see also Hoppa 1996:52, Figure 3.1 for a general model illustrated graphically). Of the 68,000 individuals buried at Christ's Church, only 968 burials were excavated in the crypt of the church. Of these 968, only 389 were both positively identified through coffin plates and preserved well enough to be analyzed.

In another example, Scheuer and Bowman (1995) discuss some factors that may have affected the size and composition of an analyzable crypt sample using data from St.

Bride's Church crypt, London, but which are applicable to a number of different specialized burials. Included in these factors are the level of preservation of skeletal material and coffin plates, excavation procedures (a major problem for St. Bride's Church crypt), and socio-economic factors that resulted in burials in special locations (ie the crypt) versus other more common areas. Although the filters may be different when comparing the analyzable sample derived from a cemetery as compared to the analyzable sample derived from a skeletal reference sample, the end result is the same. Bias is introduced or magnified in the study sample. However, through the review of the documentary records available, whether they are church records for cemetery samples or morgue records for anatomical samples, the nature of the bias can often be assessed.

Within the theoretical framework of cemetery studies, there are different levels of representativeness of the samples (Saunders and Herring 1995b). Stated another way, bias does not necessarily mean that cemetery collections cannot be useful for investigating many different research questions. The assessment of the bias is the critical step in determining the level of representativeness of the sample. If reference samples are considered in the same theoretical framework as cemetery collections (since in some ways they are highly specialized mortality samples), then there are levels of representativeness in the reference collections and the effects of the bias will be different depending on the research questions being addressed. Having an understanding of the bias in the skeletal samples means that it is possible to assess whether or not confidence can be placed in any conclusions based on data from that sample. Continuing with the example above from Christ's Church, it is possible to assess the level of representativeness which is fairly restricted in this case because,

> the data compiled from the crypt sample as a whole did not show any unexpected significant differences from the data compiled for the named sample; nor is there any reason to suppose that the named sample is selective except by presence of the coffin plate. Whether the 968 crypt burials are representative of the additional 67,000 inhumations that were buried in the church yard cannot be evaluated since none of the latter is available for study (Molleson 1995:200).

In other cases, when other sources of documentary data are available, it is possible to assess representativeness at additional levels. For example, Scheuer and Bowman (1995) used skeletal data and documentary data from St. Bride's parish, London, to reconstruct mortality age profiles. They compared the skeletal collection from the crypt (n = 131), documentary data for the entire crypt sample (n = 312) and documentary data available for 99% of all burials at St. Bride's parish (n = 4477). Furthermore, the entire parish was compared to a modern profile for England and Wales.

They found that the skeletal collection is representative of the crypt except for child deaths under 10 years of age. However, the entire crypt sample is not representative of the parish of St. Bride. When the entire cemetery profile, the crypt profile, and the modern profile for England and Wales were compared, the crypt sample followed a pattern between the other two profiles. By making such comparisons, Scheuer and Bowman found that,

> the differences between the skeletal collection and the rest of the people interred in the crypt are due to factors that are probably logistical and connected with unknown details of excavation. The much bigger differences between the crypt and the cemetery population [are] more likely due to the gap in socioeconomic status of the two groups (1995:59).

The skeletal collection from the crypt at St. Bride's Church may not be representative of the mortality in the entire parish; however, the collection may still be used to address many different research questions. If the St. Bride's crypt collection was derived from a narrow segment of the parish, then the collection is an invaluable resource for research questions related to, for example, morbidity and mortality in a more privileged segment of a parish in London in the 18th and 19th century.

The last major aspect of the theoretical framework of cemetery studies that is directly applicable to skeletal reference collections is related to sampling approaches. Once the bias is identified, it is not only possible to assess the level of representativeness of the collection, it is also possible to sample the skeletal collection in order to minimize or control the effects of the bias. Furthermore, it is also possible to sample the collection so that the bias may actually be exploited to maximize the research potential of the collection to address specific research questions. In one example of this approach, Cunha (1995) used the Coimbra Collection to investigation skeletal stress indicators.

The Coimbra Collection is both a cemetery collection and a reference collection that has been used to develop forensic identification methods and to address other research questions (Cunha 1995, Rocha 1995, Santos 2000). Several biases become obvious after a preliminary review of the documentary resources that are available for the Coimbra Collection (Cunha 1995). First, the collection is a sample of individuals from lower socio-economic classes, and second, a *relatively* restricted geopolitical area is sampled in the collection. Cunha (1995) describes the collection as heterogenous since there are individuals in the collection who were born throughout Portugal and some

Portuguese colonies in Africa and South America. However, the most frequently listed place of birth is the District of Coimbra, and 68% of individuals in the collection were born in the District of Coimbra or the five districts that share a border with Coimbra. While these might *seem* to be insurmountable biases for the development of forensic identification methods (but see Chapters 6), they can be ideal conditions "to provide important information about the etiology and validity of stress indicators" (Cunha 1995:179, see also Santos 2000). Cunha made these conditions even more ideal by exploiting the biases of the collection and sampling only those adults from the collection that were actually born in the District of Coimbra in order to control for social, economic and political differences between districts.

The New Biocultural Synthesis

When the focus of research is on reference collections, in contrast to cemetery collections, cemeteries studies theory is complemented by some of the theoretical issues described in *Building a New Biocultural Synthesis* (Goodman and Leatherman 1998a) because additional sources of bias have affected the reference collections. Goodman and Leatherman (1998:19-20) describe five interrelated concepts of the new biocultural synthesis (NBS): 1) a rejection of reductionist indicators and consideration of the effects of power relations on forming local environments; 2) the importance of the links between the local and the global, and the interaction of the two; 3) local and global historical context are critical to understanding the direction of social change, and the biological consequences of change including evolutionary change; 4) the environment only takes on meaning in relationship to the subject, and thus, humans shape, and at the same time, are shaped by their environment; 5) the ideology, knowledge, and perspectives of subjects and scientists are essential to understanding human action.

The NBS approach provides additional explanatory power for unraveling and interpreting the sources of variation in reference collections. These five critical issues that are at the core of the NBS are clearly interconnected; however, some of these issues are more applicable than others when the focus of research involves reference collections and the goal is *not* to reconstruct the population from which the collection was derived. The third issue with an emphasis on historical context and the fifth issue with an emphasis on the ideology, knowledge, and perspectives of the researchers are described, further developed, and illustrated with examples below. Both issues are considered together because they are so closely linked.

With the exception of problems associated with excavation methods, the filters that influence the analyzable cemetery samples are imposed by the cultures and societies that interred or entombed those individuals and not by the researchers studying the samples (Hoppa 1996). With most research involving cemetery collections, the researcher is usually temporally, and often culturally distant from the research subjects. Socio-economic and cultural criteria such as religion, age, wealth, status, occupation and views of gender of the research subjects at the time of death resulted in inclusion in a cemetery or special entombment in a crypt (Hoppa 1996). Similar criteria have had an effect on which individuals are included in a reference collection; however, the researchers (Drs. Todd, Terry, Trotter, Huntington, etc) working in the context of their research discipline and society (which is also the society of the research subjects) have played an important role in determining which individuals were included in the reference collections. Thus, the fifth issue (researcher bias) described by Goodman and Leatherman (1998), particularly with a historical perspective (issue three) may have the greatest effect but seems to be ignored in most research involving reference collections.

The ideology, knowledge and perspective of past and current researchers have had and continue to have direct effects on theoretical and methodological issues including (1) construction of the demographic profile of the reference collections; (2) what documentary data were systematically collected, catalogued, and made readily available for research; (3) approaches to sampling the collections; and (4) research questions that have been investigated using the collections. One of the major biases, that bridges all four of these issues and the history of physical anthropology, is a racial approach to investigating human variation.

PHYSICAL ANTHROPOLOGY AND THE RACE CONCEPT

The Current Status of the Race Concept

In two complementary chapters in *Building a New Biocultural Synthesis,* Armelagos and Goodman (1998) and Blakey (1998) review the race concept and the use of racial classification in anthropological research and demonstrate the effects of researcher bias. Using a historical approach, they demonstrate that power relations in a society have an enormous impact on researchers' views and what is accepted by academics and the general public as valid research; and that science and ideas of what is natural were used to perpetuate and justify poverty, racism, and inequality. Armelagos and Goodman review the race concept and its use in anthropology because,

> the history of race so clearly and centrally illustrates the role of sociopolitical forces in the origin, adoption and persistence of ideas. We suggest that the new biocultural anthropology should include sophisticated understanding of the history of concepts. Furthermore, a new area of biocultural collaboration involves the determination of the social and biological costs of racism (1998:359-360).

Reviewing the use of the race concept in a historical context, along with current genetic and skeletal data, provides very strong evidence for rejecting the continued use of racial categories or equivalent typological approaches in the study of human variation (Keita and Kittles 1997, Cartmill 1998, Templeton 1998, Armelagos and Goodman 1998, Lieberman 2001, Molnar 2002).

Culturally constructed folk taxonomies that are based on pre-Darwinian and pre-Mendelian concepts are not useful for the study of human variation (Keita and Kittles 1997, Cartmill 1998, Templeton 1998, Armelagos and Goodman 1998, Lieberman 2001, Molnar 2002). Rather, the continued use of racial categories (or equivalents with other terminology) for the study of human variation only further supports current structures of power (Armelagos and Goodman 1998). The term folk taxonomy is in contrast to biological or scientific taxonomy. Biological taxonomy is derived from the Linnaean system of classification where living (or once living) things are classified in hierarchical taxa (species, genus, family, order, class, phylum, kingdom). Late 20th century approaches to biological taxonomy rely on more biological criteria than Linnaeus' early classification, but species are still organized in a hierarchical manner. Similarly, folk taxonomies are hierarchical classifications; however the categories are defined by cultural criteria. Folk taxonomies of "natural" or "unnatural," "edible" or "inedible," and "us" and "them" vary with cultural context and are not based on scientific (testable) assumptions. Familiarity with multiple folk taxonomies for classifying humans and the historical circumstance that led to their development is essential for understanding this cultural basis for the classifications (Fish 1999). For example, the classification system in the U.S.A. is based on "blood" which classifies anyone with "one drop of Black blood" as "Black" even if one parent is described as "White" (Fish 1999). In other words, all "hybrids" are considered "Black" regardless of actual ancestry, genotype or phenotype (beyond *possibly* skin pigmentation)[2]. Keita and Kittles (1997) provide a detailed critique of how this racial thinking is a major methodological impediment to the study of human genetic variation and the development of evolutionary theory with its assumptions regarding discrete "races," original "pure races," and admixture in American "Negroids" and (geographically) Africans.

The debate over the usefulness of the race concept for investigating and explaining human variation is older than the discipline of physical anthropology (Caspari 2003). Boas (1912, 1940) is widely cited as one of the few mainstream anthropologists from the early part of the 20th century who actively questioned the race concept (Brace 1982, Armelagos and Goodman 1998, Caspari 2003, Gravlee et al 2003). Recently, two

---

2 In other parts of the world such as Haiti, "one drop of White blood" may result in a classification of "White" (Fish 1999). In this context the "hybrids" are usually described as "White." Still in other areas, for example Salvador, Brazil, ten different offspring of the same parents may be classified into ten different racial groups or *tipos* (Fish 1999).

independent re-analyses of Boas's original data have been published that come to two very different conclusions. Sparks and Jantz (2002) conclude that when Boas's data are analyzed using modern statistical methods, the data do not demonstrate that environmental plasticity is more important than genetics in cranial variation. In contrast, again using modern analytical methods that were not available to Boas, Gravlee and colleagues (2003) conclude that Boas's (1912) conclusions are supported by a re-analysis of his data.

The stated goal of both studies was to test Boas's conclusions using modern statistical methods and his own data in their original form. The discrepancies in statistical significance between these two re-analyses are problematic. Some of the discrepancies may be because the studies use different statistical tests and different sub-sets of Boas's data. Holloway (2002) raises some questions regarding how Sparks and Jantz selected their sample of 8,000 to 10,000 from Boas's sample of about 13,000. On the other hand, Gravlee and colleagues's (2003) sample is much closer in size to Boas's original sample. A total sample size is not provided but can easily be totaled ($n > 12,600$) from the available information in their Table 2. Additionally, there are some differences in the research design of the two studies. Gravlee and colleagues (2003) state that they are testing the three most important of Boas's conclusions. Sparks and Jantz (2002:14636) state that they intended to fill a void in the literature since Boas's "findings have never been critiqued in a systematic way."

Rather than trying to assess whether Boas was "right" or "wrong" or "why Boas published such seemingly erroneous conclusions" (Sparks and Jantz 2002:14638), addressing two underlying fundamental questions might be a better approach. First, is cranial morphology plastic? Trying to assess which contributes more to variation, genes or environment, ignores the interactive effects of both in producing an observable phenotype. Second, does the cephalic index or other cranial morphology correspond with racial categories? In regard to the first question, both re-analyses are constrained by their attempt to reproduce Boas's original work using modern statistical approaches. A better approach might be to consider Boas's data within a modern theoretical framework for understanding human variation. Both re-analyses use the data in Boas's original groupings. The criteria for defining those groups are not consistent and thus problematic. Some groups are defined by religious criteria (Hebrew), while others by different levels of political boundaries (Scottish, Polish, Hungarian and Slovak), and still others by regional/geographic criteria within nations (island of Sicily and central Italian). These categories were not used in an analysis by Gravlee and colleagues (2003) when they tested the third hypothesis "There are significant differences in head form between U.S.-born children and their *own* immigrant parents; these differences are greater than those between foreign-born children and *their* parents" (2003:128; citing Boas 1912; emphasis added). This approach, where children are compared to their parents, controls for real genetic variation (as compared to ancestry of a population). Gravlee and colleagues (2003:133) found that their analysis supported Boas's original conclusion: "foreign-born descendants are notably more similar to their parents than U.S.-born descendants are to theirs." Correlation coefficients are consistently much larger between foreign-born descendants and their parents than U.S.-born descendants and their parents. The pattern is the same if children are compared to either the mother, the father, or both.

With regard to the question of the race concept in its current form in physical anthropological research, neither re-analysis provides any conclusive evidence. Very early in his anthropological career, Boas rejected the essentialism of the race concept and he questioned the validity of "sub-racial" types (Hebrews, Hungarians, etc) but he never totally abandoned the essentialist ideas regarding "major races" (Caspari 2003). In other words, the data for this immigrant study were collected to address issues regarding the race concept in the context of the first decade of the 20th century and not the first decade of the 21st century. At the time when Boas conducted his study, the cephalic index was considered to be stable and *the* tool for classifying human variation into clear categories (Holloway 2002). Again constrained by their attempts to reproduce Boas's original results, both re-analyses (Sparks and Jantz 2002, Gravlee and colleagues 2003) fall into an early 20th century typological methodology and focus on group means with no consideration for the range of variation within each group.

Sparks and Jantz's conclusion that ancestry (group) is an important source of variation is not warranted considering the problems with how the groups are defined. The results of Sparks and Jantz (2002: Figure 1), and Gravlee and colleagues (2003: Figure 3) clearly show that the mean cranial morphology of the Scottish sample is intermediate between the Sicilian sample and the central Italian sample. More similarities should be expected between the two Italian groups, not because they are Italian, but because they have more cultural similarities (suggesting contact between regions) and are geographically closer to each other, which makes gene flow more likely (Relethford

2001).

Both of these re-analyses of Boas's data contribute little to current debate regarding the race concept. With the exception of Coon's (1962) publication of *The Origin of Races*, professional physical anthropologists have rejected the essentialist and deterministic aspects of the race concept that has been misused by third parties (Caspari 2003). While there are still some social implications of the continued use of the race concept (Armelagos and Goodman 1998, Blakey 1998, Lieberman 2001), the current debate regarding the race concept (or equivalents) may be more broadly framed as a question of the utility of this concept when investigating human variation. Since Boas's early publications, a large body of literature presenting genetic and craniometric evidence clearly illustrates that human variation does not cluster into racial categories and these categories are not useful for the study of either phenotypic or genotypic variation (Keita and Kittles 1997, Armelagos and Goodman 1998, Cartmill 1998, Templeton 1998, Brown and Armelagos 2001, Relethford 2001, Lieberman 2001, Relethford 2002, Molnar 2002). In other words, regardless of the degree of heritability of cranial morphology, neither genetic or craniometric variation corresponds with racial categories or continental origin.

Two recent studies summarize and synthesize much of the evidence that has been amassed since the adoption of the modern evolutionary synthesis after World War II by physical anthropologists. In a review of numerous studies that have investigated genetic and/or protein variation in humans, Brown and Armelagos (2001) demonstrate that racial categories account for very little genetic variation. They found that all genetic/protein studies continue to support the conclusions regarding race and genetics established by

Livingstone (1962) and Lewontin (1972). Additionally, using both genetic evidence and craniometric data, Relethford (2002) confirmed these conclusions: 1) there is much more variation within any "race" than between "races;" 2) only 6-13% of genetic and craniometric variation can be attributed to "race" or continental origin; 3) human genetic and craniometric variation does not correspond with racial categories, continental origin or skin pigmentation. Relethford concludes his paper by stating,

> it is ironic that one of the most visible human characteristics and the one that has dominated in racial classification is the least illuminating about underlying patterns of global human genetic diversity. The high level of among-region variation seen in skin color does not apply to genetic or craniometric traits, thus weakening the case for using skin color as a proxy for other biological traits (2002:397-8).

Despite some of these limitations of the race concept, racial categories and/or continental origin categories continue to be used extensively in physical anthropological research. There is some debate as to how extensively the race concept is used. Lieberman and colleagues (2003) are somewhat critical of Cartmill's (1998) conclusion that there has been little change in the use of the race concept in the *American Journal of Physical Anthropology* (AJPA) between 1965 and 1996. A review of the AJPA by Armelagos and Van Gerven does indicate a drop in the use of race in the journal after 1996. Lieberman and colleagues (2003) suggest that there has been a qualitative rather than quantitative change in the use of the race concept in the journal since 1965. As an example, Lieberman and colleagues (2003) compare the use of the race concept by Feldesman and Fountain (1996) with earlier publications from 1965 and 1931. Feldesman and Fountain (1996) conducted a meta-analysis of previously published research to investigate the relationship between "race" and femur-stature ratio[3]. As Lieberman and colleagues (2003) note, Feldesman and Fountain (1996:211) state "we recognize... that this grouping was crude and did not capture the more substantial store of geographic variation characteristic of modern humans." Despite their qualifications about the racial categories, Feldesman and Fountain's *still* used these categories for their analysis. Their approach echoes the pattern of use of the race concept that is found in the earliest publications describing forensic identification methods including age and sex determination methods.

The focus of the discussion (Lieberman et al 2003) on how to measure the use of the race concept obscures two critical issues regarding the continued use of this approach in physical anthropology. First, despite some of the obvious changes after World War II when the ranking of racial differences was dropped, and some additional qualitative changes since 1965, more than one third of the papers published in the AJPA since 1965 use race (or other typological approaches) for the investigation of phenotypic and genotypic variation (Cartmill 1998, Lieberman et al 2003)[4]. Second, Relethford (2001, 2002) demonstrates two things very clearly: human craniometric variation does not correspond with racial categories, continental origin or skin pigmentation; *and* the unit of analysis used to investigate variation will have an effect on the results. The use of racial

---

3 Feldesman and Fountain (1996) use quotation marks around racial terms in their paper.

4 The prevalence of racial categorizations or continental equivalents in physical anthropology may actually be grossly underestimated by Lieberman and colleagues (2003), Armelagos and Van Gerven (2003) and Cartmill (1998). The overwhelming majority of forensic anthropology papers published in last three decades of the 20th century were published in the *Journal of Forensic Sciences* and not the AJPA (Buikstra et al 2003: Figure 2).

or continental groups as a unit of analysis may result in far worse than a "crude" measure of human variation as Fledesman and Fountain (1996) suggest. Using racial groups as a unit of analysis, particularly when combining data from different sources, may completely obscure the actual pattern of variation in some investigations (Albanese 2000, Chapter 5). Thus, while some of the implicit and explicit social implications of the continued use of the race concept may have changed (Lieberman et al 2003), there are likely a number of methodological costs associated with the continued use of the race concept.

The race concept in various forms continues to be used extensively in skeletal biology and forensic anthropology for investigating human variation and for defining the range of applicability of identification methods. A review of the use of the race concept clearly indicates that biological determinism has been abandoned. In most human skeletal research after World War II, human variation is not ranked by racial group. For example, Gill (1990:viii), a strong proponent of the race concept as biologically meaningful, denounces "some unfortunately serious political abuses of the concept." Major changes in the use of the race concept (or its outright abandonment in some cases) corresponded with major social changes: the aftermath of the Holocaust after World War II, and the civil rights movement in the U.S.A. (Caspari 2003). However, even at the end of the 20th century, various forms of racial approaches to skeletal (and other anthropological) research have been criticized for still maintaining the early 20th century typological and essentialist aspects of the race concept (Armelagos et al 1982, Goodman 1995, Arlemagos and Goodman 1998, Armelagos and Van Gervan 2003, Caspari 2003). Others (Gill 1990, Kennedy 1995, Owsley and Jantz 2001) have refuted this criticism and some physical anthropologists continue to use racial groupings for investigating patterns of human variation (for example, Ousley and Jantz 1998, Meadows Jantz and Jantz 1999), and for establishing the parameters for the applicability of sex and age determination methods (for example, İşcan et al 1984, 1985; İşcan and Miller-Shaivitz 1984, Holman and Bennett 1991, Marino 1995, Seidemann et al 1998). Independent of the political reasons for abandoning the race concept (Armelagos and Goodman 1998, Blakey 1998), there are methodological benefits to considering (1) the historical and political context of the race concept (Brace 1982), (2) this concept's influence on the methodology for the development of age and sex determination methods, and (3) this concepts influence in the construction of the reference collections that have been used so extensively to develop these methods.

Reference Collections, The Emergence of Physical Anthropology in the United States, and The Race Concept

The history of reference collections (and how they were collected) cannot be separated from a review of the use of the race concept in physical anthropology, and the emergence of physical anthropology as a separate discipline in the United States. The historical review of the race concept in physical anthropology is also an invaluable source of information for assessing the effects of bias associated with the social and scientific context on the collecting of skeletons.

The race concept had an enormous impact on the direction of physical anthropological research in the period before World War II (Blakey 1987, Blakey 1998, Arlemagos and Goodman 1998, Armelagos and Van Gervan 2003, Caspari 2003). Blakey is critical of many of the preeminent physical anthropologists from this period,

> Hrdlicka, Hooton, and Coon were producers of the apologetic explanations for the status quo of economic inequality, racial oppression, and imperialism. Eugenics was viewed as the comprehensive application of mainstream, evolutionary and racial studies in the United States and Europe... These were not mad, renegade or peripheral pseudoscientists. These were leading representatives of mainstream American anthropology practicing the scientific method with all its intrinsic subjectivity (1998:384).

In this passage, Blakey describes the physical anthropologists who were the peers of Drs Terry and Todd, and the paradigm under which much of physical anthropology operated at the time when over 80% of the Terry Collection and all of the Hamann-Todd Collection were amassed.

Blakey names Ales Hrdlička, Earnest Albert Hooton, and Carleton Stevens Coon because they had a great influence on physical anthropology in the period between the two World Wars. Rankin-Hill and Blakey (1994) refer to this group with its theoretical, research and political views as the "Harvard-Washington Axis." Hrdlicka is credited as the founder of physical anthropology as a separate discipline in the United States; Hooton was the individual who supervised the doctoral dissertations of many of the individuals who filled the positions of this new discipline; and Coon was one of Hooton's first students (Spencer 1981, 1997a, 1997b) and an outspoken pro-racial physical anthropologist even after World War II when he served as president of the American Association of Physical Anthropologists (Caspari 2003). While racial and at times racist perspectives dominated much of the physical anthropological research of the period

(Blakey 1987, Rankin-Hill and Blakey 1994, Goodman 1995, Armelagos and Goodman 1998, Blakey 1998), scientific racism was not necessarily simple, overt, always intentional, nor should all physical anthropologists from this period be painted with the same coarse brush[5].

W. Montague Cobb describes several telling incidents that involve both Aleš Hrdlička and T. Wingate Todd from this critical period when physical anthropology was being established as a discipline separate from anatomy, and reference collections were being actively collected (these incidents are described by Cobb 1959, Cobb 1981 and in interviews with Cobb published in Rankin-Hill and Blakey 1994). After Cobb graduated from Howard University as a M.D. in 1929, he was approached by the Dean of that medical school, Dr. Numa P. G. Adams, with an offer to consider a full-time academic career at the medical school. Cobb accepted on the condition that he could study anatomy. The problem was finding an institution where he could pursue a Ph.D. As Cobb describes,

> A place for advanced training was sought which could afford sound work in gross anatomy and its expansion to physical anthropology. It first appeared that there was no institution in the United States where this kind of training could be obtained by a Howard man [African-American]. Then Dr. Adams remembered that he had met a man in Cleveland... It was a great joy to discover that Dr. Todd was

---

5 Despite Hooton's views about race and eugenics, he was actively anti-racist and even worked with Franz Boas to combat some of the scientific racism at the time (Caspari 2003). Caspari (2003) describes an incident where in the early 1930s Hooton sent a letter to the seven leading physical anthropologists in the U.S.A. for their endorsement. Among other things, the letter stated that there was insufficient evidence to link physical features with mental ability or an evolutionary rank to races. Only Hrdlička signed the letter.

> already so advanced in his thinking and had amply demonstrated by his published work that prevalent American concepts in respect to race and human potential had no place in his laboratory (Cobb 1959:237; also reprinted in Rankin-Hill and Blakey 1994:80).

To place into context how advanced Todd's thinking was, Cobb was the only African-American to be awarded a Ph.D. in physical anthropology in the United States before the Korean War (Rankin-Hill and Blakey 1994).

While at Western Reserve University, Cobb met Hrdlička for the first time in 1930. Hrdlička was looking for someone to inventory all human materials available for anthropological research at institutions in the United States (Cobb 1930). The obvious candidate to complete this task was T. Dale Stewart (Cobb 1981), who had been on staff at the Smithsonian Institution in various capacities since 1924 (Ubelaker 1997). Todd, who was the chair of the committee of the American Association for the Advancement of Science (AAAS) charged to assign this task, appointed Cobb instead since he would soon be returning to Howard University in Washington, DC.

Todd's appointment of Cobb to conduct this inventory of research materials was not simply because of Howard University's proximity to the Smithsonian Institution. In an interview decades later Cobb stated, "now I have just understood in the last couple of years or so, why Todd gave me the job of the committee of which he was chairman. He turned that over to me because I was right down here near Hrdlička. But he also wanted to see how Hrdlička would treat me..." (Interview with Cobb from 1985, in Rankin-Hill and Blakey 1994).

Hrdlička became the first curator of physical anthropology at the Smithsonian

Institution in 1903, and his efforts throughout the 1920s eventually led to the establishment of the American Association of Physical Anthropologists (AAPA) in a meeting of Section H (Anthropology) of the AAAS in 1928[6] (Spencer 1981, Stewart 1981). Hrdlička's name literally cannot be separated from the field of physical anthropology in the United States: "founded by Aleš Hrdlička, 1918" appears on the cover of every issue of the AJPA. In the early years of the journal, Hrdlička set the editorial policy[7], established the format, solicited papers, edited papers, handled subscriptions, dealt with printers, paid for deficits, and contributed significantly to the journal with his own work (Stewart 1981). At the same meeting of the AAAS that the AAPA was established, the AJPA was accepted as the new association's official organ (Spencer 1997a). Hrdlička had pressed ahead to established the journal with such an unstable financial footing before the end of World War I because Boas's students gained control of the American Anthropological Association and its journal *American Anthropologist* during the War (Spencer 1981). The establishment of the AJPA is a landmark in the history of physical anthropology in the United States for three reasons,

---

6 Hrdlička served as the chair of the committee to establish the new association. Other members of the committee included Earnest Hooton and Robert Terry. Conspicuous by his absence from the committee was Todd, who was against the separation of physical anthropology and anatomy since most practicing physical anthropologists at the time were M.D.s and held positions at departments of anatomy (Stewart 1981, Rankin-Hill and Blakey 1994). Eventually Todd served as President of the association in 1938 (Spencer 1997a).

7 In the first issue of the AJPA Hrdlička (1918:18) wrote, "the paramount scientific objective of physical anthropology is the gradual completion... of the study of the normal white man living under ordinary conditions (also cited in Armelagos and Van Gerven 2003).

> (1) it had the immediate effect of securing the discipline's identity, (2) it provided Hrdlička with an opportunity to codify the discipline in broader and essentially modern terms, and (3) it gave him the political platform from which to continue his campaign for the recognition of physical anthropology as a legitimate and independent discipline (Spencer 1981:358).

Despite these major distinctions, only rarely has any of Hrdlička's work been cited after World War II by physical anthropologists because he, "firmly believed in the innate superiority of academicians (Hrdlička 1941), men (Hrdlička 1925a), and 'native' or 'old American' whites (Hrdlička 1925b)" (Ranken-Hill and Blakey 1994:77). Using craniometric data, Hrdlička was overt in his intentions in trying to explain and justify the social inequalities in the United States in the first half of the 20th century (Blakey 1987, Rankin-Hill and Blakey 1994). By "native" or "old American" Hrdlička was referring to American citizens whose ancestors had immigrated from Europe at least 80 years and closer to 150 years before 1925 and whose parents and all four grandparents were born in the U.S.A. (Hrdlička 1925b). In contrast to Boas who viewed race as independent of language or culture, Hrdlička viewed race as a constraint on culture (Aremelagos and Van Gerven 2003) and wrote "the real problem with the American Negro lies in his brain, and it would seem, therefore, that this organ above all others would have received scientific attention" (Hrdlička 1927:208-209; also cited in Aremelagos and Van Gerven 2003:56). Brace (1982) describes some of Hrdlička's major influences and the foundation for his research interests but in some ways his research pursuits are very baffling. Hrdlička was born in Humplec, Bohemia (now the Czech Republic) and immigrated to the United

States when he was about 12 years old (Spencer 1997b)[8].

The perspective advocated in the NBS is not required to see that there are social, political, and economic implications of researchers at prestigious government funded institutions, such as the Smithsonian Institution, trying to establish a real or perceived biological basis for inequality, or the political implication of Todd's appointment of Cobb for the inventory project. Considering the spectacular careers of Stewart and Cobb, skill and proximity to the Smithsonian Institution (since Stewart was on staff there) probably were not issues for deciding who should conduct the inventory project.

The political implications of a federally funded institute for physical anthropology was not lost on some of the students of Franz Boas. Besides, the debates in print and at conferences, and over editorial control of journals, the "Boasian school" and the "Harvard-Washington axis" constantly fought over control of various committees of the National Research Council (NRC) and the National Academy of Science (Rankin-Hill and Blakey 1994). Some of Boas's former students were instrumental in blocking two separate attempts to secure funding from the NRC for an institute of physical anthropology under Hrdlička's control at the Smithsonian Institution (Spencer 1997b).

At their first meeting after Cobb returned to Washington, Hrdlička told Cobb,

---

8 Perhaps the fact that he was male and an academician superceded the fact that he was an immigrant? Or, as Cobb suggests in this anecdote, there really was no biological basis for the ranking of groups in this manner, "Then one day I said 'Dr. Hrdlička, you accept me alright, why do you have the restrictive ideas about so called pure Negroes?' 'Well the Negro is alright if he's had the hardships the white man has had. You have the vigor of the hybrid,' he said. Well anytime you see something you cannot explain, you invent an explanation" (Cobb 1985 interview cited in Rankin-Hill and Blakey 1994).

"You have been out there with Professor Todd. He is a brilliant man but he has far too many radical ideas. Now if you will come down from time to time, I will show you the way," (Hrdlička cited in Cobb 1981:519). Cobb and Todd were very much against the mainstream concepts of racial determinism but could hardly be called Boasian in their approach. Neither showed interest in socio-cultural anthropology or a four-field approach to research (Blakey and Rankin-Hill 1994). Throughout the 1930s, Cobb actively published several papers (but not in the AJPA) where he used a physical anthropological approach to counter the hypotheses that there were differences between "Caucasoids" and "Negroids" and he followed a biocultural and medical anthropological approach to investigate the effects of racism, segregation and poverty on health (Rankin-Hill and Blakey 1994). One paper, "Race and Runners," was published after Jesse Owens won multiple gold medals at the 1936 Olympics in Berlin. Cobb collected anthropometric data from Owens and other athletes and compared it to the detailed anthropometric data collected by Todd from cadavers that were to be included in the Hamann-Todd Collection, and conclusively demonstrated the non-concordance of so-called racial traits. Training, and not race, was the reason why so many world records continued to be broken (Cobb 1936, also summarized in Rankin-Hill and Blakey 1994).

Though it is often ignored, possibly one of the most significant ideas in the NBS approach is that *all* researchers bring a bias to their research and that all research can have political, economic, and social implications (Blakey 1998, Goodman and Leatherman 1998b, Roseberry 1998; Singer 1998). The implications of Todd welcoming Cobb to study anatomy at Western Reserve University in 1930 and Cobb's publications were as overtly political as Hrdlicka's research. The debates in print, at conferences and for the control of the NRC by the "Boasian school" and the "Harvard-Washington axis" were political struggles as much as they were scholarly debates (Rankin-Hill and Blakey 1994).

The incidents described above involving Cobb and Hrdlička are presented to illustrate that this bias in past (and current) research can be divided into three interconnected levels: biases of the individual researcher, biases in the field of research, and biases in the greater culture or society. A federally funded institute of physical anthropology headquartered in Washington, DC, and headed by someone with such overtly racial views, has political and social implications (Blakey 1987, Rankin-Hill and Blakey 1994). However, Hrdlička had three *personal* visions for physical anthropology in the United States: the establishment of an association of physical anthropologists, the establishment of a journal of physical anthropology, and the establishment of a national institute of physical anthropology (Spencer 1981). Hrdlička attributed the rejection for funding for the institute to "personal animosities" between him and some of Boas's students on the NRC committee responsible for the funding of anthropological research (Spencer 1981:359). Furthermore, despite their obvious differences, Cobb described Hrdlicka in this way: "his single minded devotion to his field was never in dispute" (1981:519).

The biases in the field of research from that period were dominated by a racial approach that had an impact on the research questions that were addressed and the approaches to conducting that research. Even Cobb and Todd, who were quite overt in their research against racial determinism and outright scientific racism conducted their research within a racial paradigm. In the series of ground breaking papers from 1920 and 1921 on determining age from the pubic symphysis, Todd divided his sample by sex and "race" as is evident in the titles of the papers: "Age changes in the pubic bone: I. The male White pubis," (1920); "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," (1921). Todd's approach to interpreting pelvic variation was different from earlier work by Turner (1885) and later work by Schultz (1930) who clearly ranked various "races" as more or less "primitive." Todd did not rank the variation he saw, but the division of his sample by race reflects the racial thinking that dominated physical anthropology at the time. Similarly, Cobb fought against racism in his anthropological publications and in other avenues, but he also wrote,

> the defects of modern European civilization are so obvious, particularly in respect to its dependency on exploration and period slaughter and its failure to adjust population size and caliber to resources, that while its material achievements excite amazement, its social organization hardly evokes excessive admiration (1939:324, reprinted in Rankin Hill and Blakey 1994:87).

Rankin-Hill and Blakey (1994) cite this passage by Cobb as an example of his "thorough demonstration of human equality" since Cobb not only presented the positive attributes of Afro-Americans but he also balanced his argument by recognizing weaknesses that were also present in European society. I would argue that, positive or negative, these are all racial stereotypes even if Cobb was substituting "European" for "White".

The reason to cite this passage by Cobb is not to criticize him but to illustrate that although Cobb's (and Todd's) approach may have been the only approach possible to combat the racism at that time, they still operated within the same racial paradigm that permeated the field of physical anthropology and the greater society in the U.S.A. It is ironic in that the same views that made it necessary for Todd to collect anthropometric data (and skeletons) by "race," were the same forces that drove Cobb to write the paper, "Race and Runners." The paper would not have been possible if Todd had not collected the anthropometric data by "race" in the first place. Todd and Cobb may have been radicals by Hrdlička's standards but they conducted their research firmly within a racial framework.

The decades between the World Wars is a critical period in the development of physical anthropology in the United States (Spencer 1981), and thus, on the discipline in the rest of the world. The AAPA is both the oldest and the largest association in the world devoted exclusively to physical/biological anthropology with a number of international members (Spencer 1997a). This inter-war period is when physical anthropology coalesced into a separate discipline with the expansion of the AJPA, the establishment of the AAPA, and the awarding of doctorates in physical anthropology (Spencer 1981). The collections that have been used for so much research, including the development of so many identification methods (age, sex, stature, etc.), could not have been amassed at any other time in the history of the discipline. Furthermore, the methodology for the development of these identification methods arose in this period. The collections, the methods, and the methodology all bear the stamp of the period. Soon after Todd's death, Shapiro wrote a brief obituary where he stated that "Todd regarded the skeleton as a kind of auto-biographical record of the individual..." (1939:459). Following this metaphor, the collections themselves and how those collections were and are used are biographical records of physical anthropology.

Todd's views, which reflected the mainstream views of physical anthropology at the time, are illustrated in the following passage:

> there is no large series available of pure African material or indeed of any other non-White stock with accurate information as regards to age. Hence we are driven to use American Negro-White hybrid material for comparison with the observation upon the male White skeleton. Actually this is not of practical disadvantage because the differences between the White and Negro stock are sufficiently marked to be picked out even in the hybrid material... in the course of present studies one is convinced that the negroid type of pelvis is retained in the hybrid population. Since then in morphological features the Negro-hybrid pelvis is distinctly negroid one must expect that the features of growth and metamorphosis of the pelvis bones will also bear the testimony of the negroid origin (1921:3).

Although Todd did not rank the differences between races, he certainly followed the mainstream assumption that races were real biological categories that were the sources of variation. This was a view that was *not* supported by the research presented in the same paper on age changes in the pubis. He concluded section II of his paper on the "Negro-White hybrid" by stating that "we are forced to conclude from this review that, so far as the Negro and White pelves are concerned stock has strikingly little influence upon pubic metamorphosis," (Todd 1921:26).

This view on racial categories, reflected in Todd's comments and still evident in the Hamann-Todd Collection, is echoed in a publication by Terry 11 years later:

> the name 'American negro' is used here to imply the high probability of the presence of some racial strains other than negro recently introduced in the series under investigation

> [from the Terry Collection]. Aside from the hybridization of negro stocks that has been in process for an unknown period in Africa, negroes brought to the western hemisphere have been in contact with the native Indian population and with white colonists and their descendants since the early years of the sixteenth century... For the purpose of studies such as the present one [on clavicular morphology], it is perhaps best to acknowledge at the start that only a rough diagnosis of purity or admixture is possible from the evidence at hand... (1932:352-353; quotes around American negro appear in the original publication).

Despite the obvious contradictions in the racial categories, both collectors religiously documented "race" along with age and sex for each skeleton that was catalogued. These are the three fields of data that continue to be consistently used to define sub-samples for research (see Appendix A for examples of this approach in sex determination methods).

These quotes are not presented to criticize their work when it is considered out of the context in which they conducted research in the 1920s and 1930s. As Brace notes, "historical assessment does no good if it simply turns into an exercise of self-flagellation," and does not lead to action that can provide a sturdier foundation for future research (1982:24). These quotes are presented to illustrate the context for physical anthropological research in the 1920s and 1930s; to illustrate that the efforts of these two pioneers working in this context had an influence on the collections they constructed; and to illustrate that the research context in the 1920s and 1930s continues to have an impact on some of the current research in physical anthropology in general, but particularly when samples from the Terry and Hamann-Todd Collections are used. For example, 25 years after Todd stopped collecting skeletons and just before collecting for the Terry Collection ended, Giles wrote regarding the Terry Collection:

> categorization of specimens as Negro on the collection records undoubtedly reflects American cultural standards, not genetic ones. It seems reasonable to assume that any person exhibiting phenotypic evidence of Negroid admixture was considered a "Negro." The Negro sample thus has an indefinite white American and possibly American Indian component, while the white sample represents American whites of European descent (Giles 1964:129).

More recently, Ously and Jantz state,

> Genetic exchange among the various populations of *Homo sapiens* means that there will always be overlaps in distributions, resulting in some degree of misclassification. American Blacks and Hispanics are known to incorporate genes from at least two ancestral populations. One can observe that in most analyses involving all groups [American Blacks, American Whites including European born individuals, Chinese males, Hispanic males, Japanese, and Vietnamese males], Hispanics assume a central position... Here it should be emphasized that FORDISC can only return answers based on metric information, but the questions are often phrased in social terms. The progeny of a Black-White couple may have been socially known as black, but may possess more genes from white than black sources... Since a person's social race is not based on objective metric criteria, metric traits are an imperfect method for assessing probable social race (1996:19; FORDISC User Guide).

Although the terminology varies from quote to quote (population replaces or is used interchangeably with stock, racial strain, continental origin etc), there are similar implicit assumptions that echo early 20th descriptions of human variation: overlap between groups is due to gene flow (or mixing of stock) between these "populations;" individuals in the overlap range are "hybrids" and distance from the group mean is interpreted as distance from the ideal genetically-based type; there were original pure races which are the source of "White" genes and "Black" genes; continental origin is correlated with genetic and morphological variation; possible misclassification of individuals is due to the discrepancy between socially defined race and morphology (hybrids intermediate), and not problems with the classification system. None of these assumptions are consistent with the genetic and craniometric data regrading the distribution of human variation (Keita and Kittles 1997, Armelagos and Goodman 1998, Cartmill 1998, Templeton 1998, Brown and Armelagos 2001, Relethford 2001, Lieberman 2001, Relethford 2002, Molnar 2002).

A History of Age Determination Methods and Research Methodology

The effects of racial thinking of the period on the identification methods and physical anthropology methodology are seen in Todd's papers on age changes in the pubic symphysis. In these papers, Todd (1920, 1921) not only developed new age determination methods, he also established the methodology and rationale for the development for an entire class of age determination methods that have been developed since 1920. Todd was not the first to note that there were age changes in the pubic symphysis (see Todd 1920:292-297) but he was the first to develop a system of phases (McKern and Stewart 1957). Furthermore, over eight decades ago, Todd wrote,

> in the investigation of large series of human skeletons various modifications of the bones are met with which hitherto have not been accurately checked up against age. Many of these modifications which appear successively during adult age are on the border-line between the anatomical and the pathological. Some cannot be so regarded, but must be classified as purely anatomical changes... it is this prime importance of the symphysis as an age indicator which calls for its description first of all the age features... [but] no individual part of the skeleton

> however is infallible, and the most accurate estimate of age can only be made after examination of the entire skeleton. (Todd 1920:287-288).

In the decades that followed, Todd and Lyon (1924, 1925a, 1925b, 1925c) developed a series of methods using cranial sutures which was re-evaluated and greatly modified by Cobb (1952); Brooks (1955) presented a modified version of Todd's Method; McKern and Stewart (1957) investigated age-related changes in the entire skeleton using a sample of American male Korean War dead with the pubis symphysis method being divided into a three component system; Gilbert and McKern (1973) used a female autopsy sample and also divided the changes in the pubic symphysis into three components; Lovejoy and colleagues (1985a, 1985b) developed a multifactorial method, and a method involving the auricular surface; Meindl and colleagues (1985a) suggested modifications to the original Todd pubic symphysis method; İşcan and colleagues (1984, 1985, 1987) applied the same approach to the sternal end of the rib; and Brooks and Suchey (1990) presented a modified version of Todd's method using a very large autopsy sample.

The basic methodology, particularly with respect to sample selection, has changed little since Todd's papers, and in many cases in the research that followed, samples are divided by sex and "race" (McKern and Stewart 1957, Gilbert and McKern 1973, İşcan et al 1984, 1985). The division by sex has been justified on the basis of changes related to pregnancy (Stewart 1979). The division of the sample for race-specific methods has not been substantiated[9]. Gilbert and McKern discuss the possibility of racial differences but

---

[9] Only İşcan and colleagues (1987) have found racial differences in age-related changes in the fourth rib; however, because of the small sample used to initially develop the methods (n = 118 males and n = 86 females divided over 6 decades)

based on their respective small sample of American "Negroes," they did not pursue the development of racially specific methods. In summary volumes, Gilbert and McKern's methods have been described as both race-specific (for example, Moore-Jansen et al 1994) and not race-specific (for example, Krogman and İşcan 1986) since the samples were composed predominantly of "Whites" and proportionally very few "Negroes."

Lovejoy and colleagues (1985a, 1985b, including Meindl et al 1985a) and Brooks and Suchey (1990) did not develop race-specific age methods and the emphasis on racial variation is much less prevalent in their research. However, the extra-publication documentation that accompanies the pubic bone cast prepared by France Casting for the Suchey-Brooks method and the Kent State University photos/slides of the auricular surface method is dominated by references to "race." In the documentation that accompanies the instructional pubic bone cast (not the reference casts) each individual is described based on certain criteria in a specific order: "race," sex, place of birth, and finally age. In the documentation that accompanies the instructional auricular surface photos, the "race" of each individual is prominently noted beside the age and specimen number of the individual from the Hamann-Todd Collection.

Based on Brooks and Suchey's (1990:237) statement regarding their sample, "the Suchey-Brooks method is based on a large multiracial sample of individuals of diverse socio-economic backgrounds," it is likely that racial and geographic origin data were included in this instructional kit in order to demonstrate that the method can be widely

---

and the small sample used to test the methods, other possibilities cannot be excluded.

applied. Still, the assumption is that either one of these variables, geographical origin or race, are factors potentially affecting human variation. Geographic origin at the state level such as Tennessee or New Mexico, is not the source of biological variation. The variation is the result of a *combination* of genetic factors and the socio-economic and politico-legal issues that have an effect on growth, development and health of individuals in specific communities. As discussed above, genetic variation does not correspond with racial variation. Brooks and Suchey (1990) do refer to these issues in their paper ("diverse socio-economic groups") but these issues are not noted in the documents that accompany the casts. Students who are learning these age assessment methods assume, as I did when I learned to use the casts of the pubic bones, that race is somehow relevant when determining age and a possible source of variation.

In many ways Todd's approach was brilliant and, as he mentions, could not have been possible without a large and well documented collection[10]. Evidence since then has shown that the underlying assumption for using these specific joints for assessing age is not as anatomical (as compared to pathological) as Todd first stated, and that there are differences between chronological age and physiological age (Cobb 1952, Brooks and Scuhey 1990). Unfortunately, the racial assumptions derived from the greater society and reflected in the demographic structure of the Hamann-Todd Collection have been perpetuated in the methods developed since 1920. In other words, more recent methods

---

[10] There are some problems with the accuracy and the distribution of the age data for the Hamann-Todd Collection (Cobb 1952, Lovejoy et al 1985b). Following a critical review of the documents associated with the collection, Meindl and colleagues (1990) have selectively sampled the collection to overcome many of these problems.

that follow Todd's methodology reflect some of the major theoretical concepts of race and human variation from the early 20th century despite Todd's own conclusion that there are no significant differences between "races" in symphysis morphology in his large sample.

A History of Sex Determination Methods and Research Methodology

The impact of Todd's groundbreaking research on age changes in the pubic symphysis is a good example of bias in the field of research. The methodology established by Todd in the early 1920s was still followed at the end of the 20th century. Todd's methodology continues to be accepted as the gold standard in the discipline for developing age determination methods even though his specific age determination methods in their original form are not widely used. Just as Todd established the methodology for the development of age determination methods, Washburn (1948) did the same for metric sex determination methods using the ischium-pubis index. Washburn's method was not the first metric method nor the first to use a pelvic index for determining sex (St Hoyme 1957). However, when Washburn published the ischium-pubis method, he established the standard methodology for the development of metric sex determination methods that is still widely used.

Metric approaches for sex determination in the United States have their origins in the last decades of the 19th century. The first use of an index for determining sex from the pelvis (but not the ischium-pubis index) can be traced to Matthews and Billings in 1891 (St Hoyme 1957). Other pelvic indices were used earlier by Turner (1885) and others but these focused on assessing racial differences (St Hoyme 1957). Around the same time other metric approaches for sex determination using various bones such as the femur and the humerus were also being developed (for example, Dwight 1894, Dorsey 1897, Dwight 1905; all discussed in detail in Stewart 1979). These early methods established a pattern where a sectioning point was calculated to discriminate between smaller females and larger males. After Dwight's publication (1905), there was a lull in the development of metric sex determination methods until Letterman's attempt to develop a method using measurements of the sciatic notch (1941). The poor performance of the method for discriminating between males and females is likely the reason why it has been largely ignored. The next major metric sex determination method published in the U.S.A. was Washburn's use of the ischium-pubis index in 1948.

Schultz (1930) first used the ischium-pubis index (along with many other measurements and indices) to compare variation in the axial skeleton in humans and non-human primates. Schultz used data that he collected and previously published data from a number of different collections. Whenever the information was available, Schultz always divided his sample by "race" or nationality but also provided overall means for the entire species of *Homo sapiens*. The three sub-samples that recur throughout the 135 page paper are "Negro," "White" and "Japanese." These groupings are artifacts of the organization of the collections and published data that Schultz used rather than any meaningful sub-sample division. The criteria for inclusion in any given sub-sample are not consistent. "Negro" and "White" are racial categories presumably based on phenotype or more likely geographic origin. However, inclusion in a group described as Japanese is based on geopolitical criteria.

Schultz also did not hesitate to assign biological rank in some cases, for example, "the sacrum of negroes is clearly more primitive and anthropoid in character than that of whites, whose sacrum is the most specialized among all primates in regard to its width as well as to its concavity" (1930:360). Schultz was describing sexually dimorphic characteristics in the morphology of the human sacrum which he attributed to "race" for two reasons. His samples are minute (n = 6 "Negroes" and n = 4 "Whites") and he did not control for sex in the group means. The tiny sample size alone could be the source of this variation between "races." When the tiny sample size is combined with the lack of control for sex, the slight difference of even one individual in the sex proportions in each sample will result in *apparent* differences between "races." When Washburn developed the ischium-pubis index into a sex determination method, he continued Schultz's approach of segregating samples but dropped the ranking of "races" and species, and he did not provide a species mean for *Homo sapiens*. The other major departures from Schultz's approach are in the sample size and the collections used. Washburn collected data from 300 individuals from one source: the Hamann-Todd Collection.

With the high accuracy of the ischium-pubis index for sex determination (reported at better than 90%), Washburn ushered in a revival of a metric approach to sex determination from the last decades of the 19th century. The metric approach for sex determination took off after Thieme and Schull (1957) confirmed Washburn's results for the ischium-pubis index and added discriminant function analysis to the methodology, an approach that has become the default multivariate statistical approach in physical anthropology for sex determination methods (see Appendix A for examples). Thieme and Schull were not the first to use discriminant functions for sex determination. Pons (1955) used a sample from the Ferraz de Macedo Collection (Lisbon) to develop sex determination methods using discriminant functions[11]. Thieme and Schull cemented the discriminant function approach to the sex determination methodology. After 1957, the standard methodology for developing metric sex determination methods was set and has largely remained unchanged. The age of the methodology itself is not necessarily the problem. The problem arises from the fact that this methodology is based on racial assumptions regarding human variation that predate genetic and evolutionary theory.

The gold standard for developing sex determination methods established by Washburn (1948) and then expanded and reinforced by Thieme and Schull (1957) has features that can be grouped into three major characteristics: (1) with few exceptions only "race" and sex data are considered at the exclusion of all other documentary information available for reference collections; (2) discriminant functions are used when multivariate statistical approaches are required; and (3) "race" is the critical criterion for defining the sample (or dividing the sample into sub-samples), and the accuracy of the method is presented by "race." All three characteristics are closely connected with the pre-

---

11 The Ferraz de Macedo Collection is sometimes referred to as the "old Lisbon collection" and is an entirely different skeletal collection than the current collection in Lisbon: Luís Lopes Collection. With the exception of 33 skulls, the old Lisbon collection was destroyed in a large fire in March, 1978 (Hugo Cardoso, personal communication).

categorization of human variation into racial categories [12].

The table in Appendix A is a summary of many of the metric sex determination methods developed in the United States since the end of the 19th century. The table is comprehensive but not exhaustive since duplicates are excluded (for example, there is only one entry for FORDISC rather than separate entries for each version 1.0 and 2.0). The approach to research, with pre-categorization of samples into racial groups, is evident from the start (Dwight 1894, Dorsey 1897, Dwight 1905; also summarized in Stewart 1979). While many of these features are easily listed in a table format, the important differences in Washburn's paper may not be obvious from Appendix A. With Washburn's (1948) publication, a new style and format was established for the writing and structuring of a paper that presented a metric sex determination method. This difference is easily demonstrated with a brief comparison between Washburn's (1948) paper and the paper by Letterman (1941). As with most publications investigating human variation before 1948 (for example, Schultz 1930), Letterman presented the sex determination method as a product of an investigation of sex (and "race") differences. The title of the paper is "The Greater Sciatic Notch in American Whites and Negroes." From the outset, Letterman does clearly state that the paper will deal with sex

---

[12] These characteristics for the development of sex determination methods are seen most often in metric sex determination methods developed in the U.S.A. using a sample from the Terry Collection and/or the Hamann-Todd Collection. However, some of these characteristics, particulary the emphasis on "race" are still present in the methodology used by researchers from other countries using other collections (for example, Asala 2001, used a sample from the Dart Collection), and in the methodology for the development of morphological sex determination methods (for example, Phenice 1969).

determination methods involving the sciatic notch; however, the utility of the sex determination method is presented in the context of a discussion of sexual dimorphism and human variation (Letterman 1941:101). In many ways, Letterman's approach has more in common with Schultz's (1930) investigation of variation in the skeleton. In contrast, Washburn's paper has a much more rigid structure that is followed in most subsequent publications where metric sex determination methods are presented (Theime and Schull 1957, Giles and Elliot 1963, Steele 1976, Dibennardo and Tailor 1981, etc.). Washburn does present some context for the variation in the pelvis and he does explore some of the reasons for the success of the index; however, the focus is on the efficient presentation of the index as a sex determination method and on its high accuracy with minimal elaboration. For these reasons, Wahsburn's (1948) paper is emphasized as the critical change in approach rather than Thieme and Schull's (1957) paper, despite the fact that Washburn did not use discriminant function analysis.

The first major characteristic of this gold standard for sex determination methodology is that the documentary data (that makes the reference collections so valuable) are ignored with the exception of information on sex and "race." Other basic information such as age at death, year of birth, cause of death, etc are never considered or are only very superficially addressed (for example, Stewart 1979, Dibennardo and Tailor 1981, France 1983, İşcan and Miller-Shaivitz 1984, Schulter-Ellis et al 1983, Holman and Bennett 1991, Marino 1995). In some cases (Giles and Elliot 1963, Giles 1964, Holland 1986, Holland 1991, Smith 1997) mean age at death or the age range for the sample is presented but these data are of limited value. When dealing with age data in this context, the *distribution* of the ages is critical, not the mean or the range. The mean for the sample does not reveal if the sample is well distributed over a wide range of adult ages and thus samples a wide range of variation associated with age-related changes. A normal distribution of ages cannot be expected, assumed, or even desired. Yet age at death is a critical source of variation when investigating issues related to sexual dimorphism and the development of sex determination methods. Giles noted over 30 years ago that age is a factor in allocation accuracy,

> Boulinier (1969) and subsequently Giles (1970[a]) have shown that in Giles' cranial sample of 408 American Negroes and Whites, where each specimen's age and sex are known from written records, the discriminant function method tends to misclassify younger males and older females. This is a tendency only; it would seem to have little practical effect and might, in fact, be present but undetected in visual sexing (1970b:102).

Though he downplayed the importance of age when methods are applied to a large sample, the impact on a single case may be significant. Ironically, it is the "visual sexing" methods where the effects of age have been recognized as a clear liability (Meindl et al 1985b, Walker 1995).

Walker argues that these age effects have been further confounded by cultural influences on skeletal biological research that he describes as "sexism in sexing" (1995:36). Walker found complex age-related changes in cranial sexual dimorphism when developing scoring systems for the mastoid process, mental eminence, supraorbital area, orbital margin and nuchal crest that could be used for sex determination[13]. Using

---

13 These age effects are not limited to the cranium. Age effects on the pelvis, particularly the pubis are described in detail on pages 189-191 and are not

300 crania of known age and sex from St. Bride's Church, the Terry Collection and the Hamann-Todd Collection, Walker found that age changes occurred in most of these cranial traits but in various degrees. The age effects were most pronounced in the supra-orbital area. In summary, a relatively high percentage of males younger than 30 years of age tended to have what has been considered a female-like pattern. In females over 45 years of age, the frequency of what has been considered a male-like pattern is higher.

Since Walker relied on data from three different sources, these complex age-related patterns are not due to differences in pattern of sexual dimorphism between different groups or biases in any one skeletal collection. Walker argues that ignoring these age-related factors has contributed to the apparent excess of males in some historical samples because post-menopausal females are identified as males. This problem is described as "sexism in sexing" because, "the error seems to be rooted more in cultural stereotype of 'typical' female morphology than in an appreciation for the complex biological reality of human cranial sexual dimorphism" (Walker 1995:36).

Beyond the age-effects, with few exceptions (Ousley and Jantz 1996), the potential effects of secular changes on sexual dimorphism have also been ignored. Univariate sex determination methods such as those presented by Stewart (1979) are particularly susceptible to secular change in stature (and overall body size) since univariate methods are size-specific as much as they are population-specific.

Thieme and Schull (1957) are exclusively responsible for the second characteristic: the use of a discriminant function as a statistical approach for sex

---

repeated here.

determination (for example, Giles and Elliot 1963, Giles 1964, Dibennardo and Tailor, 1979, Dibennardo and Tailor 1981, Schulter-Ellis et al 1983, France 1983, İşcan and Miller-Shaivitz 1984, Holland 1991, Holman and Bennett 1991, Marino 1995, Smith 1997). Discriminant function analysis continues to be the preferred statistical approach for developing sex determination methods even though logistic regression can be more robust when predicting a binary dependent variable such as sex (Norušis 1990, Saunders and Hoppa 1997).

The early popularity of discriminant function analysis is likely due to two reasons. First, the metric methods in general (including discriminant functions) were considered as a more objective approach that required less expertise than morphological assessments of sex (Giles 1970b). Second, discriminant functions are much easier to calculate manually than logistic regression. In the decades before mainframe computers were available at various institutions, calculations had to be performed manually. Thieme and Schull (1957) provide detailed step by step methodology for these manual calculations of discriminant functions. After access to mainframes was more readily available and the subsequent explosion in the availability of desktop computers with statistical packages, discriminant function analysis continued to be used because it is not restricted to predicting binary dependent variables as with logistic regression. Because race is still such a dominant concept in physical anthropology and racial differences in sexual dimorphism are assumed, discriminant functions remained the engine of various identification methods including FORDISC. The assumption is that there are racial differences in sexual dimorphism, and thus an unknown individual must be allocated to one of four (or occasionally more groups): "Black" female, "White" female, "White" male, or "Black" male. The problem is not with the use of discriminant function analysis since the approach is statistically robust when various statistical assumptions are met (Norušis 1990). The problem is with underlying methodological reasons for using discriminant function analysis: pre-categorization of samples into racial categories when developing sex determination methods.

The last characteristic of the gold standard is seen in the earliest metric sex determination methods and is most affected by the racial thinking that dominated physical anthropology in the first half of the 20th century. The emphasis on racial differences from the period was in part responsible for the approach used by Washburn, and the apparent power of the ischium-pubis index in determining sex reinforced the mainstream racial views. Since the end of the 19th century and throughout the 20th century, all samples were religiously divided by "race" or if only one "race" is sampled then "race" is used to define the sample (for example, Dwight 1894, Dorsey 1897, Dwight 1905, Giles and Elliot 1963, Giles 1964, Dibennardo and Tailor, 1979, Dibennardo and Tailor 1981, Schulter-Ellis et al 1983, France 1983, İşcan and Miller-Shaivitz 1984, Holland 1991, Holman and Bennett 1991, Marino 1995, Smith 1997). Washburn divided his sample from the Hamann-Todd Collection into "Negroes" and "Whites." Thieme and Schull (1957) only sampled "Negroes" from the Terry Collection and defined the group on the basis of "race."

The underlying and *unsubstantiated* assumption was and continues to be that there are racial differences in sexual dimorphism and/or that race is a source of variation. The irrelevance of race when determining sex was summed up by Giles over 30 years ago: "in many cases, *discriminant functions based on various racial samples will concur* when they are used to sex [sic] a particular individual," (1970b:102, emphasis added). Yet, "race" and sex are still discussed together and "race-specific" sex determination methods continued to be developed (for example, St. Hoyme and İşcan 1989, Ousley and Jantz 1996, Smith 1997).

The pattern for explaining the accuracy and applicability of methods reflects this emphasis on race and can be traced back to Washburn,

> the sex of over 75% of the pelves can be determined by the [sciatic] notch alone, therefore theoretically well over 95% of skeletons can be sexed using the [ischium-pubis] index and an observation of the notch...*Obviously, major racial groups must be treated separately* (1948:205; emphasis added).

Washburn does not explain why this is obvious or provide any evidence to support this assumption. Less than a decade later, Thieme and Schull state, "...we can be confident that this formula would probably give high accuracy for any other American Negro population or possibly any Negro population..." (1957:253). Thieme and Schull echo the distinctions and terminology of Terry (1932) and Todd (1921) when distinguishing between "pure Negroes" and "hybrid Negroes" that retain their "Negro" traits. Twelve years later and using a sample of "White femora" Dibennardo and Taylor concluded that,

> though our conclusions must be limited to the population sampled, there seems ample reason to suppose that femoral circumference might prove equally effective in sexing [sic] other groups. We will test this supposition as soon as we are able to collect data on a sufficient sample of North American Blacks. This will extend the forensic applicability of the methods (1979:636).

Dibennardo and Taylor suggest that not only might separate methods be required for different "races," but that the pattern of sexual dimorphism is so different between "races" that the dimensions used to determine sex for one "race" may not even be useful for developing sex determination methods for another "race." In the 1980s and the 1990s, the methodology remained the same; however, the approach to publishing methods changed. Researchers tended to select samples of "Blacks" and "Whites" and develop race-specific sex determination methods in one publication rather than in two (for example, Holland 1986, Holman and Bennett 1991, Marino 1995, Seidemann et al 1998). With these sex determination methods, an equally strong statement about the applicability of the methods was implied when race-specific methods were developed.

Differences in sexual dimorphism between various groups should be expected and may have an effect on the applicability of metric sex determination methods. However, there are some unsubstantiated assumptions about human variation that are at the heart of the methodology for developing race-specific sex determination methods. Aside from the possible social implications of the continued use of racial categories, these assumptions may seriously affect the reliability of methods. It has been assumed that a race constitutes a homogeneous group with a single pattern of sexual dimorphism and dividing a sample by race controls for differences in the pattern of sexual dimorphism. Even when qualifiers such as "American Blacks" are used, this assumption is flawed because, as mentioned above, patterns of craniometric and genetic variation are not concordant with racial categories, continental origin, or skin pigmentation (Relethford 1994, 2001, 2002; Brown and Armelagos 2001). Controlling for "race" does not control for either genotypic or phenotypic variation, and thus race-specific methods are misleading indicators for establishing the true parameters of applicability of the methods.

St Hoyme (1957) noted several decades ago that the focus on racial differences in the pelvis had been an impediment to identifying and quantifying sex differences since the end of the 19th century. More recently, in a review of sex and race determination methods that included a brief synopsis of the major American anatomical collections (Terry, Hamann-Todd, Cobb and Huntington Collections), St. Hoyme & İşcan (1989:61) cautioned that racial designations in these collections are "social or legal, not biological, assessments, based on local custom," and "there is, therefore, the possibility of mistaking results of rickets, scurvy, or other stigmata of poverty for normal expression of sex and race." Yet, the racial approach continues to dominate the methodology for developing sex determination methods, although the utility of race-specific methods is questionable. For example, Marino (1995) developed general and race-specific models for sex determination methods using the first cervical vertebra. If all that is available for analysis is a first cervical vertebra, how will a race-specific equation assist in the identification of the unknown individual? How can "race" be assessed from the cervical vertebra in order to decide which sex determination method should be applied?

CONCLUSION

The continued use of the race concept throughout the 20th century in the development of sex and age determination methods is due, in part, to the fact that "race" data were recorded for the main reference collections that are still extensively used for research: the Terry and Hamann-Todd Collections. As with Todd's work on age determination methods, the development of sex determination methods could not have been possible without large well documented collections. The methodology for the development of sex determination methods that has been followed since before World War II corresponds with a shift to the Terry Collection and/or Hamann-Todd Collection as a source of data. The assumptions of the period under which the collections were amassed is reflected in the prominent recording of "race" data. Furthermore, the collections themselves, along with how they have been used, have strengthened the emphasis on assessing variation in racial terms. This emphasis on race has reinforced the assumption regarding the patterns of variation that led to the prominent recording of race in the first place.

Just as Todd was progressive and even "radical" in his views in the pre-war period, Washburn is widely credited for ushering in the modern period of physical anthropology. Their actions before and after World War II, respectively, were explicitly anti-racist. Washburn's (1951) paper "The New Physical Anthropology" encouraged the adoption of the synthetic theory of evolution by physical anthropologists. Washburn argued that physical anthropology was little more than a "technique" and that,

> there has been no development of theory in physical anthropology itself, but the dominant attitude may be described as static, with emphasis on classification based on types. Any such characterization is oversimplified, and is intended *only to give indication of the dominant techniques, interests and attitudes of the physical anthropologist...* Much of the method was developed before the acceptance of the idea of evolution, and all of it before the science of genetics (1951:298; emphasis added).

Washburn stressed the importance of considering these evolutionary and genetic concepts, along with cultural issues for understanding human genotypic and phenotypic variation. While the language used in his arguments follows the pattern for the period, Washburn provides several examples that are still highly relevant 50 years after the paper was first published. He argued that the adoption of new genetic terminology alone would not solve any problems but instead physical anthropology,

> must change its way of doing things to conform with the modern evolutionary theory. For example, races must be based on the study of populations. There is no way to justify the division of a breeding population into a series of racial types. It is not enough to state that races should be based on genetic traits; races which can not [sic] be reconciled with genetics should be removed from consideration (Washburn 1951:299).

Following this simple yet effective criterion and considering genetic data (Brown and Armelagos 2001) and craniometric data (Relethford 1994, 2001, 2002), *all* racial categories (Black, White, etc) or continental equivalents (European, African, Asian) are not useful units of analysis for understanding patterns of human variation.

In a subsequent paper Washburn (1963) did not argue against the existence of biological races or the continued use of racial classifications; however, he strongly questioned the unsubstantiated claims that various researchers continued to make regarding the origins of races, the adaptive significance of morphology that was attributed to various races, and the *motivation* of some researchers for dividing *Homo sapiens* into races[14]. Additionally, he described the serious effects of *racism* on health and life

---

14 Washburn's (1963) paper was a version of the presidential address he delivered at the AAA Annual Meeting in Chicago, November, 1962. The published version is

expectancy, and he was highly critical of the emphasis on race for understanding the origin and evolution of the entire species of *Homo sapiens*. Despite Washburn's (1951, 1963) arguments, it is ironic that Washburn's (1948) paper on the ischium-pubis index has had such an impact in entrenching, in a much more subtle way, pre-evolutionary and pre-genetic concepts of human variation in skeletal biological research.

"much less harsh, focusing on the limited use of race as a valid object of study and the lack of scientific support for any claims of racial inferiority" and does not include a "scathing" denouncement of Coon's (1962) *The Origin of Races*, which was published earlier the same year (Caspari 2003:65-66).

CHAPTER 3

The History and Demographic Profile of the Robert J. Terry Anatomical Collection

by
David R. Hunt, Ph.D.[1] and John Albanese, M.Sc.

ABSTRACT

Dr. Robert J. Terry began collecting human skeletal remains in the St. Louis, Missouri area for research and educational purposes in 1898. He continued collecting skeletal specimens in the Anatomy Department at Washington University until his retirement in 1941. Mildred Trotter succeeded Terry as anatomy professor and continued the collecting and strove to balance the demographic distribution of the collection. In 1967, after her retirement, the collection was moved to the Smithsonian Institution's National Museum of Natural History, and as with several other well documented collections, the Terry Collection is still widely used for a diverse range of anthropological and medical research. Despite its extensive use, there has been limited discussion of the collection's history and incomplete description of holdings and associated materials of

[1] Collections Manager, Physical Anthropology Division, Department of Anthropology, MRC 112, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, (202) 786-2501 (tel), (202) 357-2208 (fax), email: HUNT.DAVID @ NMNH.SI.EDU.

this collection.

In this paper, the historical background of the collection and the collection process is described; the demographic profile of the collection, and a detailed description of the documentary and supporting data are presented; and the quality and accuracy of the available documentary data are assessed. The Terry Collection consists of 1728 individuals. Age at death ranges from 14 to 102 years, with the majority in the range of 20 to 80 years. Year of births range from 1828 to 1943. The mean age at death for males is 53 years and for females it is 58 years. The mean year of birth for males is 1880 and for females it is 1884. Terry's strict protocol for processing cadavers and for recording documentary data on the individuals still make the Terry Collection a valuable resource for anthropological and medical research.

INTRODUCTION

The Robert J. Terry Anatomical Collection is one of the most widely studied skeletal collections in North America and around the world. The Terry Collection has been used in countless theses, dissertations, articles and monographs, and has been the primary collection used in developing sex determination criteria and methods; techniques in age at death estimation; for palaeopathological identification and interpretation; as a comparative modern human reference in human evolutionary studies; and as the basis for medical and dental training and research in normal skeletal variation for bio-medical implants. The references are too numerous to list. While there are a few sources describing details of the life of Robert J. Terry (see Trotter 1966, 1981), surprisingly little has been written about the widely used collection that bears his name (Tobias 1985). And although the collection has been used extensively for research published in numerous professional journals, only cursory reports of the history and demographic profile of the collection and associated materials have been published (Cobb 1952, Thompson 1982, St. Hoyme & İşcan 1989, İşcan 1990, Trotter 1981, Hunt 1999). This paper is intended to fill this void.

The first section of this paper will briefly describe Robert J. Terry's background and the influences and impetus for his collecting, and describe his methodology for collecting. This section will also include a brief discussion of the interwoven connection of Terry's efforts in systematic skeletal collection with other human anatomical collections throughout the world. In the second section, the basic demographic profile of the collection is presented along with a detailed description of the documentary and supporting data which are available for the specimens in the collection. In the third section, the quality and accuracy of the available documentary data are assessed.

HISTORY OF THE TERRY COLLECTION

The Early History of the Collection

Robert J. Terry (1871-1966) was keenly interested in human anatomy, particularly in normal and pathological variation in the skeleton. He was aware there was an absence of documented human osteological/anatomical specimens from which skeletal biology, anatomy and pathology could be investigated. This awareness was kindled in 1893 by one of his mentors, Dr. George S. Huntington (1861-1927), when Terry began his medical training at the College of Physicians and Surgeons, New York. Huntington was a strong proponent for saving skeletons for skeletal biological research from the documented human cadavers from the medical school. He amassed about 3800 skeletons from his dissection classes during his tenure at the College. Huntington's influence on this early period of physical anthropology has been overlooked and the Huntington Collection has not been extensively used for research in the past because of its once encumbered accessability. Huntington not only had a significant influence on Robert Terry, but also on two other medical students who studied under him at the end of the 19th century and early in the 20th century: Ales Hrdlička (1869-1943) and T. Wingate Todd (1885-1938). After Dr. Huntington's retirement, Hrdlička worked to have Huntington's skeletal collection transferred to the Smithsonian Institution (Hrdlička 1937). The Huntington Collection is currently located in the Department of Anthropology at the National Museum of Natural History (NMNH) and is available for research. Todd was responsible for assembling the Hamann-Todd Anatomical Collection from cadavers from the Western Reserve University Medical School, now housed at the Cleveland Museum of Natural History.

After two years in New York, Terry continued his medical training in St. Louis at the Missouri Medical College (which later became part of Washington University Medical School) where he spent the rest of his career: first as a student, then as a demonstrator of anatomy, then as an assistant professor, later as chair of the Anatomy Department, and finally as professor emeritus (Trotter 1981). On one short period away from his *alma mater* at the end of the 19th century, Terry met a second important influence. In 1898, Terry spent almost a year in Edinburgh and studied under Sir William

Turner (1868-1916) (Trotter 1981, Tobias 1985). Just as Huntington had been collecting identified skeletons in New York, Turner had amassed a large collection of identified skeletons in Edinburgh.

Tobias (1985) describes some of the conversations he had with Terry in 1956 while on a research leave in St. Louis. Terry informed Tobias that the current Terry Collection was actually the third attempt at establishing a skeletal collection. After returning from England, Terry held the position of demonstrator of anatomy and began collecting skeletons with the permission of Dr. A. V. L . Brokaw, Terry's immediate superior at the Missouri Medical College at that time (Trotter 1981). The process of collecting that was begun in 1898 was expanded in 1900, when Terry was appointed assistant professor. Unfortunately, this early collection was destroyed by fire. Recognizing the teaching and research potential of a large identified skeletal series, Terry almost immediately began a second collection. This second collection was also rendered useless, for while Terry was at Harvard in 1906-7 as an Austin Teaching Fellow, the second collection was "interfered with, dispersed, left lying around" (Terry quoted in Tobias 1985:24). Very little else is known about these first two attempts at collecting skeletons except that none of the skeletons were included in the collection in its current form. The few details that are known about these early attempts do attest to Terry's insistence on careful documentation and cataloguing. From the two failed attempts, Terry developed the protocol for collecting and documenting skeletons which he implemented after he was appointed chair of the Anatomy Department in 1910.

In the second decade of the 20th century, Terry was able to devote more time and departmental resources to collecting human skeletons from cadavers used in the medical school's anatomy classes. These cadavers were primarily obtained from local St. Louis hospital and institutional morgues. A smaller portion of the cadavers came from other institutions throughout the state of Missouri (Overholser et al 1956). There is little doubt that the majority of the collection derives from the lower socio-economic classes from St. Louis and Missouri. The cadavers predominantly consist of individuals who were not claimed by relatives at local morgues, became property of the state, and would have been buried at the Missouri taxpayers' expense. These bodies were made available to the medical school to be used for anatomical instruction, a common practice for most jurisdictions in the U.S.A. and Canada before World War II.

The Collection Process

By the end of the first decade of the 20th century, Terry had a well established uniform protocol for the collecting, cataloguing, maceration and storage of the collection. Cadavers slated for the osteological collection were used in the anatomy classes only for soft tissue dissection to preserve the bones in their whole state. In some cases the calvarium was cut for brain dissection (920 specimens) and in fewer cases the cranium was sagittally sectioned (312 specimens). In the postcrania, generally only the ribs and occasionally the sternum were cut for access to the internal organs. In a few cases vertebral bodies were sampled for histological examination. Dr. Terry was specific in the complete representation of the range of human skeletal variation. Thus, he was diligent in inclusion of all forms of "normal" individuals and did not focus on pathological specimens.

Maceration consisted of stripping the bone of as much of the soft tissue as possible without damaging the bone, soaking the skeleton in hot water for 72 hours, brushing, and then drying the bone. Each hand and foot was placed in a cotton glove with each digit in the appropriate sleeve of the glove and tied at the open end (Trotter 1981). De-greasing of the bone was accomplished by exposure to benzene vapors for a period of time to remove some of the fats using a pressurized heating unit. This unit and much of the processing was accomplished in a service shed located on the roof of the anatomy building (Roy Peterson, personal communication). Terry was explicit that he did not want the bone void of fats because he felt the bone would preserve better with some fats still present. The long-term survival of this collection despite its extensive use demonstrates Terry's foresight in the preservation of the skeletal elements by the maceration protocols.

After completion of the skeletal processing, the catalogue number was written on each bone, with the exception of tiny elements such as ear bones. Furthermore, each individual phalanx and sesamoid of the hand and foot was numbered according to its position. In most cases the hand and foot bones are colour-coded to side: red ink on bones of the right side and black ink for bones of the left side. Throughout the anatomy class and the skeletal processing careful records were kept to ensure that there was no mixing of remains from different skeletons (Trotter 1981). At the end of the anatomy course, there was strict confirmation that the anatomy students turned over the entire skeleton to the technician responsible for preparation. Each bone in the skeleton was inventoried at various points in the processing, such as before and after soaking in water and before and after the exposure to de-greasing agents.

Very few individuals from the early years of collecting (for the third collection) are still included in the present Terry Collection. Trotter (1981) notes that by about 1920, the collection numbered several hundred. However, only five individuals currently in the collection (or 0.3% of the collection) have years of death before 1920. Terry was committed to providing practical instruction in human anatomy and allowed his students access to the skeletal collections and other laboratory resources. Decades of handling by hundreds of students took their toll, and damaged and incomplete skeletons were gradually replaced. Some of these removed skeletons, along with their associated documentation are still utilized in the Osteology Laboratory of the Anthropology Department at Washington University.

Specimens which were "retired" were replaced with other specimens in the collection. But to avoid unused or blank catalogue numbers, the replacement specimen was given the retired specimen number and was denoted by the "R" suffix to the catalogue number. Clear catalogue numbers on all bones on each individual attests to the fact that none of the replacements were mixed with the specimens which originally had a given catalogue number. Replacements always have an "R" suffix on all documents and bones that carry a catalogue number. Thus, the catalogue numbers are sequential with no gaps. For example, the entire skeleton number 9 was replaced by the complete skeleton of number 9R. All accompanying documents and each bone of the replacement skeleton are labeled with the catalogue number 9R. Some of the retired skeletons were included in Terry's and Trotter's "Osteological Series" (O.S.) used for teaching purposes in the anatomy classes, while others were used by Trotter in burnt bone experiments. A few of the skeletons were sent to other institutions. Terry sent 6 complete skeletons to W. W. Howells, which are now part of the collections at the Peabody Museum, Harvard University. Thirteen crania were sent to Aleš Hrdlička and added to the NMNH collections in 1927. A series of skeletons were also sent to Raymond Dart at the University of the Witwatersrand in exchange for five Basuto from Dart that are still part of the Terry Collection (catalogue numbers 421 to 425). The impact of the addition or removal of these skeletons on the demography of collection are minor when considering the overall size of the collection and some of the modifications to the collection that followed.

The Spread of an Idea

While there has been some debate whether T. Wingate Todd at Case Western Reserve or Terry began collecting skeletons first (see for example Trotter 1981:505 and Shapiro 1939: 459), both certainly had a mutual influence on each other throughout the first four decades of the 20th century when most of the Terry Collection and almost all of the Hamann-Todd Collection were amassed. A description of the history of the Terry Collection cannot be separated from the history of other skeletal reference collections amassed in the 20th century, including the Hamann-Todd Collection and the Dart Collection. The contact between the individuals involved is important in understanding at least some of the research context that resulted in these invaluable resources for research.

T. Wingate Todd was a lecturer under Sir Grafton Eliot Smith (1871-1937), at the

University of Manchester. While at Manchester, Todd was also responsible for processing and cataloguing the skeletons Smith had acquired as part of the Nubian Archaeological Survey (Shapiro 1939). Todd returned to the U.S.A. to replace Carl August Hamann (1868-1930) in 1912 as Professor of Anatomy at Western Reserve University (later Case Western Reserve University) after Hamann became dean of the medical school (Shapiro 1939, Cobb 1981). When Todd took up the appointment, he immediately began expanding Hamann's teaching collection which already numbered over 100 skeletons. By the time of Todd's premature death in 1938, there were over 3300 skeletons in what became one of the largest identified research collections in the world. Todd collected demographic data, and where possible, medical data for each cadaver, as well as a series of anthropometric measurements (Cobb 1952, 1981). Smith's influence on Todd is also reflected in the inclusion of over 900 ape and monkey skeletons in the Hamann-Todd Collection (Krogman 1939) since Smith was interested in primate brain evolution. Todd's drive to collect reference human skeletal series was extended to William Montague Cobb (1904-1990), one of Todd's students. After Cobb was appointed professor of anatomy at Howard University in 1932, he established a skeletal collection through the anatomy department. This collection now numbers over 700 documented individuals and is located at the Cobb Laboratory at Howard University (Rankin-Hill and Blakey 1994).

A few years later, Raymond A. Dart (1893-1988) was a senior demonstrator at University College under G. E. Smith. It was Smith that prompted Dart, in 1920-21, to travel to the U.S.A. as one of the first of two Rockefeller Foreign Scholars (Tobias 1985), first to Cleveland to work with Todd for a short period of time, and then for six months with Terry in St. Louis. Impressed by Terry and Todd's approaches and the value of a large reference collection, Dart began collecting skeletons in 1923, after he was appointed chair of the Department of Anatomy at the University of the Witwatersrand, Johannesburg (Tobias 1985, 1991). By Dart's retirement in 1958, there were over 2000 skeletons in the collection (Tobias 1985).

After Dart's retirement, Dr. Philip V. Tobias named the collection after Dart, and continued the collection process. Dart's efforts influenced Dr. A. Galloway, a senior lecturer for Dart, to establish a similar skeletal collection at the Makerere College in Kampala, Uganda. One of Galloway's senior lecturers, Dr. David Allbrook, and one of Tobias's senior lecturers, Dr. Leonard Freedman, successively became heads of the Anatomy Department at the University of Western Australia in Perth, and they formed a collection of the skeletons from anatomical cadavers at that institution[2]. Thus, "in almost biblical fashion, it is possible to trace the odyssey of an idea and an ideal down the generations," (Tobias 1985:25).

Mildred Trotter's Influence on the Terry Collection

When Terry retired in 1941, Dr. Mildred Trotter (1899-1991) continued collecting skeletons until her retirement in 1967. It was under Trotter's direction that the name of the collection was changed from the Washington University Collection to the Terry Collection. Trotter also had a long career at Washington University, first as a student and

---

2 There are no longer any human skeletal collections at the Anatomy Department at the University of Western Australia.

then later as a member of the faculty (Conroy et al 1992). Trotter arrived at Washington University in 1920. She received her Masters in Science (Sc. M.) in 1921 and her Ph. D. in 1924. After a short time at Oxford on a fellowship, she returned to St. Louis in 1926 and held the position of assistant professor under Terry. After Terry's retirement, she was awarded the title of "Coursemaster of Human Anatomy" which she held until her retirement. She was promoted to the position of full professor in 1946.

Over 80% of the collection had already been amassed by the time Trotter moved into Terry's position. The most significant contribution to the collection by Mildred Trotter is her efforts to balance the collection's demographic profile. Because social and economic factors affected the number of females available for dissection in the early 20th century, few were included in the skeletal collection. Trotter (1981:507) states, "I can recall many years when it was difficult to arrange to have even one or two females among the 40 [cadavers] required for our class of 80 students." Trotter focused her collecting on "White" females, and younger individuals (Trotter 1981). Besides the selective additions to the collection in the 1950s, Trotter also "re-instated" to the collection approximately 90 skeletons to increase the number of females and younger individuals which had been removed from the Terry Collection. In one case, there are skeletons with catalogue numbers 15, 15R, 15RR, and 15RRR, each of which is a different female. Although there is no way to estimate the total number of skeletons processed during the entire collecting project, there are currently 1728 documented skeletons in the collection.

In the late 1950s and into the 1960s, the Anatomy Department at Washington University began to change the focus of its research to brain morphology and function

(Roy Peterson, personal communication). With decreased interest in continuing support for the Terry Collection, and anticipating her retirement, Trotter began correspondence with her long-time friend Dr. T. Dale Stewart (1901-1997) concerning the possibility of transferring the collection to the Smithsonian Institution for permanent curation. In this transfer in 1967, the NMNH Anthropology Department also received Dr. Trotter's collection of burnt bone from her studies of bone ash weight, as well as her collection of hair samples from her studies on hair morphology and identification (Trotter 1930, 1938, 1943; Trotter and Duggins 1948, 1950; Duggins and Trotter 1950, Trotter, Duggins, and Setzler 1956).

DEMOGRAPHY AND RESOURCES AVAILABLE FOR RESEARCH

Documentation of Individuals and Associated Resources

The Terry Collection presently consists of 1728 skeletons. Various documentary data are available at the NMNH for each individual on a series of forms. Because of the extended collection period and a shift in research focus after Terry's retirement, not all documentary forms are available for all individuals. By far the most important form is the morgue record. All individuals in the collection have a morgue record of some form. The majority of the specimens also have a dental chart and anthropometric data form. Other documents that are available for some individuals include a skeleton index card with some supplementary information (such as the date the individual was added to the collection after maceration) and a summary check list of documents available for each individual which was added after the collection was moved to the NMNH. Some documents and records were not transferred to the NMNH in 1967. Receipt of transfer, death certificate copies, anatomical catalogue cards and collection records, and Terry's and Trotter's papers and correspondences are at the Washington University Medical School in the Medical Archives, the Anatomy Department's office files, and in the Anatomy Department Morgue.

The format of the morgue record varied during the 45 years of collecting currently represented in the collection, but the same basic information was consistently collected, including name, sex, age, "race," cause of death, date of death, morgue or institution of origin, permit number, and various dates and records related to embalming and processing of the cadaver. The skeletal inventory has information regarding damaged or absent bones for each individual and sometimes includes biological age estimated from morphological traits, cranial sutures, pubic symphyses, bone texture, and arthritic lipping. Pathological and normal osteological variants observed during the autopsy appear on the bone inventory list which was completed after the maceration process. The dentition was charted for most individuals.

For over 68% (1182 specimens) of the collection there is an anthropometric data form. The format of these forms and the number of measurements taken of the cadaver varied over time. One of the earlier anthropometric forms used was developed by Martin (a version is available in Martin 1957) and is in German. In most cases where an anthropometric form is available, height and weight data, and some cranial measurements were collected. Other cadaver measurements, such as head length and chest circumference were more sporadically recorded but are available for some individuals.

Terry (1938, 1940) reported that anthropometric data were collected on a regular basis starting in 1924 but both pre-mortem and post-mortem factors significantly altered the cadavers making anthropometric measurements problematic. The most critical pre-mortem factor was considerable weight-loss related to causes of death. Two important post-mortem factors included the loss of water after death and the loss of muscle tone. Both of these post-mortem factors where further complicated by the amount of time that elapsed between death and the receipt of the cadaver at the Washington University Medical School morgue which varied from several days to several weeks. Experiments were conducted by Terry (1938, 1940) to compensate for the loss of water by injecting the cadaver with embalming fluids. Detailed notations were made on the morgue record regarding the amount and type of fluid used, and method and pressure used to inject the fluid for each individual. As both Krogman (1939) and Terry (1940) note, Todd took similar steps to try to correct for some of the post-mortem changes in cadavers in the Hamann-Todd Collection. Todd personally collected a series of anthropometric measurements on the first 2500 individuals that were included in the Hamann-Todd collection (Krogman 1939). Terry and his assistants were more selective about which anthropometric measurements were collected and devoted resources to other types of data.

Aside from the documents, other resources that are available for the Terry Collection at the NMNH include 947 cadaver photographs or photo negatives, 836 plaster death masks prepared prior to maceration, and 1089 hair samples. Skin samples were also collected but were discarded in the 1960s when renovations to the Anatomy Department

Building required the demolition of the cold room where these tissue samples were stored (Roy Peterson, personal communication). Photographs, plaster death masks and hair samples are associated by catalogue number.

Demography of the Collection

Age at death data are shown in Figure 3.1 and year of birth data are available in Figure 3.2. Age at death ranges from 14 to 102 years, with the majority in the range of 20 to 80 years. Years of birth range from 1828 to 1943. Birth year is not readily available for most individuals on the documentation that is available at the NMNH. For this study, year of birth was calculated by subtracting age at death from year of death for those individuals whose age is certain. The mean age at death for males is 53 years and for females it is 58 years. The mean year of birth for males is 1880 and for females it is 1884. Two major forces described above had an enormous impact on the demography of the collection and make the differences in age and year of birth data between males and females difficult to interpret: the source of the cadavers before World War II and the change in collecting strategy by Trotter after Terry's retirement. Figure 3.3 clearly shows the effects of both influences. Before World War II, there were fewer females than males available for anatomical instruction. After World War II, through Trotter's efforts in balancing the collection, the proportion of females in the collection is much higher than most other reference collections. The ratio of males to females is about 1.4:1.

Figure 3.1: Age at death for males and females with age certain in the Terry Collection (n = 658 females; n = 950 males; total = 1608).

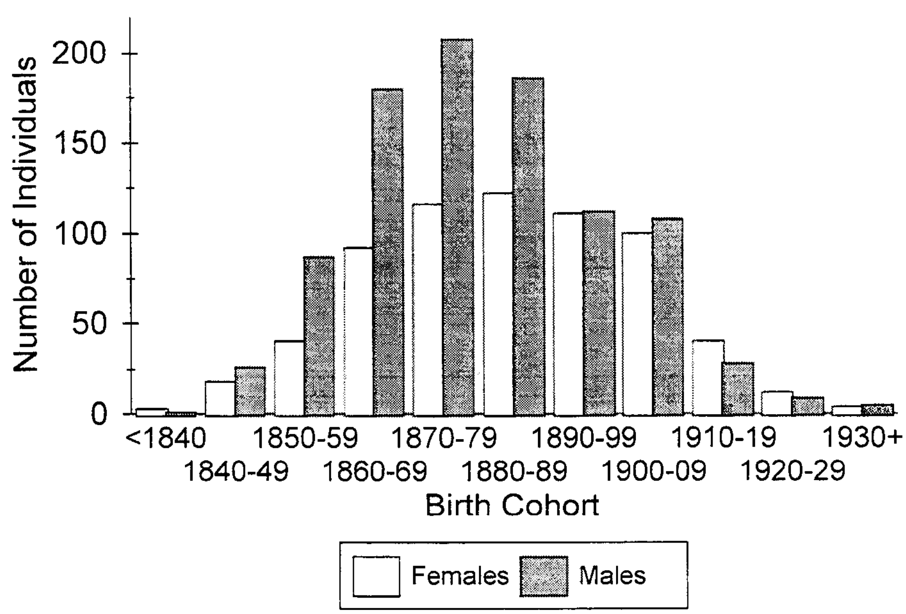


Figure 3.2: Years of birth for males and females from the Terry Collection with age certain (n = 950 males and n = 658 females; year of birth = year of death - age).

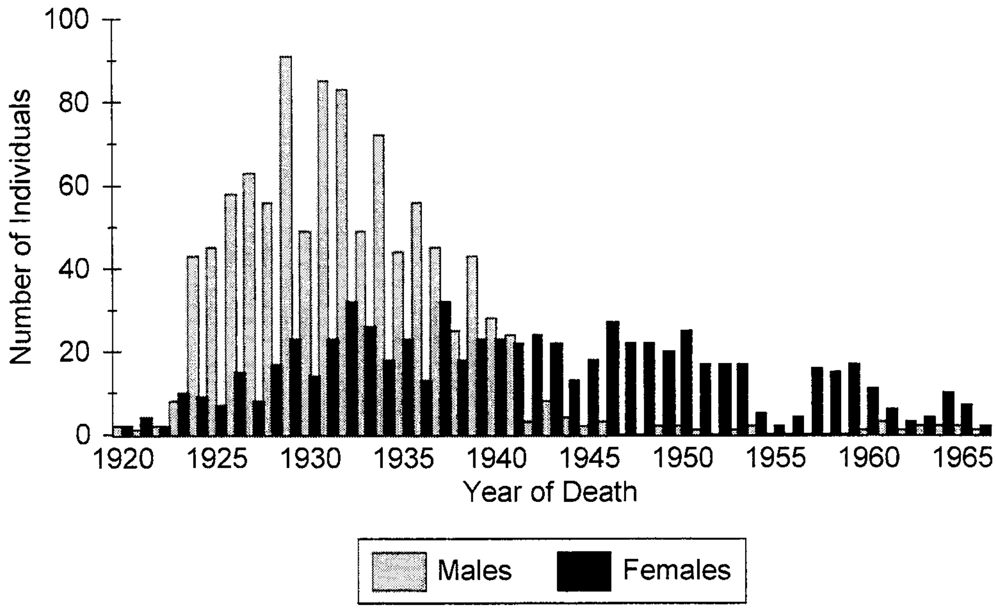


Figure 3.3: Number of males and females included in the Terry Collection by year of death (1920-1966). Note the change after Terry's retirement in 1941 (n = 949 males and 654 females).

Unfortunately, the other result of Trotter's strategy is that very few older adult males who were born after 1900 were included in the collection.

Terry's retirement had an impact on the collection in other ways. Figure 3.4 is a graph of the number of individuals for whom stature data were collected plotted by year of death. Note the sudden change around the time of Terry's retirement. Although the number of individuals for whom stature was collected had been in decline from the peak in 1932, there is a sharp decrease in measuring the cadaver after 1941. Collection of other data such as photographs/slides, death masks, hair samples, and anthropometric data follow a similar pattern (not shown graphically).

The decade after World War II was a period of great social and economic change

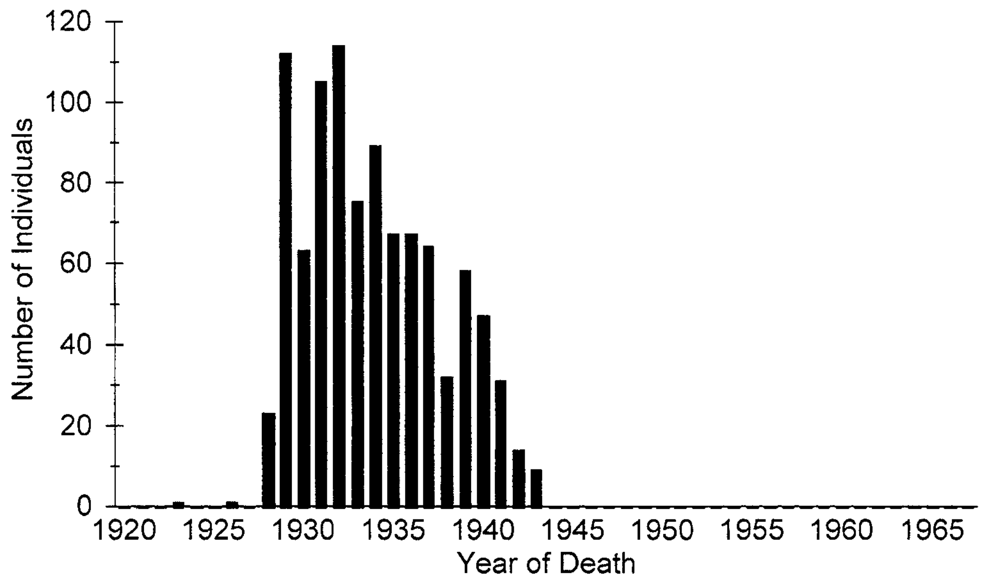


Figure 3.4: Stature data available by year of death (n = 972). After Terry's retirement in 1941, there was a sharp decrease in the collection of anthropometric data. By 1944, stature data were no longer collected although skeletons were still added to the collection by Trotter until 1966. Photographs/slides, death masks, hair samples, and other anthropometric data follow a similar pattern of availability for the collection.

in the United States. Because of the economic boom after World War II, a higher standard of living throughout the U.S.A. resulted in fewer people who were too poor to claim and bury their family members. At the same time, substantial changes in social views towards anatomical instruction and the advent of tissue transplant surgery resulted in new legislation in many North American jurisdictions allowing for testamentary bequests of bodies for scientific research (anatomical instruction, transplant surgery, etc). While it was not illegal to will one's body for scientific research in most states in the first half of the 20th century, courts in various states were divided on the issue of testamentary bequests for disposition of the body (Pregaldin 1958). The confusion most likely arose from the lack of statutes and regulations for such practices before World War II. In these transitional years in the 1950s when Trotter was instructor of human anatomy at Washington University and President of the Missouri State Anatomical Board, the demography and economic status of the pool of cadavers available for anatomical study changed (Overholser, et al 1956).

THE QUALITY OF THE DOCUMENTARY DATA AVAILABLE FOR THE COLLECTION

A documented skeletal collection is useful if documentation is credible and the records associated with each individual in the collection are consistently accurate. Two examples will be used to highlight the high quality and accuracy of the documentary data associated with each individual and in the Terry Collection as a whole. First, as noted above (and as indicated, by Terry himself) there are problems with some of the anthropometric measurements for some of the individuals in the collection, particularly the weight data. Despite Terry's and Todd's separate attempts to compensate for loss of water in both collections, there are factors related to cause of death that still present some serious problems with interpreting the anthropometric data. Terry notes,

> ...these bodies commonly bear the marks of undernourishment and in many cases of the wasting effects of chronic ailment that brought death. Whereas these conditions scarcely affect at all the longitudinal measurements they render some of the transverse and circumferential measurements of questionable value (1940:435).

Yet, the weight data available for the Terry Collection should not necessarily be abandoned. Because Terry and his colleagues had established such a thorough record-keeping protocol, some of the problems associated with the weight data can be avoided through careful sample selection. For example, if cause of death, time elapsed between death and measurement, photographic documentation and even embalming data are used to select a sample, weight may be used to arrive at meaningful results.

Recognizing the potential of length measurements, Terry devoted considerable effort to collecting stature data from a large number of cadavers that would eventually be included in the skeletal collection (Figure 3.4). Terry (1940) describes the importance of the posture of the cadaver when collecting anthropometric data, particularly (and obviously) with stature. Terry (1940) conducted several experiments to show that measurements of the cadaver varied depending on the position of the cadaver: supine or "standing." He experimented with two different measuring tables that allowed the cadaver to be placed in the standing position of a living person for stature measurement and standardized photography. Terry (1940) provides empirical data from living subjects and cadavers that show that both of these tables were useful for more closely recreating the living stature of the subject. Thus, the stature data that Terry collected very closely and consistently approximated true living stature. Trotter and Gleser (1951) did have to correct for some errors in stature measurements in 11% of their sample from the Terry Collection because they noted that in some of the photographs the soles of the feet were not flat on the baseboard of the stature measurement table. The correction of stature is possible because Terry's careful and consistent approach to documentation. He had a protocol for standardized photography of the cadavers with a clear scale in each photograph. With Terry's emphasis on documenting stature data for the collection, it is no coincidence that Trotter is co-author of several landmark publications on stature estimation (see Trotter and Gleser 1951, 1952, 1958, 1977).

The second example highlighting the quality of the documentary data of the Terry Collection is also related to the thoroughness of the record-keeping protocol and Terry's efforts to confirm various types of data from independent sources. Aside from data on sex, probably the next most important demographic field of data available for a documented collection is age at death. Terry went to great lengths to have a complete and accurate record for each individual in the collection, even after his retirement. Copies of letters sent by Terry to hospitals, coroners, and the various institutions throughout Missouri, and their replies are available at the Washington University Medical School. In these letters, Terry requested confirmation of age data, as well as information about place of birth, and occupation. For example, in a letter dated March 19, 1943, the Medical Librarian from St. Mary's Infirmary in St. Louis wrote, "the records of St. Mary's Infirmary show that on December 6, 1939, [name] died at this institution. She was a widow of 48 years of age; birthplace - Missouri; occupation - none."

In another letter to the superintendent of the Koch Hospital in Koch, Missouri, written by Terry and dated March 16, 1943, he asked for information on occupation, place of birth, and the parents place of birth (state) for 11 different individuals who died between 1921 an 1939. This letter was not signed by Terry and is likely a copy retained for his records. Correspondences such as these are evidence of Terry's rigorous approach to collecting data and cross-checking its accuracy for each individual.

Terry made efforts to clearly indicate when age data were not reliable. In the infrequent incidents when age at death could not be confirmed, age is clearly noted as an estimate with "ca." (*circa*) or as questionable with a question mark (?) next to the age on the morgue records. Consequently, these individuals can be easily excluded from a sample where accurate age at death data are necessary. Figure 3.5 is a bar graph with age at death data. Note how clear spikes at ages that are multiples of five are reduced if questionable ages are excluded.

With any documentary data, transcription errors are almost inevitable and should be expected. One of the authors (JA) reviewed 21 fields from all available documents for 356 individuals in order to assess the accuracy and reliability of the data. The data were cross-referenced between forms and with skeletal data (for age and sex), and were checked for internal consistency on each form (for example, confirming that the date of receipt of the cadaver at the morgue is after date of death). Seven errors or suspected errors were found out of over 7000 separate data (21 variables multiplied by 356 individuals). These errors were obviously transcription errors and in most cases were easily recognized and could be corrected or left out of an analysis. With any investigation involving skeletal data from the Terry Collection, the available documentary in the Terry Collection should be reviewed to identify possible transcription errors when a sample is selected for analysis.

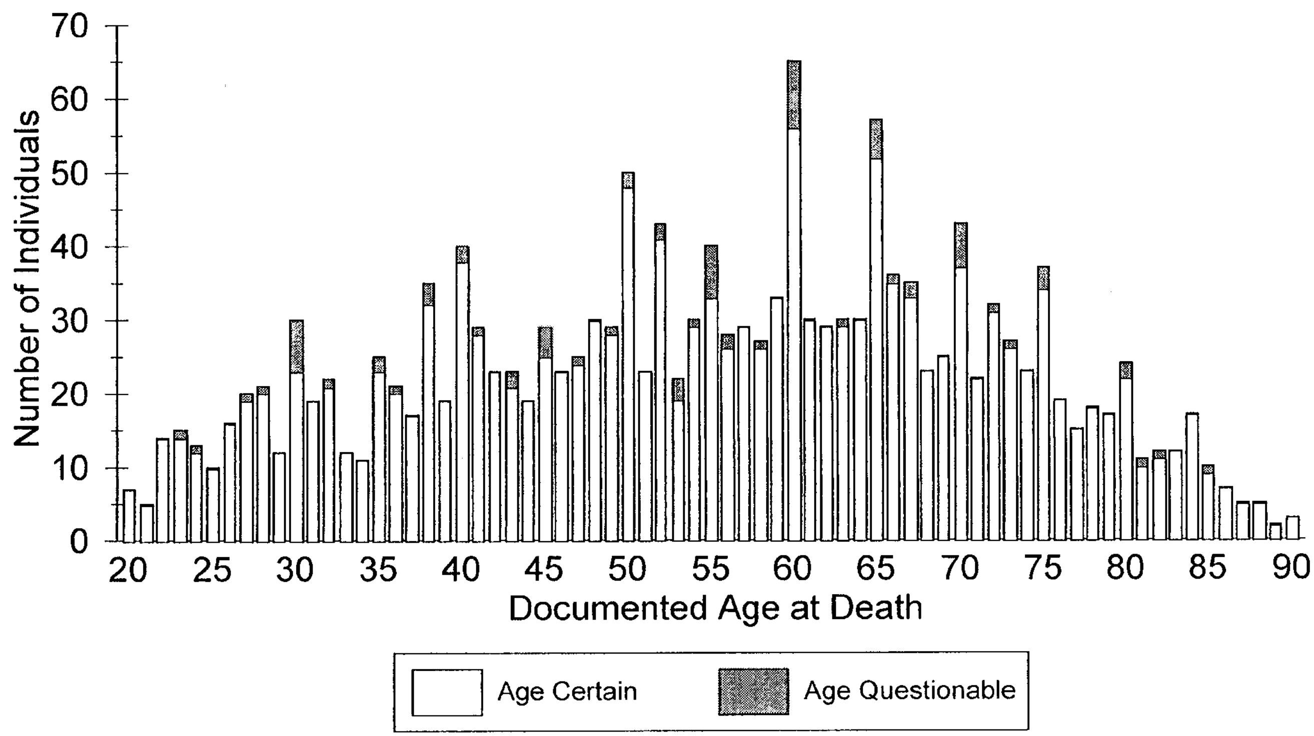


Figure 3.5: Accuracy of age at death data for the Terry Collection (n = 1608 with age certain and 84 with age questionable; only those aged 20-90 year are presented). Note how clear spikes in ages are reduced if those skeletons with questionable ages are removed from the sample.

DISCUSSION AND CONCLUSION

It has been assumed for decades that the Terry Collection and other similar collections are not representative of populations in the U.S.A. (Giles 1964) and may no longer be useful for the development of forensic identification methods (Ousley and Jantz 1998). However, with only a very few exceptions (for example, Ericksen 1982), this lack of representativeness has never been thoroughly investigated and the impact of the biases of the collection on skeletal variation has never been assessed. The origin of the cadavers is likely an important source of bias that contributes to the lack of representativeness of the Terry Collection. Terry noted several decades ago that, "at the outset a remark is in order to remind that the material of the dissection laboratory can hardly be taken as a sample of the living population from which it has been derived..." (1940:435). Other important issues have played a role in the demographic profile and skeletal variation in the Terry Collection as it is currently composed. Only a relatively small number who died after 1955, about 10% of the entire collection, were people who willed their bodies for medical research. It has been assumed that this 10% of the collection is likely more representative of middle class populations in the U.S.A. (Angel 1976, Ericksen 1982). However, Ericksen (1982) found that differences between the willed and the non-willed females were not conclusive using the proximal femur. Ericksen suggested that this lack of difference between sub-samples could be, in part, due to her assumption that the willed individuals were of a higher socio-economic status throughout their lives. In other words, they may not have been of a higher socio-economic status during their growth and development. In fact, about 20% of the willed individuals had part of their growth period during the Great Depression.

There is little doubt that those who were included in the collection before World War II came from the lowest socio-economic strata in Missouri. As is evident from Terry's quote above, he recognized that the 90% of the cadavers that are currently part of the collection were drawn from a very specific segment of the population in St. Louis and Missouri. However, the majority of individuals in the collection died during the Great Depression, and may have been of very low socio-economic status *only at the time of their deaths*, and may not necessarily have lived in poverty during their growth period. Despite the low socio-economic status and the early years of birth, about 15% of the males for whom stature data are available are 180 cm or taller. The origin of the cadavers alone, either willed or not willed, are not good indicators of socio-economic status, patterns of skeletal variation, or representativeness.

It is very likely that the Terry Collection is not representative of the living population in St. Louis in the first half of the 20th Century. The Terry Collection is not an "osteological census." The collection may not even be very representative of the pattern of mortality in St. Louis, although the mortality patterns of the Terry Collection have never been compared to those for St Louis or Missouri. All reference collections will be biased because of the collection process and not just the source of the collection, which does have some obvious limitations, which Terry pointed out several decades ago. The selective nature of the collection process does necessarily affect the demographic profile of the collection, and thus, the skeletal variation in the collection because age and year of birth are sources of variation. Trotter's shift to collecting younger females to solve a major problem that plagues most reference collections also resulted in new biases associated with racial classification and the lack of older males over 40 years of age born in the 20th century. Terry's and Trotter's approach to collecting, and the context under which they added to the collection must be considered when the representativeness of the collection is discussed. The socio-economic, and legal changes after WWII, allowed Trotter a wider choice of whom to include in the collection, while the shift in the focus of research at the Department of Anatomy at Washington University after Terry's retirement, and the historical separation of physical anthropology from anatomy throughout the middle of the 20th century, all had a clear impact as well (see Figures 3.3 and 3.4). Regardless of who was directing the collecting, some selection was occurring with respect to which cadavers were or were not included in the skeletal collection, and which skeletons remained as part of the collection.

The collection *as a whole is biased* and random sampling of the collection will result in a biased sample. For example, in a study of sexual dimorphism or development of sex determination methods, a random sample will result in a comparison of female and male sub-samples with different age and YOB effects on variation. The female sample will have proportionally more younger females born in the 20th century and the male sample will have proportionally (though not absolutely) more older individuals born in the 19th century. Calculating a mean age and YOB for the male and female sub-sample will not reveal much useful information about the distribution of age and YOB for the sample. The result will be a skewed pattern of sexual dimorphism as age and YOB effects are attributed to sex differences. Paradoxically, Trotter's efforts made the collection both more representative by increasing the proportion of females overall, but also less representative because those females tended to be younger and born in the 20th century. The higher proportion of younger females born in the 20th century is a result of the collecting process, and not because women in St. Louis were dying at a younger age in the 20th century.

The age and YOB biases will be further magnified if the sub-samples are divided by "race" because Trotter not only tried to include as many younger females as possible, but there was a clear emphasis on including younger "White" females. The age and YOB effects are even more exaggerated for each "race" and sex sub-sample, and thus, age and YOB effects may be misinterpreted as sex and "race" effects. A simple example will illustrate this sex and "race" bias: using logistic regression, it is possible to predict the "race" of females in the Terry Collection (n = 671) with an accuracy of 69.5% using only age at death and year of birth (no skeletal data) as independent variables (Albanese 2003).

There is much debate about the relevance and usefulness of the race concept in physical anthropology (for example, see Armelagos and Goodman 1998, and Ousley and Jantz 1998 for some contrasting views and uses of the race concept). Regardless of how the race concept is currently viewed by any given researcher, racial categorization of specimens is a considerable source of bias in the Terry Collection (St. Hoyme & İşcan 1989). It must be acknowledged that the "race" affinities, associated with individuals in the collection, were designated at the time of collecting between 1917-1966 and reflect the concept of race for ***that*** time (see Terry, 1932 and Todd, 1921 for their view of the race concept when these collections were amassed). The race classifications are based on social categorization and less accurately on any biological/genetic reality (Giles 1964, St. Hoyme & İşcan 1989). Furthermore, it is likely the criteria for inclusion in given racial category varied over the decades of the collection period as popular and academic views regarding race changed in the U.S.A. A simple comparison of the use of the race concept in physical anthropology from a series of publications, for example, Hrdlička (1925), Schultz (1930), Cobb (1936), St. Hoyme (1957), and other more recent uses of racial designations in *The American Journal of Physical Anthropology*, illustrates that the changes in the use and application of the race concept are complex. Despite a long history of use of racial categories by physical anthropologists to defined sub-samples from the Terry Collection, caution should be exercised when interpreting results using racially defined sub-samples in future research.

The demographic profile as a whole, and by default skeletal variation in the Terry Collection, are not representative of the living population or the mortality profile in St Louis, the state of Missouri or the entire U.S.A. in the first half of the 20th century. However, the collection as a whole does not have to be representative because the goal of most research is NOT to reconstruct the populations in the U.S.A. using Terry Collection skeletal or demographic data. If an alternative approach is pursued, and representativeness is considered in shades of grey rather than as black and white, then the level of representativeness of any collection will vary depending on the research question. Furthermore, representativeness can be maximized and biases can be minimized or even exploited in reference collections if alternative approaches to sampling are used in order to address specific research questions. Because of the size of most reference collections, the entire collection is almost never used and it is possible to sample around the various biases that are present in the collection, provided that the wealth of documentary information associated with the collections is used. Meindl and colleagues (1990) used a very selective sampling approach to overcome some of the problems with the accuracy and distribution of ages in the Hamann-Todd Collection when testing age determination methods. Despite some of the problems with the accuracy of age data for the Hamann-Todd Collection, the size of the collection allowed for the construction of a test sample of individuals with confirmed ages that was patterned after known ethnohistorical mortality profiles. With the Terry Collection, the collection process has influenced the demographic profile, and thus the variation in the collection since age at death and year of birth are sources of skeletal variation. Using data from the Terry Collection in conjunction with data from another reference collection, one of us (JA) used age and YOB criteria to sample both collections and develop a highly accurate sex determination method with an allocation accuracy of 98% (Albanese 2003). Both YOB and age criteria must be considered for sample selection when investigating sexual dimorphism using Terry Collection data. If only YOB criteria are used for sample selection in the Terry Collection, the resulting sample will have an age bias. All the individuals born in the 20th century will necessarily be younger because of the timing of the collection period and Trotter's emphasis on collecting younger individuals after Terry retired in 1941.

The Terry Collection continues to be used extensively for research in physical anthropology and other disciplines. Both the demographic profile and the historical context of collecting have had an impact on the variation in the Terry Collection and must be considered when selecting samples. The value of the Terry Collection and all reference collections comes from the documentary data. There is no doubt regarding the high quality and accuracy of the documentary data available for each individual and for the entire Terry Collection. Terry's strict protocol for processing cadavers, and just as importantly, for collecting documentary data continue to make the Terry Collection a very valuable resource for anthropological and medical research. The Terry Collection is biased, as are all reference collections; however, the collection will continue to be an invaluable source of data for many decades because it is possible to sample around the bias using the wealth of information that was documented by Robert Terry and Mildred Trotter.

ACKNOWLEDGMENTS

Research by DRH at the Washington University Medical Center library and Medical Archives and the University's Department of Anthropology were in part funded by the Department of Anthropology, NMNH. DRH is indebted to Dr. Paul Anderson, Archivist at the Washington University Medical Center Library for his help with the papers and manuscripts of Drs. Terry and Trotter; to Drs. Glen Conroy and Jane Phillips-Conroy for their help in the Department of Anatomy and their interviews; to Dr. Roy Peterson for his interview to gather first hand experiences with Drs. Terry and Trotter and concerning the history of the skeletal collection; to Dr. Tab Rasmussen for access to Washington University Anthropology Department's Osteology collections; and a personal indebtedness to Gayle King for his insight concerning the operations of the Medical Center's morgue, access to files and for relaying his personal experiences with Mildred Trotter.

Research by JA was supported, in part, by a fellowship from Social Science and Humanities Research Council of Canada (grant # 752-99-1267), by an Ontario Graduate Scholarship, and funds from the School of Graduate Studies, McMaster University. JA would like to thank Drs. Shelley Saunders and Ann Herring for their suggestions and comments, and Sandra Albanese for her technical assistance with database management.

REFERENCES CITED

Albanese J. 2003. A new metric method for sex determination using the hipbone and the femur. Journal of Forensic Sciences 48:263-273.

Angel L. 1976. Colonial to modern skeletal change in the U.S.A. Am J Phys Anthropol 45:723-736.

Armelagos GJ, Goodman AH. 1998. Race, racism, and anthropology. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p. 359-378.

Cobb WM. 1936. Race and runners. Journal of Health and Physical Education 7:1-9.

Cobb WM. 1952. Skeleton. In: Lansing AI. (Ed) Cowdry's Problems of Aging, Biological and Medical Aspects. Williams and Wilkins, Baltimore, 791-856.

Cobb WM. 1981. Thomas Wingate Todd, 1885-1938. American Journal of Physical Anthropology 56:517-520.

Conroy G, Phillips-Conroy J, Peterson R, Sussman R, Molnar S. 1992. Obituary: Mildred Trotter, Ph. D. (Feb. 2, 1899-Aug. 23, 1991). American Journal of Physical Anthropology 87:373-374.

Duggins OH, Trotter M. 1950. Age changes in head hair from birth to maturity. II. Medullation in hair of children. American Journal of Physical Anthropology 8:399-415.

Ericksen MF. 1982. How "representative" is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 59:345-350.

Giles E. 1964. Sex determination by discriminant function analysis of the mandible. Am J Phys Anthropol 22:129-136.

Hrdlička A. 1925. The Old Americans. Baltimore: Williams and Wilkins.

Hrdlička A. 1937. Biographical Memoir of George Sumner Huntington, 1861-1927, Memoirs of the National Academy of Sciences, 11, Vol.18:243-284.

Hunt DR. 1999. History and demography of the Robert J. Terry Anatomical Collection. American Journal of Physical Anthropology Supplement 28:156-7 (abstract).

İşcan MY. 1990 A comparison of techniques on the determination of race, sex, and stature from the Terry and Hamann-Todd Collections. In: Gill GW., Rhine S. (eds) Skeletal Attribution of Race: Methods for Forensic Anthropology. Maxwell Museum of Anthropology Papers, 4. University of New Mexico, Albuquerque, pp.73-81.

Krogman WM. 1939. Contributions of T. Wingate Todd to Anatomy and Physical Anthropology. American Journal of Physical Anthropology 25:145-186.

Martin R. 1957. Lehrbuch der Anthropologie. Stuttgart: Gustav Fischer Verlag.

Meindl RS, Russel KF, Lovejoy CO. 1990. Reliability of age at death in the Hamann-Todd Collection: Validity of subselection procedures used in blind tests of the summary age technique. Am J Phys Anthropol 83:349-357.

Ousley SD, Jantz RL. 1998. The forensic data bank: documenting skeletal trends in the United States. In Reichs KJ, editor, Forensic Anthropology. Second Edition. Springfield: CC Thomas, p. 441-458.

Overholser MD, Alexander WF, Trotter M. 1956 Can Missouri continue the teaching of human anatomy effectively. Missouri Medicine June:474-76.

Pregaldin AJ. 1958. Comments: Property in corpses. St. Louis University Law Journal 5:280-297.

Rankin-Hill LM, Blakey ML. 1994. W. Montague Cobb (1904-1990): Physical anthropologist, anatomist, and activist. American Anthropologist 96:74-96.

Schultz AH. 1930. The skeleton of the trunk and limbs of higher primates. Am J Phys Anthropol 2:303-438.

Shapiro HL. 1939. Thomas Wingate Todd. American Anthropologist 41:458-464.

St. Hoyme LE. 1957. The earliest use of indices for sexing pelves. Am J Phys Anthropol 15:537-546.

St. Hoyme L, İşcan MY. 1989. Determination of sex and race: Accuracy and assumptions. In İşcan MY, Kennedy KAR (ed). Reconstruction of Life from the Skeleton. Alan R. Liss, NY. pp53-93.

Terry RJ. 1932. The clavicle of the American Negro. Am J Phys Anthropol 3:351-379.

Terry RJ. 1938. Short Articles: Measuring the cadaver. American Journal of Physical Anthropology 24:235.

Terry RJ. 1940. On measuring and photographing the cadaver. American Journal of Physical Anthropology 26:433-447.

Thompson DD. 1982. Forensic Anthropology. In Spencer F, editor. A History of American Physical Anthropology 1930-1980. New York: Academic Press, p. 357-369.

Tobias PV. 1985. History of Physical Anthropology in Southern Africa. Yearbook of Physical Anthropology 28:1-52.

Tobias PV. 1991. On the scientific, medical, dental, and educational value of collections of human skeletons. International Journal of Anthropology 6:277-280.

Todd TW. 1921. 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.

Trotter M. 1930. The form, size, and color of head hair in American whites. American Journal of Physical Anthropology 14:433-445

Trotter M. 1938. A review of the classification of hair. American Journal of Physical Anthropology 24:105-126

Trotter M. 1943. Hair from Paracas Indian mummies. American Journal of Physical Anthropology 1:69-75.

Trotter M. 1966. Robert James Terry, M.D. - Obituary. American Journal of Physical Anthropology 25:97-98.

Trotter M. 1981. Robert James Terry, 1871-1966. American Journal of Physical Anthropology 56:503-508.

Trotter M, Duggins OH. 1948. Age changes in head hair from birth to maturity. I. Index and size of hair of children. American Journal of Physical Anthropology 6:489-501.

Trotter M, Duggins OH. 1950. Cuticular scale counts of hair of children. American Journal of Physical Anthropology 8:467-484.

Trotter M, Duggins OH, Setzler, FM. 1956. Hair of Australian Aborigines. American Journal of Physical Anthropology 14:649-659.

Trotter M, Gleser GC. 1951. Trends in stature of American whites and Negroes born between 1840 and 1924. American Journal of Physical Anthropology 9:427-440.

Trotter M, Gleser GC. 1952. Estimation of stature from long bones of American whites and Negroes. American Journal of Physical Anthropology 10:463-514.

Trotter M, Gleser GC. 1958. Re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. American Journal of Physical Anthropology 16:79-123.

Trotter M, Gleser GC. 1977. Corrigenda to "Estimation of stature from long bones of American whites and Negroes. American Journal of Physical Anthropology (1952)". American Journal of Physical Anthropology 47:355-356.

CHAPTER 4

Approaches for Selecting Reference Samples

INTRODUCTION

For this study, documentary, demographic, skeletal and historical data were collected from two different but complementary skeletal collections: the Terry Collection and the Coimbra Collection. The differences in structure, size, and history of the Terry and Coimbra Collections required different approaches to sampling skeletal variation in each collection. This chapter is a detailed explanation of the various approaches used to select samples from both collections to investigate the effects of age at death and year of birth, and other potential confounding factors on the pattern of skeletal variation. A new comprehensive model that combines the theory and methods of cemetery studies theory (Saunders and Herring 1995a) and the new biocultural synthesis (Goodman and Leatherman 1998a) is presented and used to select these samples.

There are three major sections in this chapter. In the first section, a brief history and description of the Coimbra Collection is presented to complement some of the information present in Chapter 3 on the Terry Collection. In the second section, a proposed comprehensive approach is presented for constructing reference samples from reference collections. In the third section, the detailed methodology for the selection of samples from the Terry and Coimbra Collections is presented.

HISTORY AND DEMOGRAPHIC STRUCTURE OF THE COIMBRA COLLECTION

There are two series of identified skulls and one series of identified full skeletons at the Museum of Anthropology at the University of Coimbra (Cunha 1995, Rocha 1995, Santos 2000). In this thesis, "Coimbra Collection" refers to the series of 505 identified skeletons: the Coimbra Identified Skeletal Collection (Colecção de Esqueletos Identificados). Individuals in the collection died between 1904 and 1936 and were excavated from the common burial ground in the Cemitério de Conchada in the City of Coimbra (Rocha 1995, Santos 2000). The type of burial alone is strong evidence of low socio-economic status, and the available information on cause of death and occupation confirms this assessment (Cunha 1995). Santos (2000) found that the majority of the sub-sample in her study (31 of 38 individuals) who died after 1926 were classified as "poor" on their hospital records and were not required to pay for their medical treatment.

Collection occurred in two distinct phases (Rocha 1995). Phase one includes individuals with ID numbers 1-72 who died between 1904 and 1912, and were exhumed around 1915. Phase two includes individuals with ID numbers 80 and greater who died between 1914 and 1936 and were exhumed before the end of 1942 (Santos 2000). The recovery of fingernails in some cases attests to the rigor in the recovery and cataloguing of remains for each individual. Individuals with ID numbers 73-79 were dissected in the Anatomical Museum at the University of Coimbra and were also interred in the cemetery in a communal grave (Santos 2000). All the skeletons in the collection are complete and well preserved except for the individuals who were dissected (Rocha 1995). The dissection procedures resulted in damage to some skeletal elements and it may be safe to assume that the communal burial may have resulted in some commingling of remains. None of these individuals who were dissected were included in the skeletal sample used in this thesis.

Years of birth (YOB) range from about 1826 to 1922, with the majority of birth dates falling between 1830 and 1910. Age at death ranges from 7 to 96 years. Birth dates are not readily available for individuals in the collection and years of birth were estimated by subtracting age from year of death. Santos (2000) recommends that an additional year should be subtracted:

YOB = year of death - age - 1 year

Santos suggests this correction because it was common for births to be registered late and ages were misreported in order to avoid a late registration fine once the ages were actually registered (Santos 2000). This misreporting would suggest that the ages for *some* individuals could also be systematically one year less than recorded. For the purposes of the various research projects in this thesis, the errors associated with YOB and age are not an issue since the sample was divided into 20 and 25 year birth cohorts and 10 year (or larger) age cohorts. Other data that are available for the collection include: nativity, sex, place of death, cause of death, occupation, parents' names as well as several serial and administrative numbers that can be used to cross-reference archival information for each individual. Santos (2000) provides more details on the demographic data, and a summary of the availability of the various fields of data for the collection.

The first series of skulls, the Medical School Collection (Colecção Escolas Médicas) that is currently curated at the Museum of Anthropology, were collected under the direction of Bernardino Luís de Machado Guimarães, the Director of the Anthropology Section of the Natural History Museum at the University of Coimbra (Santos 2000). Machados' successor Eusébio Barbosa Tamagnini de Matos Encarnação directed the collecting for the second series of skulls known as the International Exchange Collection (Colecção de Trocas Internacionais) and the Coimbra Identified Skeletal Collection (Colecção de Esqueletos Identificados) (Santos 2000).

What is known about the collection processes suggests that careful planning went into the assembly of the collection for research and teaching purposes (Santos 2000) but did not result in a random sampling of either the cemetery or the greater population of the District of Coimbra. The demographic data for the collection strongly support this assessment. The ratio of males to females is about 1.1:1 which is in contrast to the North American anatomically-derived reference collections where males greatly outnumber females. Even with the impact of Trotter's efforts to adjust the Terry Collection, the ratio of males to females is about 1.4:1. Furthermore, with the exception of the youngest and oldest cohorts where samples are small, the Coimbra Collection is distributed relatively evenly by sex across a wide range of ages. Compare Figure 4.1 with a similar figure for the Terry Collection (Figure 3.1, page 83).

Lastly, there is a brief hint of some of the greater political issues that may be associated with the construction of all three collections at the Museum of Anthropology at

Figure 4.1: Distribution of the Coimbra Collection by age and sex; n = 239 females, n = 265 males; total = 504 (age is not available for one individual in the collection of 505 individuals).

the University of Coimbra: Machado served as President of the Republic, 1915-1917 and again 1925-1926; and Tamagnini served as Minister of Public Instruction (Santos 2000). As with the North American reference collections, it seems likely that the collections at University of Coimbra would not have been possible without the efforts of a few individuals who held key positions, and had access to resources and skeletal remains.

PROPOSED MODEL FOR IDENTIFYING BIAS IN REFERENCE COLLECTIONS AND MAXIMIZING THEIR RESEARCH POTENTIAL

Figure 4.2 is a graphical representation of a proposed new model developed from cemetery studies theory (Saunders and Herring 1995a, Hoppa 1996) and the new

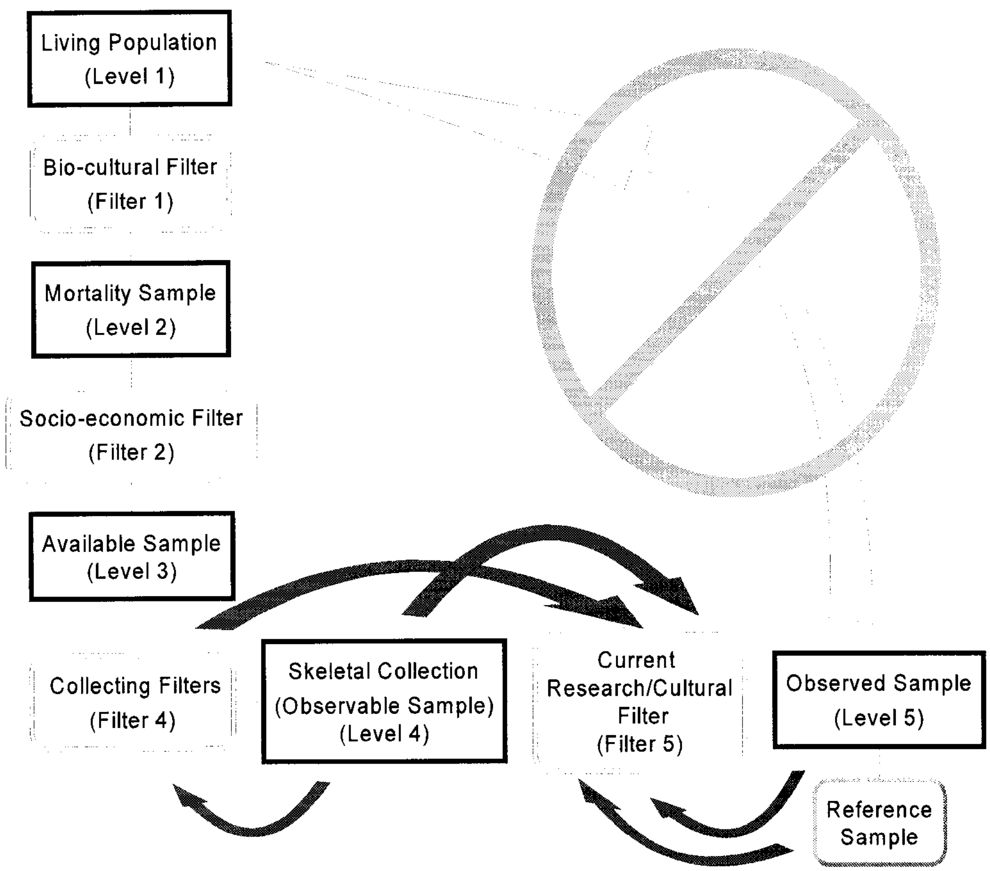


Figure 4.2: Proposed model for identifying and interpreting sources of bias and variation in reference collections (first three levels and filters are derived from Hoppa 1996:52, Figure 3.1).

biocultural synthesis (NBS) (Goodman and Leatherman 1998a) but for research involving references collections. The goal is to use a combination of theoretical and cross-disciplinary approaches to identify bias in the reference collections in order to maximize their research potential, not to discredit the collections.

Levels/Filters One to Three

Beginning at the top in Figure 4.2, the first three levels and first two filters are very similar to the levels and filters expected when working with cemetery samples. This part of the figure borrows heavily from Hoppa's graphical depiction of the filter model (1996:52, Figure 3.1). The levels (available pool of individuals) and filters are numbered to minimize confusion when referring to them and are not ranked. In these first three levels many of the reference collections mimic a highly specialized archaeological mortuary sample such as a crypt sample. Hoppa used his model to investigate the effects of these filters on the interpretations of health and demography of past populations. However, these variables (age, YOB, sex, etc.) also have an impact on the pattern of variation in the collection.

In Level 1, the living population, not everyone has the same likelihood of dying at a given age and the probability of dying is not based solely on biological issues (Hoppa 1996). The NBS emphasizes the importance of the interaction of biological and cultural factors that affect morbidity and mortality (Levins and Lewontin 1998). Poverty, restricted access to resources (when healthy or afflicted with an illness), difficult working conditions, exposure to infectious diseases, and the synergistic interaction of these and other factors may increase morbidity and mortality in different cohorts in some segments of the community.

Level 2 represents the mortality in a given community at any one time. As with cemetery samples, various socio-economic, gender and religious factors have had a filtering effect on who is more likely to die at a given age (Filter 1) *and* who is included in the specialized burial (Filter 2): crypt, dissection hall cadaver, etc. The socio-economic filters are often cited as having the greatest impact on the composition of various reference collections. However, the effects of the socio-economic filter is neither simple nor static. For example, different socio-economic issues have had various effects on the Terry Collection at various points in the collecting process. Early in the collection period, the poorest individuals, those who lacked social or economic supports in the community at the time of death, were transferred from various institutions in Missouri to Washington University for use in anatomical instruction. After the mid-1950s, there were two major contrasting socio-economic issues. First, as in the inter-war period, individuals without economic or social supports at the times of their death were still being used for anatomical instruction. However, in the post-World War II period, the economic situation in Missouri had improved to the point where very few cadavers were available from this source (Trotter 1981). At the same time, changes in social views towards dissection and anatomical instruction along with changes in legislation and regulation allowed the bequeathing of human remains for anatomical instruction in Missouri (Trotter 1981). In contrast, collecting for the Hamann-Todd Collection ended in the late 1930s with Todd's death, prior to the sweeping legal and social changes related to views of anatomical instruction. The pattern of variation in different anatomically-derived reference collections will vary despite some of the superficial similarities in the sources of skeletons (Albanese 1997a, 1997b, 1997c). Aside from the obvious differences in communities (political, social, economic differences) that the anatomical collections were drawn from, other factors have had an impact on the collection process, the socio-demographic structure of the collections and the skeletal variation in the collections.

When the Terry Collection is compared to a cemetery-derived reference collection, such as the Coimbra Collection, the importance of unraveling socio-economic issues and their impact on variation in the collection becomes particularly important for assessing the level of representativeness of reference collections. As with most of the North American reference collections, the Coimbra Collection clearly consists of individuals of the low socio-economic status in the community (Cunha 1995, Santos 2000). It may be very misleading to assume that what may be summarized in a single indicator (low socio-economic status) has the same effects on health and skeletal variation through time in St. Louis, or in Coimbra as compared to St. Louis. Variation in the effects of differential access to resources should be expected since the network of power relations are "uniquely configured, socially and historically, in particular places at particular times" (Roseberry 1998:81).

Roseberry (1998) is critical of a comparative approach where historical context and power relations are distilled out so that the basic elements that are considered to be distinctive of the "type" can be identified and compared through time and space completely devoid of context. Although Roseberry's critique is not directed at skeletal biology or research involving reference collections, many of her criticisms are directly applicable to these spheres of research. Focusing only on the distilled essential elements of the types being compared obscures the underlying processes responsible for the differences. The observed data, the presence or absence of specific features, is accepted as the reality rather than as tangible results of underlying processes.

One of these processes that may explain some the differences between the Terry and Coimbra Collections are relates to the depth of the poverty across multiple generations. Using exposure to lead as a proxy for other insults, Schell (1997, Schell and Czerwinski 1998) proposes a model that helps to explain the cross-generational nature of power relations and poverty, and their impact on growth, development and overall health.

As noted in Chapter 3, about half the individuals currently in the Terry Collection died during the Great Depression (1929-1939) and may not necessarily have lived in poverty during their growth period. The poverty may have been acute and lasted as little as a few months or a few years before death. In contrast, the individuals in the Coimbra Collection were from poor families who were unable to maintain a family mausoleum or to pay for the extended care of a grave. These families did not have the resources for expenditures across several generations. The poverty was not only chronic throughout the individual's lifetime but also spanned several generations. Investigating and unraveling the socio-economic issues in the context of the collection period can be very useful in setting parameters for sampling the reference collections in order to maximize representativeness of the sample to address a specific research question.

As is the case in cemetery analysis, gender issues are connected to economic and social factors and have had an impact on the sex ratio in the collections (Cannon 1995). However, with reference collections derived from anatomical sources, social views of dissection and anatomical instruction have also influenced the final sex ratio in these collections. In most jurisdictions in Canada and the U.S.A. anatomical dissection was legal and grave robbing had all but stopped by the end of the 19th century; however the social stigma associated with being dissected persisted into the first decades of the 20th century (Richardson 1987, Harrington and Blakely 1997). As in the grave robbing period, individuals of lower socio-economic status continued to be the source of cadavers for anatomical instruction in the first half of the 20th century. Additionally, as in the grave robbing period, community social support networks and political power also had an impact on who was dissected (Lassek 1958, Breeden 1972, Wilf 1989, Harrington and Blakely 1997). Grave robbers tended to target Black cemeteries and individuals who lacked family or friends in the community (Wilf 1989, Harrington and Blakely 1995). Anatomists risked sparking a violent riot if it came to the attention of friends and family that their recently diseased loved one was robbed from the grave and dissected (Lassek 1958, Breeden 1972, Humphrey 1973, Wilf 1987, Richardson 1987). These views of dissection likely had an effect on the sex ratio in the anatomical collections after dissection was legalized. Dissection was considered an indignity that was too horrible to be inflicted even on *poor* White women (Wilf 1989, Ginter 2001). Additionally, there may have been some sex selectivity in anatomy departments for male cadavers over females (Ginter 2001). Some selectivity by anatomical instructors certainly had an effect on the age distribution of the anatomy-derived collections. Subadults are considered less than ideal for the instruction of general anatomy because various systems (such as the lymph system) are not fully developed. It is likely that popular perceptions of dissection rather than medical selectivity had an influence on the sex ratio of cadavers available for dissection. Trotter (1981) notes how difficult it was to make special arrangements to have at least one or two female cadavers in each anatomy class. In the case of the Terry Collection, the sex selectivity was in favour of a higher proportion of females.

This connection to 19th century issues of anatomical dissection were hinted at by

Terry in a paper published 40 years after he began his first skeletal collection,

> now, with schools of medicine established around the world drawing on material from all the racial stocks, *with dissection legalized*, methods for conservation much perfected and laboratories in charge of full time staffs, the opportunities for pursuing problems of physical anthropology are considerably augmented, and it remains to be seen to what extent in the future these opportunities will be seized upon and made fruitful (1940:433; emphasis added).

Only after World War II, was there a major shift in popular views of dissection, and dissection was viewed as something that could be done regularly to others besides poor, Black or immigrant, transient men.

When dealing with cemeteries or with cemetery-derived reference collections such as the Coimbra Collection, Level 3 represents all the individuals who were interred in the common burial ground (or equivalent) of the cemetery. With reference collections derived from anatomical sources (Terry, Hamann-Todd, Grant, Huntington Collections), Level 3 is the pool of cadavers used for anatomical instruction. Below Level 3, a linear hierarchical model is not the best way to approach the investigation and identification of bias when considering skeletal reference collections.

Levels/Filters Four and Five

There is a major deviation from the cemeteries studies model illustrated by Hoppa (1996) because of the nature of Filter 4 that includes research, researcher, and greater cultural biases. In the case of reference collections, the collectors are members of the society from which the collection is derived. Furthermore, Filter 4 is affected by a feedback loop from the collection back to the Filter. There are major cultural biases that have affected reference collections: emphasis on race and outright racism in the society and popular views on dissection and anatomical instruction. Closely connected to the greater cultural biases are the research biases which include the predominance of the race concept in physical anthropology (methodology for developing sex and age determination methods) and the separation of physical anthropology from anatomy (cessation of collection of skeletal material with the retirement of specific individuals). Lastly, personal bias (including research interests) of given researchers within the context of the changing society also had an impact on the demographic structure of the collections.

One of the best examples that bridges all these levels of bias is Mildred Trotter's (1981) efforts to balance the sex ratio in the Terry Collection (reviewed in Chapter 3). As with many historic cemeteries (for example, Cannon 1995), males outnumber females in many reference collections. The ratio of males to females in the Hamann-Todd collection may be as high as 15:1 (Krogman and İşcan 1986) and in the Grant Collection the ratio is about 6.8:1[1]. Largely through Trotter's efforts, the Terry Collection was modified in structure and has a very different sex ratio of 1.4:1, with males outnumbering females to a lesser degree. Trotter did correct some of the sex bias in the collection. However, she also introduced some new biases to the collection since females and males are not matched for

---

1 The Grant Skeletal Collection was amassed in the Department of Anatomy at the University of Toronto under the direction of J. C. B. Grant between 1928 and 1950. In the mid-1980s the collection was transferred from the Department of Anatomy to the Department of Anthropology at the University of Toronto where it is currently located.

year of birth (secular change effects) and year of death (socio-economic status, among other things).

The various levels of bias (individual, field, cultural) are clearly connected since a researcher tends to conduct investigations within the parameters of the discipline which operates within a given society. Trotter's personal interests were clearly affected by the politics within an anatomy department that had been moving away from research in physical anthropology for decades and a society moving towards a wider acceptance of the use of human remains for research and education. Furthermore, within the discipline of physical anthropology, some researchers (for example, Giles and Elliot 1963) had lamented the lack of availability of "White" females in the Terry and Hamann-Todd Collections. As it became possible to include bequeathed individuals in the collection, departmental resources were not available to expand the collection at the same rate as the pre-World War II period but Trotter did correct what she and others saw as a major imbalance in the collection. The division between the personal, discipline, and cultural biases is arbitrary. However, it may be easier to dissect the biases into their component parts (while still considering the whole) in order to identify and explain some of the sources of bias.

The Terry Collection is not unique in having been significantly restructured in its demographic composition. Based on the surviving documentary evidence (identification and verification forms), the Grant Collection was much larger in the past. These collection records still exist for over 150 individuals that are no longer part of the collection. Almost all of the individuals that were removed from the collection were males whose age could not be independently verified and/or who were over 40-50 years of age. Additionally, five females (including two with age verified) were removed from the collection. This reduction in the size of the collection affected the sex ratio. Originally, males outnumbered females 10.5:1 instead of the current 6.8:1. These examples with changes in sex ratios suggest that some caution should be exercised when trying to reconstruct the sex proportions in Levels 1, 2, and 3 using Level 4 (the current composition of the collection).

In the cemetery-derived reference collections, the demographic profile of the collection has also been significantly modified from the available pool of individuals in the cemetery (Level 3). As mentioned above, the sex ratio in the Coimbra Collection seems to have been very carefully constructed and males and females seem to be age-matched (Figure 4.1).

These three interconnected levels of bias (individual, field, cultural) affected which skeletons were included in the collection. With the anatomical collections this decision was made soon after receipt of cadavers in some cases and a special dissection protocol was followed to minimize the impact on the skeleton. The single greatest procedure in the dissection process that affected skeletal research was whether the cranium was sectioned. The curators had to balance the need for anatomical instruction (brain anatomy) with the needs for future research use of the cranium after it was sectioned. At the very least, any measurements taken across a cranial cut will have an unknown error associated with them since the width of the saw is unknown. Additionally, it is possible that differential drying of the various sections may result in some warping.

The research potential of a reference collection for developing identification methods is dependant on what demographic data were documented, verified and curated for each individual in the collection. The obvious example for the anatomy-derived collections in the U.S.A. is the collection of age, sex and the prominent recording of "race." Aside from the sex and age criteria, the reasons for the inclusion of specific individuals in the Coimbra Collection are unknown. What is known is that the Coimbra Collection was established for both instructional and research purposes (Santos 2000). There are some notes in the inhumation and exhumation cemetery documents that suggest that specific individuals were selected for inclusion before exhumation (Ana Luísa Santos, personal communication). These decisions were likely based on a combination of legal issues related to the release of the remains (Santos, personal communication) and demographic criteria mentioned above.

Matters are even more complicated because Filter 4 is not necessarily static. For the Coimbra and Hamann-Todd Collection the collection period spans a relatively short period of time and may not reflect major changes in the filter. However, collecting for the Terry Collection was directed by two different collectors over several decades during a great deal of social change in the U.S.A. Views regarding the significance of race changed over the collection period: from Hrdlička (1925b) to Schultz (1930), to St. Hoyme (1957), and Thieme and Schull (1957). The ranking of "races" that might be considered overt racism by current standards was dropped by the post-World War II period, but the racial assumptions about variation were not. Furthermore, these filters were different for each sex and "race," and they varied th rough time. The reasons for including a "White" male in 1921 were very different than the reason why a few "White" males were included in the collection after World War II.

Some of the biases that were introduced in the Terry Collection (Level 4) include the coincidental correlation of "race" with age and YOB in the collection, particularly for females. Some of the bias in interpretation of the variation in the collection stems, in part, from the prominent recording of "race" on the identification documents that accompany the skeletons. A thorough, yet time consuming review of the medical school archives at Washington University (now far from the Terry Collection) would likely uncover a great deal of information for at least some individuals in the collection. A preliminary review of some archival documents[2] from Washington University suggest that data on handedness, nativity, and even parental nativity may be available for some of the individuals that are now in the collection. The emphasis on collecting parental nativity *may* be related to some of the debates that dominated physical anthropology in the first half of the 20th century such as Boas's (1912) immigrant study and Hrdlička's (1925b) research on "old Americans."

Similarly, with the cemetery-derived reference collections, certain information is curated with the collection and easily accessible. For the Coimbra Collection, nativity data are listed at the top of the identity form associated with each skeleton. These nativity

---

[2] Documents that were reviewed include copies of letters that Dr. Terry had sent to various hospitals and institutions in St. Louis; replies from record clerks, social services directors, and coroners for confirmation of identities and supplementary information; and various hand written and typed summary lists with basic information for each cadaver including name, age, year of death, cadaver number, and handedness.

data are presented at three levels: district, municipality and parish. The prominence of the nativity data may be related to the original motivation for collecting skeletal material at the University of Coimbra. As mentioned above, Tamagnini directed the collection of 1075 skulls known as the International Exchange Collection and the Coimbra Identified Skeletal Collection (Santos 2000). Part of the reason for establishing an exchange collection was for the international exchange of research specimens, but this goal was never realized (Santos 2000). In this context, the relevance of detailed nativity data becomes more clear. As with the anatomical reference collections, a great deal of documentary data are available from a number of different sources, including cemetery records and hospital records (Santos 2000). As Santos (2000) clearly demonstrates, modern research questions do not have to be limited by the filtering of documentary information that occurred when the collections were amassed.

As previously mentioned, the skeletal collections (Level 4), particulary in the U.S.A., influenced and reinforced the filters that significantly affected their structure in the first place. The under representation of young "White" females and not just young females was noted by researchers who used prominent anatomically-derived skeletal collections (Giles and Elliot 1963) while Trotter was still actively adding skeletons to the Terry Collection. Furthermore, this same racial structure of the collections had an influence on the research/cultural filters in Filter 5. The basic methodologies for developing sex and age determination methods (reviewed in Chapter 2) with divisions of samples into racial groups have changed little since the Terry and Hamann-Todd Collections became the primary sources of samples. "Race" is prominently documented for each individual along with other demographic variables, and thus, it has been assumed for decades that it must be an important source of variation, and that it must be considered and controlled. As illustrated with the historical review of the age and sex determination methods, there is clear continuity in research bias from Filter 4 to Filter 5 and supported by Level 4, the reference collection. Racial approaches to research still account for a large proportion of the publications in journals such as the AJPA (Cartmill 1998). The current methodologies reinforce the research bias and the cultural bias. Level 5 feeds back on Filter 5. In other words, the structure of the collection mirrors the social structure of the society of the collectors, and the structure of the collection has helped to maintain that social structure through the pursuits of physical anthropologists who have used the collection. A long history of selecting samples by "race" reinforces the idea that it is relevant and must be considered in future research. The collections were affected by cultural biases, field of research bias, and personal bias. But the collections and the demographic data that were retained have a feedback loop and continue to set some of the parameters of research. In a cyclical fashion, the reference collections, in conjunction with anthropological training and views about race, have limited how human variation is investigated and have greatly reinforced the underlying assumptions that researchers bring with them when using the collections.

The Reference Sample

There is a major departure in Figure 4.2 from cemetery studies theory and NBS theory. With very few exceptions (Trotter and Gleser 1951, Angel 1976, Jantz and

Meadows Jantz 1999), the reference collections have not been used to reconstruct patterns of variation of the living populations from which they were derived. Rather, reference collections have been used extensively to develop forensic identification methods, to refine the diagnoses of various diseases (such as tuberculosis in a pre-antibiotic period), and for other investigations when a reference sample has been required for comparison (such as with fossil specimens).

The lack of representativeness of some of the reference collections has been questioned because of the source of the collections (Filters 1 and 2). There can be little doubt that these filters have had an effect on the demographic structure and variation present in the collection. Additionally, Filters 4 and 5 have had an influence on the variation present and observed in the collection. However, the relevance and effects of all these filters are very different if the research goal is to construct a reference sample (for example, to develop a sex determination method) as compared to making some inferences about the original living population. The major reference collections do not have to be representative of the population they were drawn from in order to be useful for research. It is the size and nature of the reference *sample*, not necessarily the source of the reference collection, that will place limits on research on, for example, the range of applicability of an identification method. The reference collections tend to be so large and well documented that it is possible to identify the biases, sample around the biases, and construct a reference sample to address specific research questions, provided that a combination of skeletal and documentary evidence is used.

SAMPLING THE COLLECTIONS

The Terry Collection

For the purpose of this study, the Terry Collection was sampled to include a wide range of human variation related to age, YOB, sex and "race." Age and YOB criteria were used in order to assess their respective effects on sexual dimorphism. In all cases, the age of the individual was not in question based on the notes made by Terry and Trotter and based on the internal consistency of the records for each individual. The "race" criterion was used to ensure that there are about equal numbers for the sub-samples of "Blacks" and "Whites" in order to investigate whether "race" is a useful indicator of variation once age and YOB effects are considered. The analysis in Chapter 5 would not have been possible if the data had not been selected using racial criteria. A combined sample, that ignored "race," was used in the development of the sex determination methods in Chapter 6.

The last major criterion for selecting the sample is completeness of the skeleton. With very few exceptions, the skeletons in the Terry Collection are nearly complete and in good condition despite decades of handling. The one limiting factor is whether the cranium was sectioned during anatomical dissection. No measurements were taken across cuts in this study. In order to minimize the number of missing cases in multivariate analyses, an effort was made to select individuals with whole crania when a choice was available. In cases where there were two individuals that matched the age, YOB, sex, and "race" criteria and showed no evidence of gross pathological conditions, the individual without cranial damage was selected if such an individual was available.

Hypothetically, it is possible that this sampling approach for including individuals with whole crania may be a source of bias in the sample. However, the impact is very likely minimal or nonexistent for two reasons. First, the condition of the cranium was the least important selection criterion and often could not be met. About 36% (123 of 342) of the sample have cranial sections. Second, there appears to be no clear pattern for who was dissected in this manner. The age distribution for the individuals in the collection with no cranial cuts (n = 483) is very similar to the age distribution for the entire collection for both sexes.

The goal of sampling in this manner was to have at least 20 individuals in a series of sub-samples based on a combination of age, YOB, "race," and sex. The various categories are listed below in Table 4.2. Table 4.1 is a description of the abbreviation used in Table 4.2. Three sampling approaches were used depending on the availability of cases that met the selection criteria. For example, L O Wh M, means Later YOB, Older adult aged 40-79 years, "White," Male. This sub-sample was selected using Method A.

Many of the variable fields that are readily available for the individuals in the Terry Collection have been entered in a database[3]. This database was queried using the criteria mentioned above. All the documentary data available for those individuals that met the selection criteria were re-collected by the author. Method A for sample selection was used when the number of individuals that fell into each of the sub-samples was 30 or less. In this case all the individuals that matched the criteria were selected for analysis

---

[3] Special thanks to Dr. David Hunt for allowing access to this database.

regardless of the condition of the cranium.

Method B for sample selection was used when more than 30 individuals met the various criteria. In this case, the condition of the cranium was used to select 30 individuals. The cranial criteria were ranked: no sectioning, single transverse section, multiple sections. Individuals were added to the sample until a sample of 30 candidates was reached.

Method C for sample selection was used when more than 30 individuals without cranial sections met the age, sex and "race" criteria. In this case, the sample of 30 candidates was selected randomly (spreadsheet random number generator was used) from all the individuals without cranial damage. Despite an effort to minimize the number of individuals with cranial damage in the sample, about a third of the sample has at least one cranial cut. However, none of the cranial data are used in this thesis. The cranial data will be used alone and in conjunction with infra-cranial data in future research.

Data collection was divided into two phases: documentary data phase and skeletal data phase. The documentary data phase was completed first and these data were used to select the sample for skeletal analysis. Thirty individuals were selected in the documentary phase in order to have a pool from which to collect the complete skeletal data set for at least 20 individuals. It was anticipated that at least some individuals would have to be excluded from each sub-sample due to missing data because the data set collected was so large.

The result of this combination of sampling approaches is that age and YOB effects on variation can been separated for analysis. For example, there are younger and older

Table 4.1: Notes on terminology and abbreviations used in Table 4.2 for the Terry Collection sample and in the text for the Coimbra Collection.

Abbreviation	Term	Notes
E	Early	Early birth year, before 1900.
L	Later	Later birth year, in the 20th century.
Y	Younger Adult	Adult 19-39 years old. For the lower parameter, epiphyses are at least partially fused on all bones except the sternal end of the clavicle but the epiphyseal line may still be visible.
O	Older Adult	Adults 40-79 years old. The upper limit of 79 was selected to avoid missing data due to the effects of age.
Bl	Black	Racial designation on the morgue records.*
Wh	White	Racial designation on the morgue records.
M	Male	
F	Female	

*Most individuals are referred to as "Negro" in the morgue records. The database fields had already been designed with these terms and abbreviations before fieldwork began, and "Black" was used to avoid confusion in the database.

Table 4.2: Sample selection method used for each grouping.

Sampling Method	Method A	Method B	Method C
Sub-samples Selected Using Method	E Y Wh M L Y Wh M E Y Bl F E Y Wh F L Y Wh F L O Bl M L O Wh M L O BL F	E Y Bl M L Y Bl F L O Wh F L Y Bl M	E O Bl M E O Wh M E O Bl F E O Wh F

individuals born in the 19th century. If a sample is selected using only age criteria, there will be a disproportionate number of younger individuals who were also born in the 20th century because age and YOB are very closely correlated with year of death and the collecting process. In such a sample it would be impossible to separate the age effect from the YOB effect on variation.

The Coimbra Collection

Sampling the Coimbra Collection was considerably less complex since it was not necessary to construct sub-samples based on racial criteria. However, the Coimbra Collection was amassed over a much narrower time frame and has a slightly earlier distribution of YOB than the Terry Collection. Obviously if the most recent year of death in the Coimbra Collection is 1938, there can be no males or females over 40 years of age born in the 20th century. Some modifications were necessary in order to sample the collection. The goals in sampling the Coimbra Collection were the same as described for the Terry Collection: a sample with a wide, relatively even distribution of combined YOB and age criteria. Four sub-samples were selected: E O M (n = 75), E O F (n = 75), Y M (n = 50), Y F (n = 50). This distribution of sub-samples represents more than half of the adults in the collections that are between 19 and 79 years of age. All the age groups and all the YOBs after 1839 are represented. Age and YOB criteria were matched and controlled for in two large sub-samples (E O M and E O F) which can also be used for direct comparison with the corresponding Terry sub-samples. Unlike the Terry Collection, none of the individuals under 20 years of age were included in the sub-samples because they did not meet the growth criteria (see Table 4.1).

The first step in sample selection was to transcribe all the records (n = 505) into a database. The author collected all the fields available on the identity cards for each individual in the collection except the name of the individual. The establishment of a database was an opportunity to investigate the internal consistency of the documentary data. For example, the sequential identification numbers for each skeleton correspond with the date of death. Disparity between the identification number and the date of death were noted in two cases and both individuals were excluded from the study. A transcription error on the identity card available for each skeleton was the likely source of the discrepancy. Based on the review of internal consistency of the documentary data, it is easy to reconstruct that the year of death of 1936 was incorrectly transcribed as 1938 in both cases. Subsequent cross-referencing with other documentary sources (hospital records, cemetery records, etc.) confirms the death year as 1936 (Santos 2000).

The database was queried using the sub-sample criteria and the individuals for each sub-sample (50 or 75) were randomly selected from the available pool.

CONCLUSION

Despite this approach to sampling, there are some unavoidable gaps in the age and YOB criteria in samples from both collections. First, because of some of the bias introduced in the collecting process, the Terry Collection sub-samples varied considerably in size depending on the availability of individuals that met the criteria. For example, there are only three individuals that were included in the L O Bl M sub-sample and only two individuals in the E Y Wh F sub-sample. Second, because of the categorical nature of the selection process, there is some clustering of the samples from both collections around the age and YOB criteria boundaries. Despite these limitations, there is still a reasonably continuous distribution of both age and YOB for each sex, particularly when the samples from both collections are combined. Overall, the sample selection process described in this chapter did result in a wide range of age and YOB combinations being sampled from the Coimbra Collection and the Terry Collection (Figure 6.1, page 192).

Sample sizes varied with each analysis in this research depending on the skeletal measurements and the demographic variables included in the investigation. Additionally, a number of different demographic and skeletal fields of data were collected for both collections. Sample sizes, descriptions of relevant fields of data, and additional supplemental information are included with each separate analysis in subsequent chapters. Some data from the Forensic Anthropology Data Bank were also used in some analyses. The sampling approaches, sample sizes, and the fields of data analyzed are described with the respective analyses.

CHAPTER 5

Methodological Considerations Related to the Study of Secular Change Using Skeletal Data

By
John Albanese

ABSTRACT

There are several benefits to using skeletal data for the study of secular change. With skeletal data, different types of measurement error can be controlled and minimized and it is possible to investigate changes in absolute size (stature, long bone length, etc.), as well as body and limb proportions for both sexes. In studies that use skeletal data to investigate secular change, the approach in most cases has been to combine data from several sources into racial categories. Using femur length data from the Terry Collection (n = 327), the Coimbra Collection (n = 237) and the Forensic Anthropology Data Bank (n = 317), this paper illustrates and describes three aspects of sample construction that must be considered in the study of secular change using reference samples. First, combining data from different sources without confirming that the samples represent the same population can produce very misleading results. Second, dividing samples from separate sources into racial groups does not control for differences between the sources of data. There is no evidence that samples from the same data source should be divided into sub-

samples by races. The race concept is not useful in defining sub-samples for the study of human variation because it does not define biological populations nor fixed social categories. Variation through time in the social criteria for racial designation makes it problematic to use racial categories even when investigating the effects of racism on secular change related to growth, development and health using reference collections. In this context, statistical populations defined on the basis of racial criteria are not useful categories. Third, it is essential that males and females be considered separately when investigating secular change for biological reasons and for reasons related to variability over time in sex proportions in the reference collections.

INTRODUCTION

Several different sources of data can be used for the study of secular change in stature and body size including historical stature data (Steckel 1994) such as recruitment records (for example, Floud 1994, Mokyr and Gráda 1994) and stature measurements collected on cross-sectional samples specifically for research on stature (for example, Brauer 1982, Tobias 1986). Skeletal data are underutilized yet potentially useful sources of information for the study of secular change in stature as well as body proportions. With a few exceptions (Trotter and Gleser 1951, Angel 1976, Tobias and Netscher 1977, Tobias 1986, Ousley and Jantz 1998 and Meadows Jantz and Jantz 1999), osteometric data have not been used extensively for this purpose. In all of these studies a similar methodology has been used to investigate secular change using skeletal data. The purpose of this paper is to, first, critically review this methodology; and second, to use femur length data from the Robert J. Terry Anatomical Collection, the Coimbra Identified Skeletal Collection and the Forensic Anthropology Data Bank (FDB) to demonstrate that there are methodological problems with the current standard approach of combining samples from different sources into racial groups for the study of secular change using data from reference collections. Because of the similarities in the sources of data, close comparisons are made between the results from the current study and Meadows Jantz and Jantz (1999), and to a lesser extent with Trotter and Gleser (1951) and Angel (1976). However, similar concerns regarding sample construction and population parameters are applicable to other studies with similar methodology using skeletal data, particulary data from reference collections.

INVESTIGATION OF SECULAR CHANGE USING OSTEOMETRIC DATA

Long bone length data can be used to assess secular change or the long bone data can be used to estimate stature which can then be used to investigate secular change. Using stature measurements and long bone measurements, Trotter and Gleser (1951) demonstrate that trends derived from stature data and long bone data are synchronized. It is preferable to use long bone length data since the derived stature will result in additional sources of error from the stature estimation equations and efforts at correcting for the effects of age (Trotter and Gleser 1951, Tobias 1986). Additionally, long bone data may be superior to historical stature data for a number of reasons. First, changes in limb and body proportions can be investigated along with absolute changes in any given skeletal variable (Angel 1976, Meadows Jantz and Jantz 1999). Although stature measurements have been collected for centuries in North American and elsewhere as part of "an identification or registration scheme for soldiers, students, slave cargoes, oath takers, or travelers" (Steckel 1994:153), other anthropometric data are relatively scarce (a rare exception is Greiner and Gordon 1992). Second, there is precision of measurement with the use of modern, calibrated measurement equipment (Meadows Jantz and Jantz 1999). Third, accuracy related to the rounding of measurements in the historical records for the stature of any given individual or sub-sample mean is not a source of error with skeletal data. Steckel notes that "heaping, or concentration of measurements at whole feet or meters, even-numbered ages or units and at ages or units ending in zero, plagues many data sources, including some modern studies" (1994:155). Steckel does provide some evidence to suggest this is not a major problem; however, it is possible that the rounding alone may mask or magnify modest positive or negative changes in stature. Fourth, with skeletal data it is possible to test and control for intra-observer measurement error and inter-observer measurement error if more than one person is involved with collecting data. Lastly, sources of measurement error associated with age, posture, or even wearing of shoes during measurement are not an issue with long bone length data (Trotter and Gleser 1951, Brauer 1982). It is critical that the effects of age be controlled in any investigation of secular change involving living stature in a cross-sectional sample (Brauer 1982). For example, Tobias (1986) found that after correcting for the effects of age on a cross-sectional sample, a positive secular change of 17 mm over the first five decades of the 20th century was actually a negative secular change of 2 mm over the same period. Because the sample was cross-sectional, those individuals in the earliest birth cohorts were also the oldest and were affected most by age-related changes (compression of joints, posture, etc.). In an analysis of uncorrected data from a cross-sectional sample it is impossible to separate age effects on stature from secular change in stature (Trotter and Gleser 1951).

With both documented stature data and skeletal data, sampling and defining the parameters of the population can be problematic. For example, some of the biases of historical military stature data are obvious: only males are included. Other possible biases in military samples are much more complex. The level of representativeness of the military sample of the population will vary over time depending on approaches to recruiting, changes in entrance requirements, variation in socio-cultural and economic motivation for volunteering, and the phasing in and out of mandatory service (Greiner and Gordon 1992). With osteometric data, skeletal reference collections and cemetery samples are not random samples of the populations from which they were drawn. A further complication is that the acquisition and cataloguing of skeletons in reference collections has usually occurred over a short period resulting in a relatively narrow range of years of birth being included in a sample. Excavations of specific cemeteries to form collections may result in similar biases. In several cases, samples from different sources have been combined to extend the range of years of birth that are covered by any one sample. For example, Angel (1976) used data from many sources to investigate secular change in the entire skeleton in American "Blacks" and "Whites" in two periods. The first period, referred to as Colonial-American Civil War (1675-1879), was composed of samples from over two dozen family plots or small burial grounds in over a dozen states including Hawikuh Church, New Mexico; Fort Brown, Texas; Little Big Horn, Montana; West Point, New York; and various sites in Virginia and Maryland. The second period referred to as modern middle class (with years of death between 1950-1975), was composed of a small number of bequeathed individuals from the Terry Collection, and 163 forensic cases where death was accidental.

In a more recent study, Meadows Jantz and Jantz (1999) used skeletal data to investigate secular change in bone lengths and limb proportions using six long bones: humerus, radius, ulna, femur, tibia, and fibula. Following Angel's approach, they combined data from four different sources (in overlapping chronological order): the Huntington Collection, the Terry Collection, World War II Casualties, and the Forensic Anthropology Data Bank (FDB). The data from these sources were combined into two racial categories, "Black" and "White," covering a range of years of birth from 1800 to 1979. Meadows Jantz and Jantz found that secular change was significantly greater in males than females and greater in the lower limb bones than the upper limb bones. Differences were greater in "Whites" than "Blacks" but not at a statistically significant level.

Regardless of the type of data, skeletal or historical stature measurements, a critical aspect of the study of secular change is the approach to sample construction and defining the parameters of the population about which inferences can be made. There are several methodological problems that can produce misleading results when samples from different sources, which may not represent the same population, are grouped into racial categories. Amalgamating individuals from several sources into a single sample can confound interpretations of secular change because it is not clear if the different sources of data that are sampled represent the same population. There are several possible explanations or interpretations if secular change is observed in a combined sample. First, if the different samples represent the same population, the apparent change through time may be secular change due to improving or worsening conditions. Second, the change that appears to be occurring over time might be the result of genetic variation in different populations that just happen to be from different time periods. Third, the change that appears to be occurring over time might be the result of differences in living conditions between two different populations that are coincidently separate in time. Fourth, any combination of the these three explanations might also be a possibility depending on how a sample is constructed.

In contrast to Angel (1976), and Meadows Jantz and Jantz (1999), Trotter and Gleser (1951) followed a different approach. They investigated secular change using two samples in two periods: the Terry Collection (years of birth 1840-1909) and World War II casualties (1900-1924). Unlike later investigations of secular change, these authors did not combine the samples but instead compared the pattern in the different samples:

> the two periods [1840-1910 from Terry Collection and 1900 to 1924 from WW II casualties] affording data on males, jointly extend from 1840 to 1924 with some overlap at the turn of the century. However, no attempt has been made to delineate a continuous curve for either stature or bone length based on data from the Terry Collection and from military personnel. As has been suggested the *two groups are not comparable in many respects.* Nevertheless, the trends presented by separate curves for the two sources of data may be compared and are seen to differ (Trotter and Gleser 1951:437; emphasis added).

As Trotter and Gleser note, combining samples from different sources can confound interpretations; however, dividing samples from reference collections by "race" can also produce misleading results.

There is no evidence that samples should be divided into racial sub-samples. The overwhelming evidence clearly indicates that the race concept is not applicable to *Homo sapiens* and that racial categories are not useful research tools for the study of phenotypic or genotypic variation (Keita and Kittles 1997, Armelagos and Goodman 1998, Cartmill 1998, Templeton 1998, Relethford 2001, Brown and Armelagos 2001, Lieberman 2001, Relethford 2002, Molnar 2002). Classification into any racial groups has varied depending on the number of categories in the racial scheme, and the social criteria used to segregate "races" which have also varied over time and space in both popular and academic circles (Keita and Kittles 1997, Armelagos and Goodman 1998): "individuals can change their race by getting on a plane and flying from New York to Salvador or Port-au-Prince... what changes is not their physical appearance but the folk taxonomies by which they are classified," (Fish 1999:198). In a recent meta-analysis of genetic, protein and enzyme variation within and between racial groups, Brown and Armelagos (2001) demonstrate that considering samples in racial categories does not control for genetic differences. In their review, that includes several studies (Nei and Roychouldhury, 1982; Ryman et al, 1983; and Hartmann et al, 1994) where the intent was to "maximize the amount of variance accounted for by race," Brown and Armelagos (2001:34), found that *all* the studies continue to support the conclusions regarding race and genetics established over 30 years ago by Livingstone (1962) and Lewontin (1972). Using a combination of genetic evidence and craniometric data, Relethford (1994, 2001, 2002) reaffirmed these conclusions: 1) there is much more intra-race variation than inter-race variation; 2) race accounts for only about 6-13% of genetic and craniometric variation; 3) there is no concordance of human (genetic and craniometric) variation with racial categories, continental origin or skin pigmentation.

In the context of any discussion regarding skeletal reference collections, it is important to consider that terms "White," "Black," "Negroid," "Caucasoid," etc. are terms applied by the people who were collecting skeletons (or data) at the end of the 19th century and throughout the 20th century based on changing social criteria and *perceived* phenotypic criteria. The phenotypic traits are described as perceived because racial categories are in the eye of the beholder. For example, how can continuous variation be dissected into discrete categories? Or in other words, how are discrete racial groups established using the continuous variation in skin pigmentation? Using a more benign and less politically charged example, what stature separates "short" people from "tall" people?

It is problematic to assume that there is consistency in the criteria used by physical anthropologists over a century to assign any given individual to a specific racial group. It cannot be assumed that the criteria used by Dr. Huntington to classify someone in 1895 are the same criteria used by Dr. Terry in 1935, or Dr. Trotter in 1955. Even within any one collection there are likely many inconsistencies in the criteria used for racial designation. Popular and academic concepts of race changed a great deal over the six and a half decades of collecting for the Terry Collection. Racial designation is even more complex in the FDB because racial designation in the positively identified individuals in the FDB is based on self reporting on ante-mortem documents. For the Huntington collection, Dr. Huntington actually documented the nationality for individuals of European origin when it was known. It was Hrdlička (1934) who began grouping this geopolitical variation into racial categories. Hrdlička (1934) differentiated between "Whites" born in the U.S.A. and recent immigrants to the U.S.A. of European origin in the Huntington Collection after it was moved to the Smithsonian Institution. When Dr. Terry was designating race, he was trying to categorize continuous human variation (for example, Terry 1932); racial designation in the FDB is based on an individual's perceived place in a society; and racial designation in the Huntington Collection was ascribed by Hrdlička (although current researchers are not limited by his classifications). The racial terms used to describe individuals do not necessarily have the same social connotations in each of these collections or even within any one collection and it is not possible to reconstruct how any of the collectors may have used these terms when classifying any one individual. For the major American anatomical collections (including the Terry and Huntington Collections), St. Hoyme & İşcan (1989:61) describe the racial designation as "social or legal, not biological, assessments, based on local custom." Even if the same criteria for classification were used, skin colour is not a proxy for either genetic or morphological variation (Relethford 2002).

These racial categories are used in this paper to illustrate the methodological problems associated with using racial categories for the study of secular change or human variation in general. Quotation marks are used around racial terms when referring to the categorization of individuals in the various reference collections to draw attention to the point that these terms have different meanings in each of the sources of data and for different individuals within each source of data. Unless it is otherwise stated, throughout this chapter the term population refers to the statistical definition of population and not the biological meaning of the term. A biological population is group of interbreeding individuals that is relatively isolated from other similar groups in a species (Molnar 2002), whereas as statistically, "a population always means the totality of individual observations about which inferences are to be made, existing anywhere in the world or at least within a definitely specified sampling area in space and time" (Sokel and Rohlf 1973:7). The distinction between biological and statistical population is made because there is overwhelming genetic and phenotypic evidence that racial groups do not represent biological populations, but racial categories (or equivalents with different terms) are still used to define the parameters of samples and populations for research purposes.

MATERIALS AND METHODS

Sources of Skeletal Data

There is some overlap between the sample sources used by Meadows Jantz and Jantz (1999), Angel (1976), Trotter and Gleser (1951) and the present investigation. The four inquiries draw data from the Terry Collection and all but the Trotter and Gleser study use data from the FDB. For this study, data were also collected from the Coimbra Collection, a cemetery reference collection from Portugal. See Table 5.1 for details

Table 5.1: Composition of the samples from three sources by sex and “race” for each birth cohort.

	Coimbra Collection*		Terry Collection**				Forensic Data Bank**				Total			
Birth Cohort	F	M	WF	WM	BF	BM	WF	WM	BF	BM	WF	WM	BF	BM
<1875	45	46	11	20	9	14					56	66	9	14
1875-1899	59	58	20	20	44	41	1	1	1		80	79	45	41
1900-1924	15	14	39	13	48	34	11	33	18	32	65	60	66	66
1925-1949			3	7	2	2	19	52	8	19	22	59	10	21
1950+							40	52	10	18	40	52	10	18
Total	119	118	73	60	103	91	71	138	37	69	263	316	140	160

*Using current social criteria in Portugal, all the individuals in the Coimbra Collection sample would be considered “White.”

**“Race” was assigned at the time each individual was included in the collection.

F = female, M = male, WF = “White” female, WM = “White” male, BF = “Black” female, BM = “Black” male.

regarding sample sizes used in the current study. Ousley and Jantz (1998) and Meadows Jantz and Jantz (1999) describe the FDB in some detail. The FDB is located at the University of Tennessee, Knoxville, and contains data from forensic cases that were contributed by various anthropologists. The single largest contributor is Lawrence Angel (n = 182 cases), and thus, the data bank undoubtedly includes many of the individuals used by Angel (1976) in his investigation of secular change. The FDB has data from individuals from almost the entire U.S.A. with a bias towards southeastern, northeastern, and southwestern states. There are over 1700 cases in the FDB with three levels of confirmed identity that range from positive identification, tentative identification, and unidentified. About 45% of the individuals in the FDB died by homicide, suicide, or accident. Years of birth range from 1892 to 1992, although over 95% of the individuals were born in the 20th century.

The Terry Collection was collected by Drs. Robert J. Terry and Mildred Trotter at the medical school at Washington University in St. Louis, Missouri, from the end of the second decade of the 20th century until 1967. Collection practices were very different under Terry and Trotter. Terry's interest in normal human skeletal variation resulted in a large sample of individuals being included in the collection with no gross pathological conditions. After Terry's retirement in 1941, Trotter concentrated on including almost exclusively younger "White" females in order to balance the demographic profile of the collection. In the first half of the 20th century very few cadavers of females were available for anatomical instruction (Trotter 1981). This pattern changed with the enactment of legislation after World War II allowing for the bequeathing of human remains for scientific research. After Trotter's retirement, the collection, with over 1700 skeletons, was moved to the Smithsonian Institution's National Museum of Natural History in Washington D.C. in 1967. In the collection, years of birth range from 1828 to 1943 with the majority falling between 1850 and 1920.

The Terry Collection was derived from anatomy school cadavers. Most were unclaimed bodies from various hospitals and institutions from St. Louis and the State of Missouri who did not have the means to pay for their own burial. A common practice in many jurisdictions in North America before World War II was to deliver these unclaimed individuals to certified institutions for use in anatomical instruction. This fact alone strongly suggests that most individuals in the collection were of low socio-economic status, at least at the time of death. Only a relatively small number who died after 1955, about 10% of the entire collection, were people who bequeathed their bodies for scientific research. Angel (1976) drew his sample (19 females and 6 males) for the study of secular change from this small segment of the Terry Collection in an effort to sample what he described as modern middle class Americans. More detailed data on place of birth are available for a series of 107 individuals in the Terry Collection who died between July, 1926, and March, 1928. Although some of the individuals in this sub-sample were European immigrants or born outside the state, the single most common birth place is Missouri. Sixty percent of the sub-sample were born in Missouri or the semicircle of states to its south and east including Arkansas, Louisiana, Tennessee, Mississippi, Alabama, Georgia, and Kentucky.

The Coimbra Collection (n = 505), is curated at the Museum of Anthropology at the University of Coimbra, Coimbra, Portugal. This collection consists of individuals who died between 1904 and 1936 and who were excavated from the common burial ground at the Cemitério de Conchada in the city of Coimbra. The type of burial is strong evidence of low socio-economic status, and the available information on cause of death and occupation confirm this assessment (Cunha 1995). Information regarding nativity is available for 501 of 505 individuals in the Coimbra Collection. The single most frequent place of birth is the District of Coimbra, and 68% of the individuals in the collection were born in the District of Coimbra or the surrounding districts that share a border with Coimbra including Aviero and Viseu to the north, Guarda and Castelo Branco to the east, and Leiria to the south.

These three samples have overlapping ranges of years of birth that cover a period from the third decade of the 19th century to the last decade of the 20th century. The range of years of birth sampled in the current study is 1841 to 1977. The single largest difference between previous research by Meadows Jantz and Jantz (1999), and the current study is the source of data for the second and third decades of the 20th century. They use data from American World War II casualties from the Pacific Theater collected by Mildred Trotter during repatriation of the remains. This source is reflected in the large sample size (n > 460 for the femur) for "White" males for the 1910-1919 and 1920-1929 cohorts (Meadows Jantz and Jantz 1999: Table 1). A portion of these military data were originally used by Trotter and Gleser (1951) in their study of secular change.

A second difference exists between the earliest part of the range of years of birth in the four studies. In the current study, data were collected from the Coimbra Collection whereas Meadows Jantz and Jantz sampled the Huntington Collection. The Huntington Collection was amassed by George S. Huntington at the end of the 19th century and beginning of the 20th century at the College of Physicians and Surgeons, New York. After Huntington's death, the very large collection was transferred to the United States National Museum (more recently known as the Smithsonian Institution's National Museum of Natural History) in 1927[1]. There are some obvious differences between the Huntington Collection and the Coimbra Collection. The Huntington Collection, like the Terry Collection, was derived from cadavers that were used for anatomical instruction. The Coimbra Collection was excavated from a common burial ground in the cemetery in the city of Coimbra. However, for the illustrative purpose of this paper, it is argued that the Coimbra Collection is analogous to the Huntington Collection because the latter consists of "immigrants coming through Ellis Island" (Meadows Jantz and Jantz 1999:59). Thus, the Huntingdon Collection is made up of people of lower socio-economic classes who immigrated from Europe at the end of the 19th century and the Coimbra Collection consists of Portuguese people of lower socio-economic classes who remained in Europe

---

1 The exact size of the Huntington Collection has not been conclusively established. At the time of writing this chapter, the collection was being re-catalogued at the Smithsonian Institution's National Museum of Natural History in order to reassess its research potential. The collection has over 3800 catalogue numbers assigned to it, but some individuals are represented by multiple catalogue numbers, while at the same time about 30% of the catalogue numbers represent multiple individuals (David Hunt, personal communication; see also Chapter 3). Thus, a conservative estimate is that over 3000 individuals (in various levels of completeness) are represented in the collection. Despite the confusion related to the old cataloguing system, basic demographic data are not necessarily questionable on an individual level. In some cases age, sex, and other information are recorded in ink directly on specimens.

from about the same time period. Neither collection is necessarily representative of Europeans (geopolitical or racially). Rather, the point to be illustrated is that neither should be combined with "Whites" from the Terry Collection or the FDB for the study of secular change. Neither the current study, the study of Meadows Jantz and Jantz, or the study of Trotter and Gleser include data that correspond to Angel's colonial sample. Trotter and Gleser use relatively small samples from the Terry Collection for the period before 1850.

Data Collection and Sample Size

Although other researchers have used combinations of long bone lengths, cranial measurements and various indices, in this study only femur length data are used. If it can be demonstrated that there are methodological problems with combining samples from different sources into racial groups using femur length measurements then the same should be true regardless of which variable is used. The goal of the paper is to highlight the problems with how a sample is constructed and the parameters of the population about which inferences can be made, and not to investigate secular change in any given variable.

Maximum femur length was collected by the author from the Terry Collection (n = 327) and from the Coimbra Collection (n = 237), while data from the FDB (n = 317) were collected by different contributors to the FDB. The samples were divided into five 25-year birth cohorts in order to maximize the sample sizes within each cohort.

The date of birth is not readily available for most of the individuals in the Terry and Coimbra Collections. Year of birth was calculated by subtracting the age at death from the year of death for each individual in the samples. If either year of death or age at death were in any way suspect, then those individuals were excluded from the analysis. For example, for an individual from the Terry Collection, if age was preceded by "ca." (*circa*) or followed by "?" or if year of death did not correspond with other data such as date of receipt of cadaver on the morgue record available for each skeleton, then that individual was not included in the analysis. For the Coimbra Collection, the sequential identification numbers for each skeleton correspond with the date of death. Disparities between the identification number and the date of death were noted in two cases and both individuals were excluded from the analysis[2]. Individuals were included in the FDB sample only if they were positively identified. The effects of any errors when calculating year of birth in this manner are minimal in a 25-year birth cohort of adults who have completed or nearly completed their growth. For the samples from all three data sources, ages are between 18 and 80 years. Individuals from the Terry and Coimbra Collections with gross pathological conditions were excluded. Details on the composition of the three samples by sex and "race" (as described by the collectors) for each birth cohort are available in Table 5.1.

---

2 A transcription error on the identity card available for each skeleton was the source of the discrepancy. Year of death of 1936 was incorrectly transcribed as 1938 in both cases. Comparisons with other documentary sources (hospital records, cemetery records, etc.) confirm the death year as 1936 (Santos 2000).

Intra-observer and Inter-Observer Measurement Error

Femur length data were re-collected for a sub-sample of 66 individuals from the Coimbra Collection (n = 13 males and 13 female) and the Terry Collection (n = 20 males and 20 females) to assess intra-observer measurement error. Percent intra-observer measurement error was calculated using the following method:

$$\% \text{ Intra-Observer Error} = \frac{\text{Absolute Value (Measurement 1 - Measurement 2)}}{\text{Measurement 1}} \times 100$$

Percent intra-observer error was determined for each of the 66 individuals. Mean percent intra-observer measurement error is 0.06% with a range from 0 to 0.46%.

Some of the cases in the FDB (n = 135) are Terry Collection individuals who were born after 1898. Any of these individuals from the Terry Collection that are represented in the FDB and collected by this author were excluded from the FDB sample used in this study. Eight of these individuals from the FDB overlap with the intra-observer error sample collected by the author, and therefore, it is possible to assess the level of inter-observer measurement error. Following a similar approach used for the assessment of intra-observer error, the percent inter-observer error was calculated using the following equation:

$$\% \text{ Inter-Observer Error} = \frac{\text{Absolute Value (Measurement 1 - FDB Measurement)}}{\text{Measurement 1}} \times 100$$

As with the intra-observer error, the mean inter-observer measurement error is also 0.06%. The range of inter-observer error is 0 to 0.23%. Although the sample size for the assessment of inter-observer measurement is small, it follows a pattern similar to the intra-observer measurement error.

Statistical Approach

One-way analysis of variance (ANOVA) procedure was used in this study because *post hoc* tests can be used to assess significant differences between specific birth cohorts and sub-samples within each birth cohort. Statistical tests using one-way ANOVA were conducted in three separate phases using SPSS 9.0 for Windows (SPSS 1998). First, the samples from the three data sources (the Coimbra Collection, the Terry Collection and the FDB) were combined by previously defined "race" in order to look for significant differences in mean femur length over time separately for "Black" males, "White" males, "Black" females and "White" females. This first phase was an attempt to reproduce the previous approaches that have used skeletal data for the study of secular change (for example, Angel 1976, Ousley and Jantz 1998, Meadows Jantz and Jantz 1999). In the second phase of testing, data from each collection were analyzed separately in order to test for significant differences in mean femur length over time within each collection (in part, an attempt to reproduce the results of Trotter and Gleser 1951). For example, is there a significant secular change in femur length in Coimbra Collection males? In the third phase, birth cohort was held constant and variation within each cohort was compared in order to assess whether it is appropriate to divide the samples by "race." For example, are there significant differences in mean femur length between Terry Collection "Black" males, "White" males, "Black" females and "White" females, and Coimbra females and Coimbra males in the 1875-1899 cohort?

In all three phases Tukey's HSD (honestly significant difference) test was used *post hoc*. A number of different *post hoc* tests can be used with one-way ANOVA. Tukey's HSD test was selected for several reasons. First, as the name implies the test is neither too conservative (as with the Scheffe or Bonferroni tests) nor too liberal (as with the LST, least significant test) in assessing significant differences when compared to other *post hoc* tests. Second, Tukey's HSD 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). Thus, when birth cohort is held constant, it is possible to assess whether racial or other more relevant criteria should be used to define sub-samples, and it is possible to assess in which cohorts there are significant changes in long bone lengths. In Tables 5.2, 5.3, and 5.4, the significance values at the bottom of each column indicate that there are no significant differences between the sub-sample means listed in that column. These significance values should not be confused with the overall F and p values for the one-way ANOVA analysis.

This methodology and statistical analysis is in contrast to other investigations of secular change. The most common statistical approach used to investigate secular change is linear regression with some measure of time (year of birth, decade of birth, etc) as the independent variable and stature or long bone length (or combinations of long bone lengths) as the dependent variable. A different approach was used in this paper for several reasons. With linear regression, it is possible to determine whether the slope of a regression line is significant, but it is more difficult to determine when the significant secular change is occurring without additional tests such as t-tests or one-way ANOVA.

Is the change uniform in magnitude across the entire time period covered by the regression line or are there significant increases between specific birth cohorts? Or in other words, is the secular change real or is the sudden increase associated with a change in the source of data? Furthermore, with linear regression it is more difficult to assess whether the differences between two separate regression lines are significantly different. If linear regression were used from the start, the third phase of tests described above would still require one-way ANOVA. A series of one-way ANOVA tests were conducted instead of a general ANOVA model because of the very high correlation between year of birth, age at death and "race" in the Terry Collection. Race is highly correlated with year of birth and age at death and will appear as a significant factor because age and year of birth effects are attributed to racial variation. Age effects are not expected in femur length in adults who have completed their growth (Trotter and Gleser 1951), however age at death, and year of birth are highly correlated to each other, year of death, and "race" because of the collection process. Furthermore, the differences in the source of the sample (especially for the earliest cohort) for each "race" would result in the "race" factor appearing as significant because of the nature of sub-samples rather than any real "race" effect.

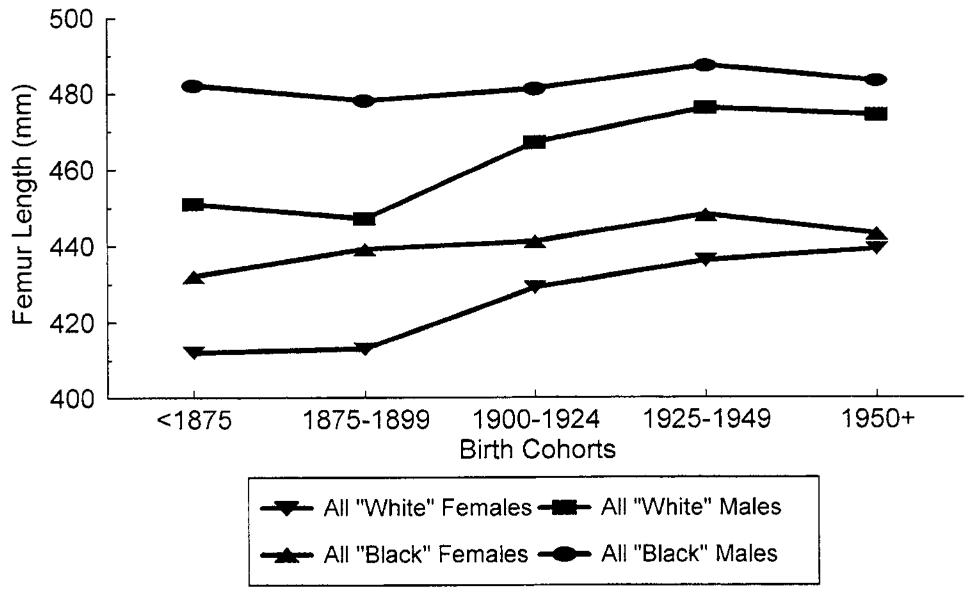
RESULTS

Phase I

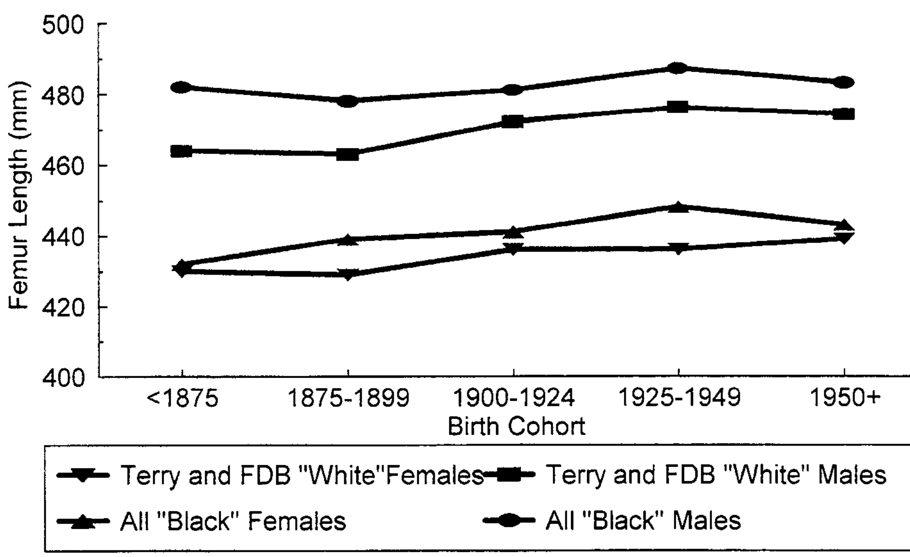
The mean femur lengths for each sub-sample spanning five birth cohorts are presented graphically in Figure 5.1a. There seems to be a positive secular change in all four sub-samples, although it is slight in some cases. Separate one-way ANOVA tests for each sex and "race" sub-sample *seem* to suggest different patterns of change over time for each "race." For "White" females, the two birth cohorts before 1900 are significantly smaller than the three cohorts after 1900 ($F = 14.168$, $p < 0.0001$). Similarly, for "White" males, the two birth cohorts before 1900 are significantly smaller than the three cohorts after 1900 ($F = 21.860$, $p < 0.0001$). "Blacks" seem to follow a different pattern than "Whites." There is no significant change in mean femur length over time for "Black" females ($F = 1.361$, $p = 0.251$) nor "Black" males ($F = 0.493$, $p = 0.741$). Homogeneous subsets for each sex and "race" sub-sample are presented in Table 5.2.

Phase II

The pattern of secular change is similar for each sex and "race" sub-sample *within* each source of data. There is no significant secular change in any of these sub-samples when examined separately. In Figure 5.2a, the means for "White" females in each birth cohort are plotted along with the means for "White" females separately by data source. The significant positive secular change in "White" females described above disappears when the sample is considered separately by data source. There are no significant changes in mean femur length in Coimbra Collection females ($F = 0.052$, $p = 0.949$), Terry Collection "White" females ($F = 0.166$, $p = 0.919$), or FDB "White" females ($F = 0.481$, $p = 0.620$). In Figure 5.2b, the means for "White" males in each birth cohort are plotted along with the means for "White" males separately by data source. As with the "White"



(a)



(b)

Figure 5.1: The mean femur lengths for each sex and "race" sub-sample by birth cohort with Coimbra Collection data (a) and without Coimbra Collection data (b). See Table 5.1 for sample sizes.

Table 5.2: Homogeneous subsets* of mean femur length for sex and "race" groups divided into five birth cohorts. All means are in mm.

"Black" Females

Cohort	n	subset 1 (mean)
<1875	9	432
1875-1899	45	439
1900-1924	66	441
1925-1949	10	448
1950+	10	443
sig.		0.483

"White" Females

Cohort	n	subset 1 (mean)	subset 2 (mean)
<1875	56	412	
1875-1899	80	413	
1900-1924	65		429
1925-1949	22		436
1950+	40		439
sig.		1.000	0.263

*Using Tukey HSD.

Table 5.2 Continued.

"Black" Males

Cohort	n	subset 1 (mean)
<1875	14	482
1875-1899	41	478
1900-1924	66	481
1925-1949	21	487
1950+	18	483
sig.		0.714

"White" Males

Cohort	n	subset 1 (mean)	subset 2 (mean)
<1875	66	447	
1875-1899	79	451	
1900-1924	60		467
1925-1949	59		476
1950+	52		474
sig.		0.915	0.156

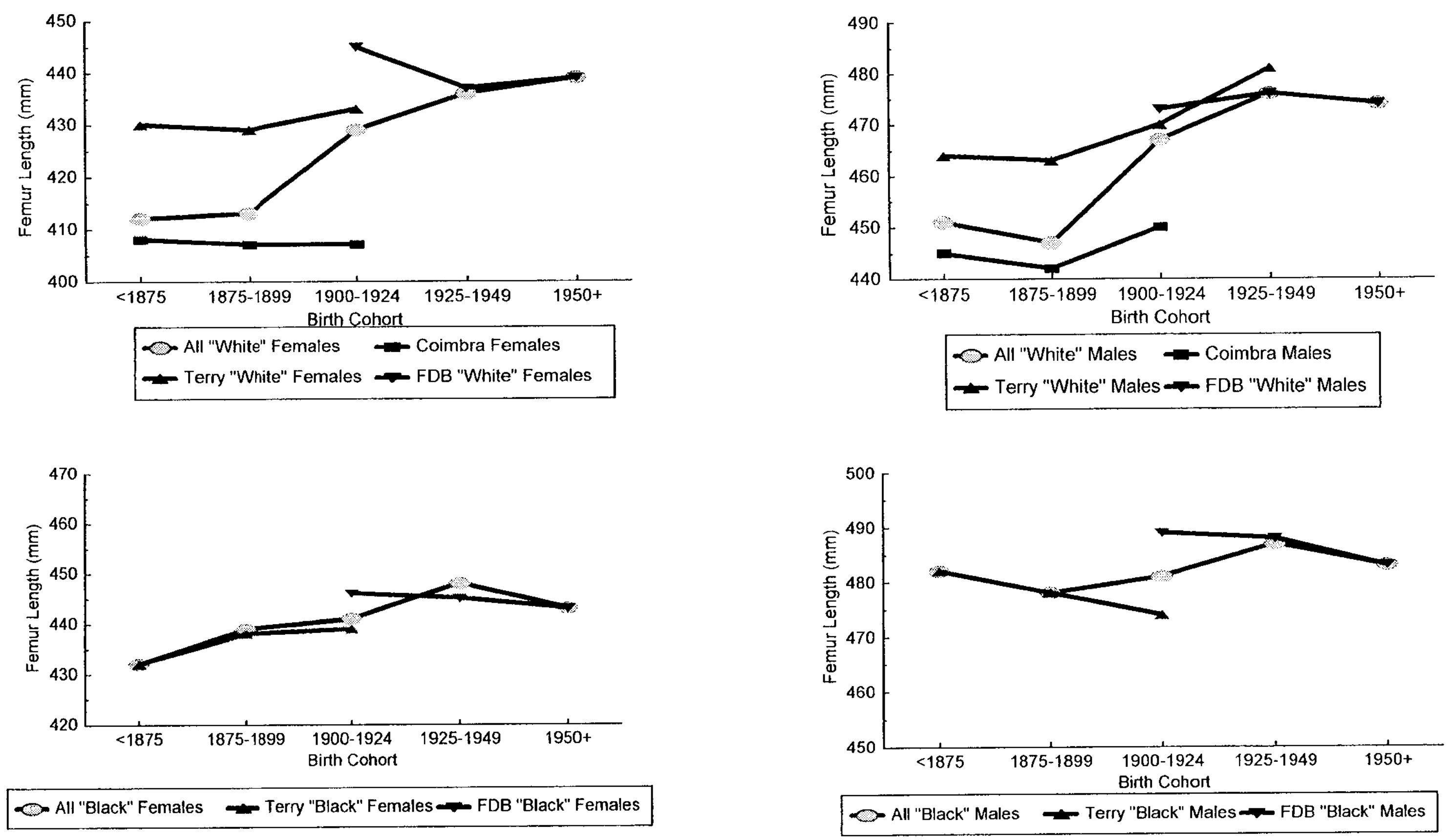


Figure 5.2: Femur length means of samples combined by "race" and separately for each source of data: Clockwise from top left: "White" females, "White" males, "Black" males, "Black" females. See Table 5.1 for sample sizes.

females, the significant positive secular change in the combined sample of "White" males disappears when the samples are considered separately by data source. There are no significant changes in mean femur length in Coimbra Collection males ($F = 1.040$, $p = 0.357$), Terry Collection "White" males ($F = 1.119$, $p = 0.349$), or FDB "White" males ($F = 0.144$, $p = 0.866$). A similar breakdown by data source for "Black" females and males are plotted in Figure 5.2c and Figure 5.2d, respectively. The lack of significant differences in the combined sample is also seen when the samples are separated by data source. There are no significant secular change in mean femur length of Terry Collection "Black" females ($F = 0.587$, $p = 0.625$) or males ($F = 0.273$, $p = 0.844$) nor are the changes significant for FDB "Black" females ($F = 0.428$, $p = 0.655$) or males ($F = 0.381$, $p = 0.684$).

Phase III

When birth cohort is held constant and the femur length means are compared, there are no significant differences between racial categories in any given birth cohort. Table 5.3 includes the homogeneous subsets of sub-samples for the 1900-1924 cohort in which all three sources of data overlap. The Coimbra females are significantly different from the "White" and "Black" females in the birth cohort ($F = 7.198$, $p < 0.0001$) and there are no significant differences between "Black" and "White" females from the Terry Collection or the FDB. The pattern is very similar for males. Coimbra males are significantly different from Terry Collection "Black" males, FDB "Black" males and FDB "White" males ($F = 5.655$, $p < 0.0001$), and there are no significant differences between "Blacks" and "Whites" from the Terry Collection or the FDB. Homogeneous subsets in Table 5.3 group Terry Collection "White" males with Coimbra Collection males and with all other males from the Terry Collection and the FDB. Stated another way, Terry Collection "White" males are not significantly different from Coimbra males nor from all other males from the Terry Collection and the Coimbra collection.

Results for the other four birth cohorts are presented in Table 5.4. The pattern is consistent for all birth cohorts except for the slight variation in the earliest cohort. In the earliest birth cohort, there are significant differences between Coimbra females and all other sex and "race" sub-samples except Terry Collection "White" females; Coimbra Collection males are significantly smaller than Terry Collection "Black" males but not Terry Collection "White" males (a pattern similar to the 1900-1924 cohort) ($F = 32.235$, $p < 0.0001$); and there are no significant differences between "Blacks" and "Whites" of either sex in the Terry Collection. In the 1875-1899 cohort, there are significant differences between Coimbra females and all other sex and "race" sub-samples; Coimbra Collection males are so small that they are not significantly different than Terry Collection females regardless of "race" ($F = 53.951$, $p < 0.0001$); and there are no significant differences between "Blacks" and "Whites" of either sex in the Terry Collection. In the two most recent birth cohorts (1925-1949 and 1950+), there are no significant differences between "Blacks and "Whites." Males from the Terry Collection and the FDB are grouped together regardless of "race" and the females from these two sources are grouped together regardless of "race."

Table 5.3: Homogeneous subsets* of mean femur length for each sex in the 1900-1924 birth cohort. All means are in mm.

Females

Group	n	subset 1 (mean)	subset 2 (mean)
Coimbra Co	15	407	
Terry Co "White"	39		433
Terry Co "Black"	48		439
FDB "White"	11		445
FDB "Black"	18		446
sig.		1.000	0.444

Males

Group	n	subset 1 (mean)	subset 2 (mean)
Coimbra Co	14	450	
Terry Co "White"	13	470	470
Terry Co "Black"	33		473
FDB "White"	34		474
FDB "Black"	32		489
sig.		0.089	0.138

*Using Tukey HSD.

Table 5.4: Homogeneous subsets* of mean femur length for four birth cohorts. All means are in mm.

<1875

Group	n	subset 1 (mean)	subset 2 (mean)	subset 3 (mean)	subset 4 (mean)
Co Fe	45	408			
Te "White" Fe	11	430	430		
Te "Black" Fe	9		432		
Co Ma	46		445	445	
Te "White" Ma	20			464	464
Te "Black" Ma	14				482
sig.		0.075	0.385	0.154	0.256

1875-1899

Group	n	subset 1 (mean)	subset 2 (mean)	subset 3 (mean)
Co Fe	59	407		
Te "White" Fe	20		429	
Te "Black" Fe	44		438	
Co Ma	58		442	
Te "White" Ma	20			463
Te "Black" Ma	41			478
sig.		1.000	0.197	0.109

*Using Tukey HSD. Te = Terry Collection, Co = Coimbra Collection, Fe = female, Ma = Male

Table 5.4 Continued.

1925-1949			
Group	n	subset 1 (mean)	subset 2 (mean)
FDB "White" Fe	19	437	
FDB "Black" Fe	8	445	
FDB "White" Ma	52		476
FDB "Black" Ma	19		488
sig.		0.695	0.407

1950+			
Group	n	subset 1 (mean)	subset 2 (mean)
FDB "White" Fe	40	439	
FDB "Black" Fe	10	443	
FDB "White" Ma	52		474
FDB "Black" Ma	18		483
sig.		0.958	0.608

DISCUSSION

Despite the differences in the sources of data (Coimbra Collection instead of Huntington Collection and lack of World War II casualty data), the results from the first phase of analysis are similar to the results of Meadows Jantz and Jantz (1999: Figure 1d). Both studies show overall positive changes in "Blacks" and "Whites" of both sexes although they are not always large increases (as with "Black" males) or significant in the current study. The greatest similarities occur in the birth cohorts starting at 1900 in both studies except for "White" females in the Meadows Jantz and Jantz study. With the exception of these "White" females, both studies show increases in femur length beginning at about 1900 (Figure 5.1a). Larger 25-year cohorts in this study make it appear as if the increase is beginning earlier when the data are presented graphically (Figure 5.1a). Despite the lack of World War II casualty data, the pattern for "White" males over the entire range of years of birth included in this study closely resembles the pattern described by Meadows Jantz and Jantz. A relatively flat or slightly negative trend changes suddenly into a large positive trend around 1900. By the fourth decade of the 20th century femur length levels off and then seems to decrease slightly.

The results from phase II of the statistical analysis clearly show that the significant positive secular change seen in "Whites" and the positive but not significant changes in "Blacks" are the result of combining samples that represent different groups with different mean femur lengths that are coincidently separated in time (Figure 5.2). It is a fair assumption that any significant positive secular change in the combined sample should also appear in each of the different sources of data when analyzed separately. For example, the significant positive change in "White" females over time should appear at least as a positive, if not significant, change when the data from the different data sources are analyzed separately. Figure 5.2a does not reflect this pattern. Although the femur length does increase slightly (3 mm) for the Terry Collection "White" females over the 85 years covered by the first three birth cohorts, the femur length in Coimbra Collection females decreases by 1 mm over a similar period and femur length of the FDB "White" females decreases by 6 mm from the 1900-1925 cohort to the most recent cohort. In all three sources of data, testing in phase II shows that slight negative and slight positive changes are not significant. The significant positive trend is clearly only an average that results from the combination of data that represent different populations rather than any real secular change in the combined sample. The pattern in Figures 5.2b and 5.2c and the lack of significance of change over time for any sub-sample within a data source suggest that a similar pattern exists for "White" males and "Black" females. The pattern for "Black" males in Figure 5.2d *appears* to be different. The two negative, though not significant, trends in the Terry Collection "Black" males and FDB "Black" males result in a false cyclical trend in mean femur length over time when the data from these two sources are combined. Not surprisingly, the flat lines for Terry Collection "White" and "Black" males mirror the results of Trotter and Gleser (1951) who also used samples from the Terry Collection.

As noted above, there is an increase in femur length around 1900 in all sub-samples described in this study and the Meadows Jantz and Jantz study except for their "White" female sample. After analyzing the sources of data separately, it is clear that in the current study the increase in femur length occurs in all sex and "race" sub-samples because the 1900-1924 cohort is the transitional cohort where the main sources of data change from the Coimbra Collection and Terry Collection around the turn of the 19th century to the Terry Collection and the FDB at the beginning of the 20th century. Similarly, the shift in the source of data to almost exclusively the FDB in the 1925-1949 birth cohort and exclusively the FDB in the most recent birth cohort results in an apparent halt in the increase in femur length or even a slight decrease in femur length (Figure 5.1a). The same pattern is visible in three of four sub-samples in the Meadows Jantz and Jantz study which suggests that the changes in the source of data are having a similar effect on their results. In the first cohort and then the fourth cohort of the 20th century, their source of data shifts as well. The shift in data source is different for males and females. For "White" and "Black" females, the shift in data source is from the Terry Collection at the end of the 19th century to the Terry Collection and FDB in the first cohort of the 20th century and then a shift to exclusively FDB data in the fourth cohort of the 20th century. Although the source of data changes in the same cohorts for "White" females, they likely follow a different pattern because of stochastic variation resulting from small sample sizes in the narrower 10-year cohorts used by Meadows Jantz and Jantz. Their sample size for "White" females fluctuates from 4 to 25 for each birth cohort from 1840-49 to 1920-29. For males, their shift in data source is from the Terry Collection to the FDB and World War II casualties. This change in data source is obvious because of the sudden jump in sample size in the 1910-1919 and 1920-1929 cohorts for "White" males (Meadows Jantz and Jantz 1999: Table 1). The sample size for "Black" males does not increase substantially over this period because there are many fewer "Black" World War II casualties in the sample collected by Trotter and Gleser (1951:429) and used by Meadows Jantz and Jantz (1999). The sudden halt in the increase of femur length in the 1930-1939 cohort for "White" and "Black" males in the Meadows Jantz and Jantz study corresponds with a second major shift in their source of data from overwhelmingly World War II casualties to exclusively FDB data. In all but one case ("White" females in the Meadows Jantz and Jantz study where sample size is very small) in both studies, the large increases and decreases in femur length, whether significant or not, correspond to changes in the sources of data. If Trotter and Gleser (1951) had also combined their data from the Terry Collection and World War II casualties, they would have found a similar sudden increase in femur length after 1900.

The first phase of analysis *seems* to suggest that there are different patterns in secular change for "Blacks" and "Whites." "White" males and females seem to show a significant secular change whereas, the apparent positive change in "Black" males and females is not significant. In Figure 5.1b, the Coimbra Collection individuals are excluded from the "White" male and female samples, and as a result, the change over time in "Whites" is very similar to the "Black" pattern. The apparent differences between "Black" and "White" females and "Black and "White" males disappear in the earliest two cohorts (compare Figure 5.1a with Figure 5.1b). As with "Black" males, the difference in mean femur length over time for this combination of "White" males that excludes the Coimbra Collection sample is not significant ($F = 2.038$, $P = 0.091$). Similarly, the difference is not significant for "White" females ($F = 0.829$, $p = 0.509$). The combination of data that included the Coimbra Collection in the "White" sub-sample is statistically significant because the average size for the Coimbra Collection males and females is much smaller and these individuals just happen to have earlier birth years on average. The Coimbra Collection individuals are so small on average that in several birth cohorts the Coimbra females are significantly smaller than the Terry Collection females and the FDB females; and Coimbra males are significantly smaller than the Terry Collection and FDB males. In fact, Coimbra males are so small that they are consistently and significantly grouped with Terry Collection and FDB females (Table 5.4). Thus, the pattern of secular change may *seem* to be different for "Blacks" and "Whites" depending on which combinations of samples are used.

Meadows Jantz and Jantz (1999:59) note, "each of the sampling frameworks [sources of data] has its own kind of bias." They go on to state that,

> the principle caveat in our results concerns sampling... In general, the female and male samples were obtained from similar sources, except for the period between 1910-1930 where there are large numbers of WWII casualties (1999:65).

This combination of factors related to the sources of data makes it impossible to interpret the cause of the greater secular change in "White" males: is it an improvement in living conditions that is greater for "White" males, sex differences in response to improved living conditions, or a statistical anomaly resulting from the inclusion of a large number of "White" males from a different data source? As Meadows Jantz and Jantz note, it is critical that males and females in a sample come from similar sources if sex differences in secular change are being investigated. However, it is absolutely essential that all the males and all the females combined into one sample from different data sources (FDB, Terry Collection, Coimbra Collection, Huntington Collection) represent the same geopolitical, socio-economic and biological population when investigating secular change.

Combining data from different sources without evidence which confirms that they represent the same population will only produce confounded or misleading results. For example, an analysis that combines the Coimbra Collection sample with the Terry Collection "Whites" provides no basis for inferences about secular change in a "White" population. Aside from the possible genetic differences in the two populations represented in these two samples, there are simply too many micro and macro socio-economic and geopolitical differences between Coimbra and surrounding districts as compared to Missouri and surrounding states. The Terry Collection (regardless of "race") and the Coimbra Collection are samples of two populations from Missouri (and surrounding states) and Coimbra (and surrounding districts), respectively, which have different biocultural histories. Any differences between the Coimbra Collection sample and the Terry Collection sample are not just due to the fact that the Coimbra Collection individuals were born a couple of decades earlier on average. Similarly, combining "Whites" from the Terry Collection (mostly Missouri and neighbouring states) with "Whites" from the Huntington Collection (mostly first generation European immigrants) who where born on average a couple of decades earlier does not reveal anything about secular change in a "White" population. It does show that the poorer immigrants who were eventually included in the Huntington Collection are smaller than poorer residents of Missouri and surrounding states who were eventually included in the Terry Collection.

Angel (1976) also does not control for other sources of variation and his sample is too small (n < 250) for such a large period of study (300 years). Angel's "White" male and female sample of 46 individuals who date from 1675 to 1879 and were excavated from numerous family plots in about a dozen states cannot be considered representative of either a single population (biological or statistical) nor representative of variation from the 18th and 19th century. It is exceedingly difficult to interpret the differences or lack of differences when comparing these individuals who date from 1675 to 1879 with a sample of 188 individuals who may be modern middle class Americans but who may or may not represent the variation present in tens of millions of people in the middle classes of the U.S in the middle of the 20th century.

The clear skeletal differences between the Coimbra and the Terry Collection when birth cohort is held constant do strongly suggest that these two samples should not be combined for the study of secular change. However, the lack of significant differences in femur length between the Terry Collection and the FDB in each of the overlapping cohorts does not necessarily suggest that the data from these two collections should be combined for the study of secular change. The trend in femur length in "Whites" and "Blacks" (Figure 5.1a and Figure 5.1b) with major changes occurring in cohorts that are transitional from one data source to the next suggest that the Terry Collection and the FDB may represent different populations. Furthermore, there are many other non-skeletal differences between these two sources of data that must be considered. The FDB is a selective sample of most of the U.S.A. whereas the Terry Collection is a selective sample of a cluster of about a half dozen states. Aside from socio-economic and geopolitical factors, complicated issues related to how the collections were amassed must also be considered. It has been stated by some authors (for example, Angel 1976, Ericksen 1982, Ousley and Jantz 1998) that the Terry Collection may not be representative of the greater American population because it was derived from a cadaver sample. The sources and magnitude of biases are actually much more complex. For example, changes in collecting practices by Trotter after Terry's retirement corrected some biases in the collection (lack of younger "White" females) while increasing others (few males born in the 20th century). These changes in collection practices are due to, first, Trotter's efforts to correct what she saw as a shortcoming in the collection, and second, a massive change in popular views of human dissection and anatomical instruction in the U.S.A. following the Second World War. The approach to collecting skeletons that was pursued by Terry and then Trotter for the Terry Collection is a very different approach to sampling skeletal variation than the approach used for the FDB which also includes a very selective sampling of the Terry Collection. The specific biases of each source of data will have a different (though not necessarily negative) impact on different research questions depending on the relative importance of age, year of birth, cause of death, and osteometric variables selected for analysis. The evidence presented in this paper does suggest that there are serious problems with combining skeletal data from different reference collections for the study of secular change. However, for some research it is very advantageous to combine data from multiple sources. For example, the differences between the Coimbra Collection and the Terry Collection can actually be exploited to construct a reference sample to develop metric sex determination methods that can be applied successfully to both small and large individuals (Albanese 2001a, Albanese 2001b, Albanese 2003).

This study shows that controlling for "race" when constructing a sample or defining the parameters of a population only further confounds the analysis of secular change. There are both social and methodological costs for using racial categories in this manner. As a methodological tool, the race concept is therefore not a useful criterion for defining sub-samples or populations for the study of secular change using reference collections. When the femur length data are presented graphically by "race," it *seems* as if "Black" males are consistently larger than "White" males and that "Black" females are consistently larger than "White" females in any given cohort. However, the results presented in Table 5.4 are very clear. In any given cohort, there are no significant differences between "Blacks" and "Whites" in the Terry Collection and the FDB and there is no evidence from the femur length data to suggest that "Whites" and "Blacks" in the Terry and FDB collections should be analyzed separately. Males are consistently grouped together regardless of "race" and females are constantly grouped together regardless of "race." The continued use of racial designation can lead to the misuse of scientific publications by third parties for the justification of social inequality even if a racial hierarchy is not set out in the original publication (Armelagos and Goodman 1998, Lieberman 2001).

In a thorough review of the concept of race in physical anthropological research, Armelagos and Goodman (1998) demonstrate clearly that race is an antiquated concept that is nothing more than a typological approach, little removed from the ethnocentrism behind concepts such as the great chain-of-being. They state,

> [this review] raises fundamental questions about the biological utility of race and the social, political, and economic factors that influence the conceptions of human differences and similarities... Race, we conclude has failed to work as a core anthropological concept; it fails to describe and explain variation. What race does, and does well is type individuals, and this typing supports the existing structures of power (1998:359).

However, they conclude that the avoidance of the use of racial categories in some cases can lead to ignoring both the biological and social costs of racism (Armelagos and Goodman 1998, see also Cartmill 1998). The study of secular change using skeletal data from different reference collections is not one of these cases. It is not safe to assume that the same criteria to describe race were used by Drs Huntington, Terry, Trotter or any of the self-reporting in the FDB. Thus, if the social criteria for classification in any given "race" have varied over time, then combining data from diverse sources that represent different populations where samples were segregated based on inconsistent social criteria reveals nothing about the impact of racism on living conditions and secular change.

While the results of this study indicate that division of samples by "race" for the study of secular change in bone size is not methodologically sound, there is no evidence to suggest that males and females should be combined for an analysis of secular change. Beyond the obvious biological basis for sexual dimorphism, there is some evidence that males and females do react differently to changes in living conditions (Greulich 1951, Greulich et al 1953, Tobias 1972, 1975; Brauer 1982, Stini 1975, 1982; Stinson 1985). It is possible that these sex differences will appear in an analysis of secular change, although 25 year birth cohorts may be too broad to detect differences related to specific, acute incidents of improvement or worsening of conditions. However, one essential reason why males and females must be analyzed separately is that equal numbers of males and females are not always available for each birth cohort when using data from reference collections. The availability of individuals of each sex in each birth cohort varies in the respective sources of data and this bias is reflected in the samples used in the present study. Although not all males are larger than all females, males are significantly larger than females in femur length and many other variables. Differences in the number of males and females from cohort to cohort would result in what might appear to be secular change.

Figure 5.3, presents combined data for all individuals from the Terry Collection sample, regardless of sex, compared with all males and all females separately. Because of Trotter's efforts after 1941 to include younger females in the Terry Collection, there are fewer males and more females in the more recent birth cohorts: in the earliest birth cohort, males ($n = 34$) outnumber females ($n = 20$); in the 1875-1899 birth cohort, the numbers of males ($n = 61$) and females ($n = 64$) are almost equal; In the 1900-1924 birth cohort, the number of females ($n = 87$) is almost double the number of males ($n = 47$). Simply increasing the proportion of females to males over time produces a negative secular change. Although in this case the change over time in the combined-sex sample is not significant ($F = 1.334$, $p = 0.265$), the results are nevertheless misleading, particularly when depicted graphically. In this context, mis-classification of small males and large females is particularly problematic when investigating secular change using data where

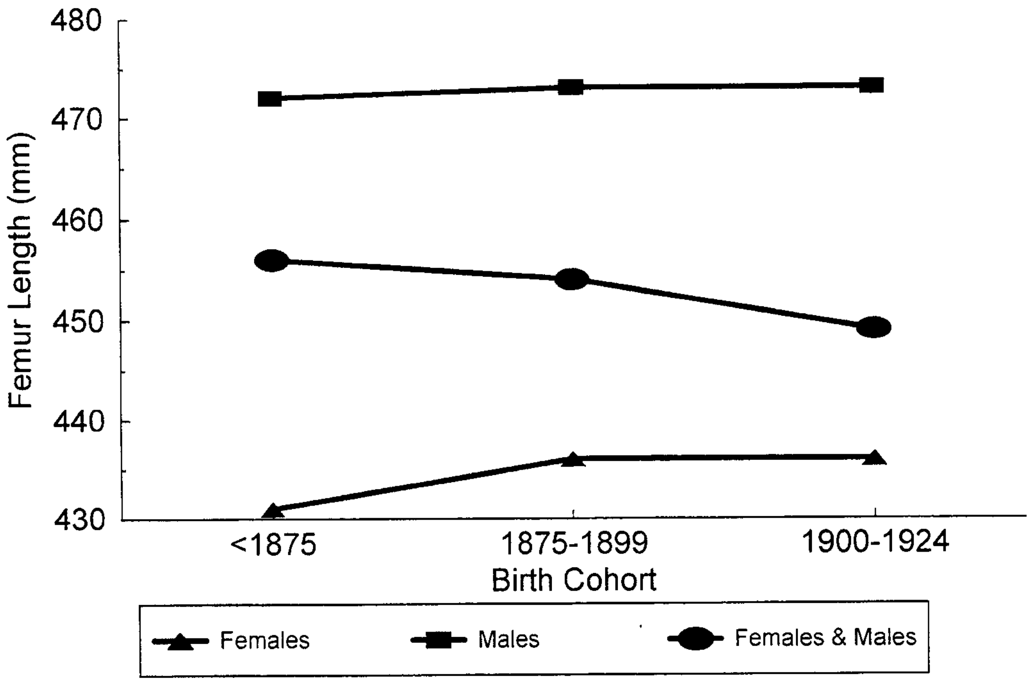


Figure 5.3: Data for all individuals of both sexes from the Terry Collection sample and Terry Collection males and females separately. Increasing the proportion of females to males over time produces a negative secular change in the combined sample even though neither females nor males have a negative trend when considered separately. Sample sizes vary by sex and cohort: In the earliest birth cohort, males (n = 34) outnumber females (n = 20); in the 1875-1899 birth cohort, the number of males (n = 61) and females (n = 64) are almost equal; in the 1900-1924 birth cohort, the number of females (n = 87) is almost double the number of males (n = 47).

individual skeletons are not positively identified. When using undocumented skeletal samples or other sources of data, a combination of several sex determination methods with minimal differences between the allocation accuracy for males and females are essential.

CONCLUSION

There are several benefits to using skeletal data for the study of secular change in stature and limb proportions. With skeletal data, different types of measurement error can be controlled or minimized and it is possible to investigate changes in absolute size, and body and limb proportions over time for both sexes. In this study, femur length data have been used for illustrative purposes; however, the issues related to how samples are constructed for the study of secular change are relevant regardless of which variables are used. This study has shown that some caution must be exercised. Combining data from different sources without confirming that the samples represent the same population can produce very misleading results. The focus in this study has been on skeletal data to highlight the differences between sources of data; however, other criteria must be considered before samples from different sources are combined. Skeletal differences between sources of data when birth cohort is held constant, such as the differences in femur length between the Coimbra and Terry Collections, do strongly suggest that samples from these sources should not be combined for the study of secular change. On the other hand, lack of skeletal differences when birth cohort is held constant, such as the similarities in femur length between the FDB and the Terry Collection, do not necessarily suggest that samples should be combined. In both cases, socio-economic and geopolitical criteria, and biases in the collection process when the data sources were amassed (which can amplify or minimize the effects of these criteria), must be considered before samples are combined for the study of secular change. Combining data from different sources without considering these factors will result in other sources of variation being attributed to secular change.

In addition, dividing samples into racial groups does not control for differences between the sample sources. There is some evidence (documented nationality, place of birth, bequeathed versus not bequeathed, etc.) that the Terry Collection may not represent a single population; however, there is no evidence that dividing samples from the same data source by “race” in any way controls for this possible lack of homogeneity. There are methodological problems with the use of the race concept to define samples for the study of human variation. Race is neither a biological concept that applies to humans nor a fixed social category. Variation through time in the social criteria used for racial designation raises problems even when investigating the effects of racism on secular change using reference skeletal collections. Finally, it is essential that males and females be considered separately when investigating secular change for biological reasons and for reasons related to possible variation over time in the sex proportions in the study sample.

REFERENCES CITED

Albanese J. 2001a. Measurement error and metric methods for sex determination. Paper presented at the 29th Annual Meeting of the Canadian Association of Physical Anthropology, Oct 25-27, Winnipeg, Manitoba.

Albanese J. 2001b. A new metric method for sex determination using the pelvis and femur. Am J Phys Anthropol 2001 Suppl 32:31.

Albanese J. 2003. A new metric method for sex determination using the hipbone and the femur. J Forensic Sci 48:263-273.

Angel L. 1976. Colonial to modern skeletal change in the U.S.A. Am J Phys Anthropol 45:723-736.

Armelagos GJ, Goodman AH. 1998. Race, racism, and anthropology. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press. p 359-378.

Brauer GW. 1982. Size sexual dimorphism and secular trend: Indicators of subclinical malnutrition? In: Hall RL, editor. Sexual Dimorphism in Homo Sapiens: A Question of Size. New York: Praeger. p 245-259.

Brown RA, Armelagos GJ. 2001. Apportionment of racial diversity: A review. Evolutionary Anthropology 10:34-40.

Cartmill M. 1998. The status of the race concept in physical anthropology. Amer Anthrop 100: 651-660.

Cunha E. 1995. Testing Identification Records: Evidence from the Coimbra Identified Skeletal Collections (Nineteenth and Twentieth Centuries). In: Saunders SR, Herring A, editors. Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press. p 179-198.

Ericksen MF. 1982. How “representative” is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 59:345-350.

Fish, JM. 1999. Why psychologists should learn some anthropology. In: Sussman, RW, editor. The Biological Basis of Human Behavior: A Critical review. New Jersey: Prentice Hall. p 197-198.

Floud R. 1994. The heights of Europeans since 1975: A new source for European economic history. In Komlos J, editor. Stature, Living Standards and Economic Development. Chicago: University of Chicago Press. p 25-38.

Greiner TM, Gordon CC. 1992. Secular trends in 22 body dimensions in four racial/cultural groups of American males. Hum Biol 4:235-246.

Greulich WW. 1951. The growth and developmental status of Guamanian school children in 1947. Am J Phys Anthropol 9:55-70.

Greulich WW, Crimson CS and Turner ML. 1953. The physical growth and development of children who survived the atomic bombing of Hiroshima and Nagasaki. Nagasaki Journal of Pediatrics 43:121-145.

Hartmann J, Keister R, Houlihan B, Thompson L, Baldwin R. 1994. Diversity of ethnic and racial VNTR RFLP fixed-bin frequency distributions. Am J Hum Genet 55:1268-1278.

Hrdlička A. 1934. Contributions to the study of the femur: The crista aspera and the pilaster. Am J Phys Anthropol 19:17-37.

Keita SOY, Kittles RA. 1997. The persistence of racial thinking and the myth of racial divergence. Amer Anthrop 99:534-544.

Lewontin R. 1972. The apportionment of human diversity. Evol Biol 6:381-398.

Lieberman L. 2001. How "Caucasoids" got such big crania and why they shrank. Curr Anthrop 42:69-95.

Livingstone F. 1962. On the non-existence of human races. Curr Anthrop 3:279-181.

Meadows Jantz L, Jantz RL. 1999. Secular change in long bone length and proportion in the United States, 1800-1970. Am J Phys Anthropol 110:57-67.

Mokyr J, Gráda CO. 1994. The heights of the British and the Irish c. 1800-1815: Evidence from recruits to the East India Company's Army. In Komlos J, editor. Stature, Living Standards and Economic Development. Chicago: University of Chicago Press. p 39-59.

Molnar S. 2002. Human Variation: Races, Types, and Ethnic Groups. Fifth Edition. Upper Saddle River, N.J.: Prentice Hall.

Nei M, Roychoudhury Ak. 1982. Evolutionary relationships and evolution of human races. Evol Biol 14:1-59.

Ousley SD, Jantz, RL. 1998. The forensic data bank: documenting skeletal trends in the United States. In: Reichs KJ, editor. Forensic Anthropology. Second Edition. Springfield: CC Thomas. p 441-458.

Relethford JH. 1994. Craniometric variation among modern human populations. Am J Phys Anthropol 95:53-62.

Relethford JH. 2001. Global analysis of regional differences in craniometric diversity and population substructure. Hum Biol 73:629-636.

Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393-398.

Ryman N, Chakraborty R, Nei M. 1983. Differences in the relative distribution of human genetic diversity between electrophoretic and red and white cell antigens. Hum Heredity 33:93-102.

St. Hoyme LE, İşcan MY. 1989. Determination of sex and race: accuracy and assumptions. In İşcan MY, Kennedy KA, editors, Reconstruction of Life from the Skeleton, p. 53-93.

Santos AL. 2000. A Skeletal Picture of Tuberculosis. Doctoral thesis. University of Coimbra.

Sokal RR, Rohlf FJ. 1973. Introduction to Biostatistics. San Francisco: WH Freeman and Company.

SPSS. 1998. SPSS 9.0 for Windows. Chicago: SPSS Inc.

Steckel RH 1994. Heights and health in the United States, 1710-1950. In Komlos J, editor. Stature, Living Standards and Economic Development. Chicago: University of Chicago Press. p 153-172.

Stini WA. 1975. Adaptive strategies of human populations under nutritional stress. In: Watts ES, Johnston FE, Lasker GW, editors. Biosocial Interrelations in Population Adaptation.The Hague: Mouton, p 19-41.

Stini WA. 1982. Sexual dimorphism and nutrient reserves. In: Hall RL, editor. Sexual Dimorphism in Homo Sapiens: A Question of Size,. New York: Praeger. p 391-419.

Stinson S. 1985. Sex differences in environmental sensitivity during growth and development. Yearb Phys Anthropol 28:123-147.

Templeton AR. 1998. Human races: A genetic and evolutionary perspective. Amer Anthrop 100:632-650.

Terry RJ. 1932. The clavicle of the American Negro. Am J Phys Anthropol 3:351-379.

Tobias PV. 1972. Growth and stature in Southern African populations. In: Vorster DJM, editor. Human Biology of Environmental Change. London: International Biological Program, p. 96-104.

Tobias PV. 1975. Anthropometry among disadvantaged peoples: Studies in Southern Africa. In Biosocial Interrelations in Population Adaptation, E. S. Watts, F. E. Johnston, G. W. Lasker (eds). The Hague: Mouton, p. 287-305.

Tobias PV. 1986. War and negative secular trend of South African Blacks with observations on the relative sensitivity of cadaveric and non-cadaveric populations to secular effects. Proceedings of the 5th Congress of the European Anthropological Association 1:451-461.

Tobias PV, Netscher D. 1977. Reversal of the usual secular trend, as evidenced by South African Negro crania and femora. Sonderabduck aus Mitteilungen der Anthropologischen Gessellschaft in Wien 103:233-239

Trotter M. 1981. Robert James Terry, 1871-1966. Am J Phys Anthropol 56:503-508.

Trotter M, Gleser GC. 1951. Trends in stature of American Whites and Negroes born between 1840 and 1924. Am J Phys Anthropol 9:427-440.

CHAPTER 6

A Metric Method for Sex Determination Using the Hipbone and the Femur

By
John Albanese

Published in *The Journal of Forensic Sciences* Mar. 2003 Vol 48(2):263-273 (Paper ID JFS 2001378_482),

ABSTRACT

Since the earliest descriptions of the pubis length measurement, it has been recognized that the location of the key landmark in the acetabulum has to be estimated. Using samples from the Terry Collection (n = 324) and the Coimbra Collection (n = 232), the purpose of this research is to, first, test the reproducibility of a new alternative to the traditional measurement of the pubis, and second, to use the best measurement of the pubis along with other measurements of the hipbone and femur to develop a metric method that can be used with confidence to determine the sex of individuals of various geographic origins and time periods.

In this study, it was found that, first, the alternative pubis measurement, known as the superior pubis ramus length (SPRL), can be collected more reliably with less mean

intra-observer error (0.57%) than the more commonly used manner of measuring the pubis (2.7%). Second, a logistic regression sex determination method using the SPRL, along with other measurements of the hipbone and femur, has an allocation accuracy of 90% to 98.5% (depending on the model used and the manner of testing) across independent samples. Third, traditional racial categorization was irrelevant to the accuracy of the method. Fourth, measurement error greater than 2% in the measurement of the pubis can mean the difference between a correct and an incorrect allocation of sex, particularly in borderline cases.

INTRODUCTION

Determination of sex of an unknown individual is one of the critical questions addressed when human skeletal remains are found in both forensic investigations and in studies of past populations. Beginning with the earliest investigations into the development of sex determination methods, the pelvis in general and the pubis in particular have been recognized as the best sources of information for determining the sex of an unknown individual (Stewart 1979). However, since the earliest descriptions of the pubis length measurement, it has been recognized that the location of the key landmark in the acetabulum has to be estimated. While the description of the pubis length measurement is well defined, relying on an estimated landmark position will result in relatively high measurement error. Using samples from the Terry Collection and the Coimbra Collection, the purpose of this research is to, first, test the reproducibility of an alternative to the traditional measurement of the pubis; second, to present an alternative approach to constructing a reference sample for developing sex determination methods; and third, to use the best measurement of the pubis along with other measurements of the hipbone and femur to develop a metric method that can be used to determine the sex of individuals of various geographic origins and time periods.

MEASUREMENT OF THE PUBIS AND METRIC APPROACHES FOR DETERMINING SEX

The length of the pubis is usually defined as the distance between the superior margin of the pubic symphysis and the acetabular margin of the pubis. Although the acetabular margin of the pubis is clearly visible in subadults, in most individuals, the pubis, ischium and ilium are completely fused by the age when the pubis is actually useful for sex determination (Olivier 1969, Stewart 1979). Fusion begins around the time of puberty and the three elements are completely fused by the mid to late teens (Steele and Bramblett 1988). Thus, while the description of the pubis length measurement may be clear, difficulty in locating the acetabular margin of the pubis in adults often results in problems with the reproducibility of the measurement.

The difficulty in locating the acetabular landmark has been recognized for as long as the pubis has been measured. In one of the earliest descriptions of the pubis length measurement, Schultz (1930) used a diagram of what he referred to as an "infantile" hipbone to highlight the landmark that is difficult or impossible to locate in adults. Many descriptions of the pubis length that have been published in North America in the last 50 years quote or paraphrase Washburn's (1948) description and suggestions for estimating the point where the pubis, ischium and ilium meet in adults (for example, see Bass 1987, Moore-Jansen et al. 1994). These "hints" include looking for irregularities and notches in the acetabulum and on the inside surface of the pubis, and holding the acetabulum up to the light to look for differences in bone thickness. In Europe, the same difficulties in locating the point are described by Olivier (1969). The confusion, which still persists in North America and Europe, surrounding the position of the acetabular landmark was summarized by Olivier over 35 years ago,

> It would appear that the [point where the ischium, ilium and pubis meet] in fully ossified bone is variously located by different authors. For Schultz it is the point A in the figure (cotyloid point) but this is a variable location, situated where the inner border of the hip bone[s] meet each other. In monkeys it is easy to find. In [humans] it is often a notch of the articular margin at this level as well as a more internal roughness of the bone. Several authors have defined this point more exactly: for Genovés it is that point of the inner articular margin which is nearest to the anterosuperior iliac spine. For Gaillard it is the intersection of the long axis of the pubis and the ischium (Olivier 1969:248-9).

Despite the problems with locating the acetabular landmark, this measurement is the only pubis measurement described in two recent measurement reference volumes (see Buikstra and Ubelaker 1994, Moore-Jansen et al. 1994).

Building on the work of Thieme (1957) and Washurn (1948), an alternative to the pubis length was described by Schulter-Ellis and colleagues (Schulter-Ellis et al. 1985, Schulter-Ellis et al. 1983) and used to develop a sex determination method that had great promise, but does not seem to be used widely. Their measurement of the pubis (PS-A) is described as the length between the superior margin of the pubic symphysis to the nearest rim of the acetabulum. They note that the main advantage of the PS-A is that the problems associated with estimating the position of the landmark where the pubis, ischium and ilium meet are avoided, but they do not provide any evidence to support this statement. They do not show that their PS-A, with a floating landmark on the acetabulum rim, is reproducible with a reasonably low margin of error nor do they demonstrate that it is more reproducible than the traditional pubis length.

The sex determination method developed by Schulter-Ellis and colleagues may not be used widely because it is needlessly complex and was not adequately tested. First, they use ratios as variables in discriminant function equations. Second, they use their *a priori* knowledge of documented sex in order to decide when to include femur measurements in cases where the pelvic discriminant functions were inconclusive or incorrect. Third, the method was tested on the same sample that was used to develop the method. Lastly, they used 100 "Whites" and 100 "Blacks" to develop two different methods for sex determination without presenting any evidence that suggests that race-specific methods are necessary. Although the allocation accuracy of their method was as high as 96%, the method of testing and other complexities suggest that their method may not perform well in actual cases. These problems may contribute to the lack of widespread use of this method, particularly when there are other good options for sex determination when the pelvis is complete.

Bruzek (1992) independently tested the method developed by Schulter-Ellis and colleagues using identified skeletons from the Paris and Coimbra Collections and found that the Schulter-Ellis method had an allocation accuracy of over 90%: Paris males, 91.5%; Paris females, 100%; Coimbra males, 93.5%; and Coimbra females 95.1%. The major shortcoming was that there was an 8.5% difference in allocation accuracy between Paris Collection males and females, a problem that is common with other metric sex determination methods (Saunders and Hoppa 1997).

MATERIALS AND METHODS

Representativeness, Sources of Variation and Sample Selection

The composition of the reference sample used to develop sex determination methods can have an impact on the applicability of the methods that are developed (Albanese 1997a, Albanese 1997b, MacLaughlin and Bruce 1990, Meindl et al. 1985, Saunders and Hoppa 1997). It has been assumed for decades that the Terry Collection and other similar collections are not representative of populations in the U.S.A. (Giles 1964) and may no longer be useful for the development of forensic identification methods (Ousley and Jantz 1998). However, with only a few exceptions (for example, Ericksen 1982, Ousley and Jantz 1998), this lack of representativeness has never been investigated in detail and the effects of the biases of the collection on skeletal variation -rather than the applicability of methods- has never been assessed. However, if representativeness is considered in shades of grey rather than as black and white, then the level of representativeness of any collection will vary depending on the research question. Furthermore, representativeness can be maximized and biases can be minimized or even exploited with alternative approaches to sampling in order to address specific research questions. With careful sampling, it is possible to construct a large reference sample from the Terry Collection or a combination of identified collections that captures a wide range of modern human variation and that can be used to develop highly accurate sex determination methods that can be applied to 21st century populations. The key is considering the demographic data (age, year of birth, etc.) and the historical details such as socio-economic, political and legal issues associated with the construction of the collections. In the past, this information has been used to dismiss the collection as antiquated and not representative. In this study the same information is used to minimize bias and maximize representativeness of human variation when constructing a reference sample. The underlying assumption in sample selection for this study is that if a greater amount of human variation is included in the sample used to develop a sex determination method, then that method can be applied with confidence in a wide range of cases. The samples used in this study are from two very different skeletal collections: the Robert J. Terry Anatomical Collection (Smithsonian Institution, Washington, D.C.) and the Coimbra Identified Skeletal Collection (Museum of Anthropology, University of Coimbra, Coimbra, Portugal). Details regarding the Coimbra Collection[1] can be found in publications by Rocha (1995) and Cunha (1995). Careful planning went into the assembly of the Coimbra Collection and supporting documentation; however, the collection is not a random sampling of either the cemetery from which the skeletons were derived nor the greater population of the District of Coimbra (Santos 2000). Despite these biases, the

---

1 There are two series of identified skulls and one series of identified complete skeletons at the University of Coimbra (Cunha 1995, Rocha 1995). In this paper, Coimbra Collection refers to the 505 identified skeletons (Colecção de Esqueletos Identificados).

collection has been used to develop and test forensic identification methods and in palaeopathological investigations (Bruzek 2002, Cunha 1995, see also Bruzek, 1992, Santos and Roberts 2001)

Information regarding the size and basic demographic profile of the Terry Collection have been described in numerous sources (for example, İşcan 1990, Ubelaker 2000). However, both the demographic profile and the historical context of the collection have had an impact on the variation that is present in the collection and must be considered in some detail when constructing a reference sample (Albanese 2002). The Terry Collection is the result of the joint efforts of Drs. Robert J. Terry and Mildred Trotter over six decades at the medical school at Washington University in St. Louis, Missouri. The collection was derived from anatomy school cadavers which were mostly unclaimed bodies from various hospitals and institutions in Missouri. Only a relatively small number who died after 1955, about 10% of the entire collection, were people who bequeathed their bodies for medical use. This 10% of the collection has been described as middle class American (Angel 1976, Ericksen 1982). However, using data from the proximal femur, Ericksen (1982) found that the differences between the bequeathed and non-bequeathed individuals are significant in some cases but do not conclusively separate these sub-samples. She hypothesizes that the practice of bequeathment can be associated with higher socio-economic status at the time of death even though the individual may have lived under different conditions during their growth and development period. Conversely, the other 90% of the collection consists of individuals who were possibly of very low socio-economic status *only at the time of death.* About 55% of the 1618 individuals who died before 1955, died during the Great Depression (1929-1939) and may not necessarily have lived in poverty during their growth period. As Terry noted several decades ago when describing the cadavers before dissection and maceration,

> ...these bodies commonly bear the marks of undernourishment and in many cases of the wasting effects of chronic ailment that brought death. Whereas *these conditions scarcely affect at all the longitudinal measurements* they render some of the transverse and circumferential measurements of questionable value (Terry 1940:435; emphasis added).

A similar impact on the skeleton should be expected. For a number of social and historical reasons, the Terry Collection, like other skeletal collections in North America, *had* a very unbalanced sex ratio. Unlike other skeletal collections and largely through Trotter's efforts, this imbalance was, in part, corrected (Trotter 1981). Trotter was instrumental in drafting a major change in the Missouri laws on bequeathment of human remains. After this change in the mid-1950s, Trotter focused on adding the skeletons of younger "White" females to the collection (Trotter 1981).

In the U.S.A., the trend in physical and forensic anthropological research in the last 60 years in the U.S.A. has been to randomly sample major collections and to develop race-specific sex determination methods in order to control for differences in sexual dimorphism between populations (for example, Giles 1964, Holman and Bennett 1991, İşcan and Miller-Shavitz 1984, Marino 1995, Ousley and Jantz 1996, Schulter-Ellis et al. 1985, Schulter-Ellis et al. 1983, Smith 1997). In these examples a major source or the exclusive source of data is the Terry Collection. Much of the variation attributed to racial differences in sexual dimorphism in various studies that have sampled the Terry

Collection may be attributable to Trotter's efforts to correct for the lack of young "White" females. In any random sample of the Terry Collection there are different proportions of males and females of different "races" from various year of birth and age cohorts. A random sampling of the collection for the development of race-specific sex determination methods will result in a poor sampling of sexual dimorphism and the erroneously elevate the apparent significance of "race" because the latter is so closely correlated to age at death, year of birth and other variables associated with how the collection was constructed. In a random sample of "Whites" there are a disproportionately high number of younger adult females born early in the 20th century compared to a male sample composed of older individuals born in the middle decades of the 19th century. In a randomly selected sample of "Blacks," the impact of Totter's approach to adjusting the collection has been less because there are many more "Black" females than "White" females that were born in the middle to the end of the 19th century. Because of Totter's approach to collecting, a comparison of "Black" and "White" females is actually a comparison of age and year of birth differences between younger females born in the 20th century who were described as "White" to older females born late in the 19th century who were described as "Negro" when they were included in the collection. Consequently, using logistic regression (see below for details), it is possible to assess "race" with an allocation accuracy of 69.5% using only age at death and year of birth (no skeletal data) in a sample that includes all the females in the Terry Collection over 18 years of age for whom age and year of birth data are available (n = 671).

Rather than control for "race" in a sample from the Terry Collection, it is argued here that the underlying sources of variation (age at death and year of birth) that are highly correlated with "race" due to the collection process need to be considered when selecting samples. Giles (1970:102) minimized the importance of the effects of age on sex determination but he found that cranial "discriminant functions tend to misclassify younger males and older females." He went on to state that this pattern of misclassification may be "present but undetected" in morphological sex determination methods. Walker (1995) and Meindl and colleagues (1985) confirmed Gile's hypothesis and have shown that age at death can be a critical factor in the level of sexual dimorphism in the cranium. Walker found that younger males have a morphology that is referred to as a typically female pattern and older females have a morphology that is referred to as a typically male pattern. After reviewing several different cranial sex assessment methods, Meindl and colleagues found that "greater age produces an increasingly male morphology" (1985: 81). Anderson (1990) has suggested that a similar but inverted pattern may occur with the ventral arc. Although his sample size is small and not distributed across a wide range of ages, Anderson reports that the ventral arc is less visible in females under 20 years of age and becomes more visible in males older than 70 years of age. When testing the Phenice method (Phenice 1969), which includes the ventral arc, Lovell (1989) found that accuracy decreased with age at death because of age-related irregularities in the pubic bone. Sutherland and Suchey (1991) reproduce Anderson's results but not Lovell's conclusions regarding age-related changes in the ventral arc. Still other research suggests that age may also be a factor in the absolute length of the pubis, where growth of the pubis continues into the third decade of life in females (Tague 1994). However, others have attributed this age related difference in pubis length as a misinterpretation of differential mortality (Fuller 1998). Lastly, the effects on sexual dimorphism of well documented secular change in the late 19th and throughout the 20th centuries are not well defined or fully understood. Because of the manner in which both the Terry and Coimbra Collections were collected, a random approach to sample selection will result in a poor and uneven sampling of variation associated with age at death and year of birth. A sex determination method developed from such a sample would be of limited use.

In this study an effort was made to include the full range of adult ages and a wide range of years of birth represented in each collection. Any variation that may possibly be related to age at death or secular change is sampled and included in the analysis even if year of birth and age at death are not included as predictor variables. Figure 6.1 is a scatter plot of age at death by year of birth of the samples from both collections. There are adults that are older and younger than 40 years of age born before and after 1900. A portion of the sample falls into each of the four quadrants of the scatter plot in Figure 6.1. All years of birth and ages are well represented. The diagonal linear appearance of the plot is the result of the limits of the Terry and Coimbra Collections in years of birth. The sample from the Terry collection consist of 324 individuals and 232 individuals are drawn from the Coimbra Collection. Years of birth range from approximately 1832 to 1913 for the Coimbra Collection sample and 1850 to 1930 in the Terry Collection sample. Ages range from 19 to 79 in the samples from both collections. The upper age limit was arbitrarily set at 79 years to avoid missing data related to extreme joint

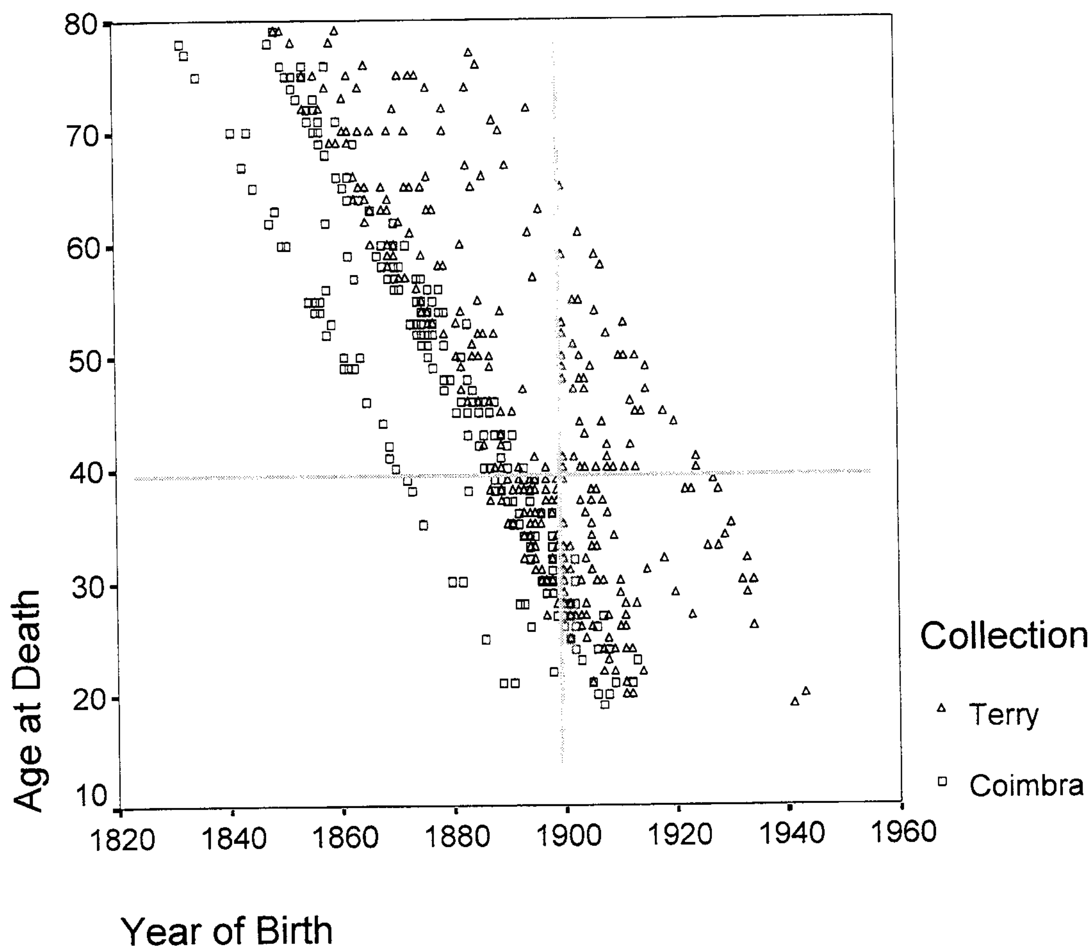


Figure 6.1: Age at death by year of birth for the entire sample from the Terry Collection (triangles) and the Coimbra Collections (squares), n = 556.

problems and misleading data that resulted from age-related loss of robusticity. Many of the individuals in the Terry Collection died before the 1935 Social Security Act and would have continued to work in manual labour employment well beyond 65 years of age. The skeletal impact of this behavior adds to the complexity of interpreting skeletal variation in the Terry Collection since it is in contrast to the wasting from malnutrition and diseases described by Dr. Terry. The lower age limit is dependent on chronological and biological criteria. First, documented age at death is 18 years or greater. Second, all epiphyses, with the exception of the sternal end of the clavicle had to be at least partially fused. In some cases, epiphyseal lines were visible but epiphyses were never separate from diaphyses on any long bones.

Measurements Collected

Several standard femur and hipbone measurements were collected for the entire sample including hipbone height, iliac breadth, pubis length, ischium length, maximum femur length, maximum femur head diameter, anterior-posterior diameter of the femur at mid-shaft, transverse diameter of the femur at mid-shaft, and epicondylar breadth of the femur (Bass 1987, Moore-Jansen et al. 1994, Olivier 1969). These measurements are defined in Table 6.1. Two new measurements were also collected. The new measurement of the pubis is referred to as the superior pubis ramus length (SPRL). The second measurement is an alternative for measuring the ischium and is referred to as the Acetabular-Ischium Length (AIL). All measurements were collected by the author to the nearest millimeter.

The SPRL is measured using sliding calipers from the superior margin of the pubic symphysis to the superior-anterior apex of the lunate surface in the acetabulum. Unlike the measurement of the pubis described by Schulter-Ellis and colleagues, the SPRL described here has a fixed, easily recognized landmark in the acetabulum. Occasionally, this landmark may be affected when there are extreme arthritic changes to the acetabulum. In most cases it is easy to measure around any arthritic lipping on the rim of the acetabulum. In the few extreme cases, using spreading calipers may be

Table 6.1: Definitions of traditional measurements*.

Measurement	Definition
hipbone height	Distance from the most inferior point on the ischial tuberosity to the most superior point on the iliac crest (1,2). Measured using an osteometric board. Also know as the innominate height.
iliac breadth	Distance from the anterior superior iliac spine to the posterior superior iliac spine (1, 2). Measured using an osteometric board†.
pubis length‡§	Distance from the superior margin of the pubic symphysis to the point in the acetabulum where the ischium, ilium, and pubis meet (1, 2, 3). Measured using sliding calipers.
ischium length§	Distance from the point in the acetabulum where the ischium, ilium, and pubis meet to the most inferior point on the ischial tuberosity perpendicular to the pubis length (2, 3). Measured using sliding calipers.
maximum femur length	Distance from the femur head to the most inferior point on the medial condyle (1, 2, 3). Measured using an osteometric board.
maximum diameter of femur head	Maximum diameter of the head of the femur at the border of the articular surface (2). Measured using sliding calipers.
epicondylar breadth of femur	Distance between the most projecting points on the medial and lateral condyles (1, 2). Measured using sliding calipers.

*Anterior-posterior diameter of the femur and transverse diameter of the femur at mid-shaft are not statistically significant predictor variables and were not used for other illustrative purposes, and have been left out.
†Some sources recommend that spreading calipers be used; however, using an osteometric board was found to be much quicker and easier while providing identical results.
‡Various definitions of the acetabular landmark for the pubis length are discussed in the text.
§Definitions of the SRPL and AIL are in the text and illustrated in Figure 6.2.
1. Olivier 1969
2. Moore-Jansen et al 1994
3. Bass 1987

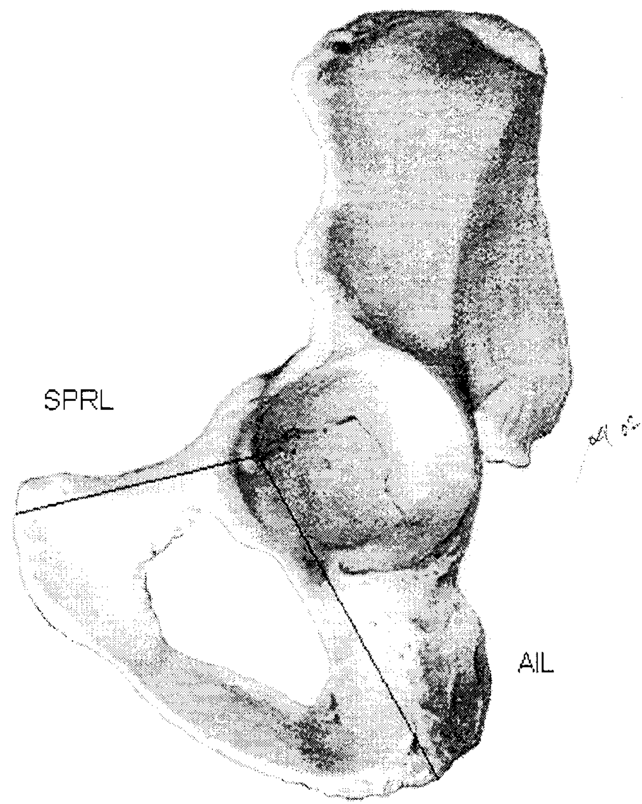


Figure 6.2: The superior pubis ramus length (SPRL) and the acetabular ischium length (AIL).

advantageous. The AIL is measured from the same landmark on the acetabulum to the most inferior point on the ischium and not perpendicular to the SPRL. The maximum length should be measured. Both measurements are illustrated in Figure 6.2.

Intra-observer Error And Measurement of The Pubis

In a randomly selected sub-sample of just over 10% of the entire sample (n = 65), all measurements were re-collected to test their reproducibility and the level of intra-observer error. The sample was divided about equally between males and females and approximately proportionally by collection: 13 males and 13 females from the Coimbra Collection, and 19 females and 20 males from the Terry Collection. To calculate the percent intra-observer error, the absolute difference between the two measurements was divided by the first measurement and then multiplied by 100 for each individual in the intra-observer error sample. The mean error was then calculated. Only the pubis and ischium data are presented here. For the traditional pubis length measurement, the mean intra-observer error is 2.7%. For the new SPRL measurement, the intra-observer error is 0.57%. Figure 6.3 is a scatter plot of the measurement error on a case by case basis for each of the 65 individuals. The error for the traditional pubis measurement by the SPRL error are graphed. For the SPRL, more than half of the sample has an error of zero, and in almost 95% of the cases, the error is less than 2% or 1 mm. For the traditional pubis length measurement, less than 30% of the cases are below 2%, while the majority have an error greater than 2%.

A review of all the cases revealed that there is no pattern to the errors for either measurement. Using the Spearman correlation statistic (SPSS 1998), correlations were weak and not significant between the measurement errors and the collection (Terry or Coimbra), the date of data collection, or sex of the individual. Data collection was consistent throughout the data collection period and there is no relationship between

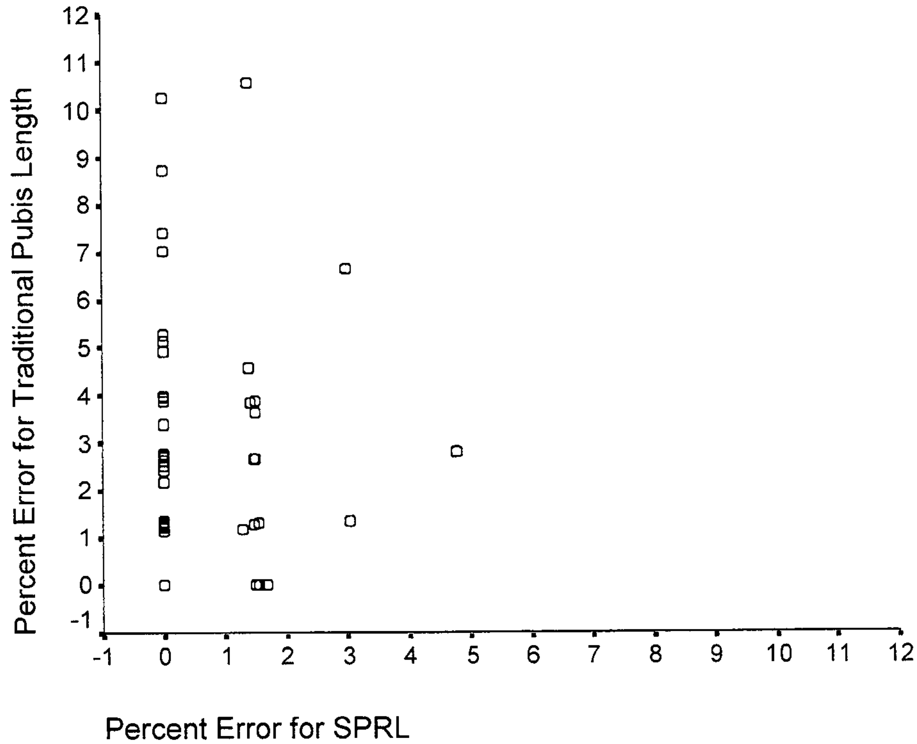


Figure 6.3: Plot of percent measurement error for the traditional pubis length with percent measurement error for the superior pubis ramus length (SPRL) for each case (n = 65). Mean intra-observer measurement error for the traditional pubis length is 2.7%. Mean intra-observer measurement error for the SPRL is 0.57%.

errors in the traditional pubis length and the SPRL measurements. The larger intra-observer error in the traditional pubis length measurement can be attributed to the difficulty in locating the acetabular landmark.

The intra-observer error of the AIL and the traditional ischium length (not shown graphically) follows a different pattern than the comparison of the SPRL and the traditional pubis length measurement. Both the AIL and the traditional ischium length measurements have a low, virtually identical intra-observer error at 0.98% and 0.94%, respectively[2].

Statistical Approach

Several multivariate statistical approaches can be used to predict a binary dependent variable, such as sex, from a group of independent variables. Discriminant function analysis is the most commonly used approach in skeletal sex determination methods; however, other more statistically robust methods are available. One underused yet very powerful approach is logistic regression.

Norušis (1990) suggests several reasons why logistic regression is a better choice than discriminant function analysis when predicting a binary dependent variable (see also Saunders and Hoppa, 1997, for a discussion of logistic regression and sex determination). First, with discriminant function analysis, there is the assumption of a normal distribution of the independent variables. Although skeletal metric data are usually normally distributed when samples are large, logistic regression does not require a normal distribution to optimize prediction accuracy and categorical variables can be used as independent variables along with metric data. Second, logistic regression analysis does not require equal variance-covariance matrices in the two groups (female and male), a condition that is necessary for discriminant function analysis. Norušis adds that "even when the assumptions required for discriminant analysis are satisfied, logistic regression still performs well" (Norušis 1990:119). Aside from the underlying statistical

---

2 Extensive tests of inter-observer measurement error of all of the pubis and ischium measurements will be presented in a future publication.

assumptions, a major benefit of a logistic regression model over a discriminant function model when allocating individuals -as opposed to discriminating between groups- is that the probability of the event is calculated. Separate posterior probability and/or typicality probability statistics must be considered for an analogous approach when using discriminant functions. With a few exceptions (İşcan et al. 1994, Kieser et al. 1992, Ousley and Jantz 1996), this approach is missing from most discriminant function sex determination methods.

The logistic model is an S-shaped function of the form

$$P = \frac{1}{(1 + e^{-Z})}$$

where P is the probability of the event (male or not male in this case) and Z is a linear combination of the independent variables such as,

$$Z = \beta_o + \beta_1 X_1 + \beta_2 X_2$$

Calculated probabilities are always between 0 and 1. If P is greater than 0.5, then the individual is classified male. If P is less than 0.5, the individual is classified female. For example, if P = 0.87, the individual is classified as male since P is greater than 0.5 and the estimated conditional probability is 87% that this combination of measurements would be found in a male. Alternatively, there is only a 13% probability (1-P) that this combination of measurements would be found in a female. Despite the clear benefits of using the logistic regression model, with only a few exceptions (Albanese 2001a, Albanese 2001b, Dudar and Dupras 1996, Saunders and Hoppa 1997) it has been underutilized in forensic anthropological research (Saunders and Hoppa 1997).

There are four main methods used to assess the fit of a logistic model to the data. First, any later version of SPSS (version 9.0 for Windows was used for this study; SPSS 1998) will automatically calculate the allocation accuracy of the model applied to the sample used to develop the model. Second, a histogram of probabilities of the sample used to develop the method can be generated to assess the range of probability scores. Ideally, the histogram should have few probabilities in the mid range and two large spikes near 0 and 1 for females and males, respectively. Third, a goodness of fit statistic known as the -2 log likelihood (-2LL) is calculated. The lower the -2LL statistic, the better the fit of the model to the data. Fourth, the allocation accuracy can be calculated for a hold-out sample not used to develop the method.

Norušis (1990) notes that all the same problems with variable selection algorithms found in regression and discriminant functions can also be found in logistic regression. The two major approaches to automated variable selection involve the Wald statistic and the Likelihood Ratio (LR) test. Either can be used in a forward or backward stepwise procedure. Both work equally well provided that sample sizes are large (Norušis 1990). In this study, experiments with both statistics in the forward and backward stepwise procedures resulted in exactly the same model. For consistency, the Forward LR was selected whenever automated variable selection was used.

In this study, a two step approach was used for variable selection. First, various possible scenarios were considered such as complete recovery of skeletal remains (hipbone and femur), dismemberment, and postmortem damage to the pubis, ischium, ilium and/or femur. Second, within each of these scenarios, the Forward LR option was used to select the best predictors from the available variables. Because of missing data for some individuals, sample size varies for each model.

The sample of over 550 individuals from both the Terry and Coimbra Collections was divided into two sub-samples. Models were developed using a sub-sample of 422 individuals (75%) from both collections which is referred to as the ***model sample***. The model was tested on a hold-out sample of 134 individuals (25%) referred to as the ***test sample*** which was not used to develop the models. The test sample was selected from the overall sample using the random sample selection feature in SPSS. The composition by sex and collection of the model and test samples are presented in Table 6.2.

Table 6.2: Model sample* and test sample by collection and sex.

	Model Sample			Test Sample			
	Te	Co	T	Te	Co	T	TOTAL
F	129	92	221	41	26	67	288
M	109	92	201	45	22	67	268
T	238	184	422	86	48	134	556

*The sample of 422 is the pool from which models were developed. Sample size varies from 401-418 for various models because of missing data. See Table 6.3 for specific sample sizes.
Abbreviations: F = female, M = male, T = total, Te = Terry Collection, Co = Coimbra Collection

RESULTS

The best-fit model which included hipbone height, iliac breadth, SPRL, maximum femur head diameter, and epicondylar breadth of the femur, correctly allocated 98% of the model sample. Allocation accuracy for males and females was identical at 98%. When the method was tested on the hold-out test sample, the results were slightly better: 98.5%

of both the males and the females were allocated correctly. Stated another way, out of 134 test cases, only one male and one female were allocated incorrectly. Rather than graph the probabilities of the model sample to test the fit of the model, probabilities for each individual in the test sample are presented in Figure 6.4. Note the near-perfect fit of the model. With the exception of the one male and one female who were allocated incorrectly, all the males have high probabilities (≥0.80) and all the females have low probabilities (≤0.2). In 93% of the test cases, the scores indicated that there is 90% or greater probability of a correct allocation. Thus, there are both consistently high probability scores for each individual allocation and a high overall allocation accuracy. A summary of the various models and allocation accuracies are presented in Table 6.3. Coefficients for each model are presented in Table 6.4.

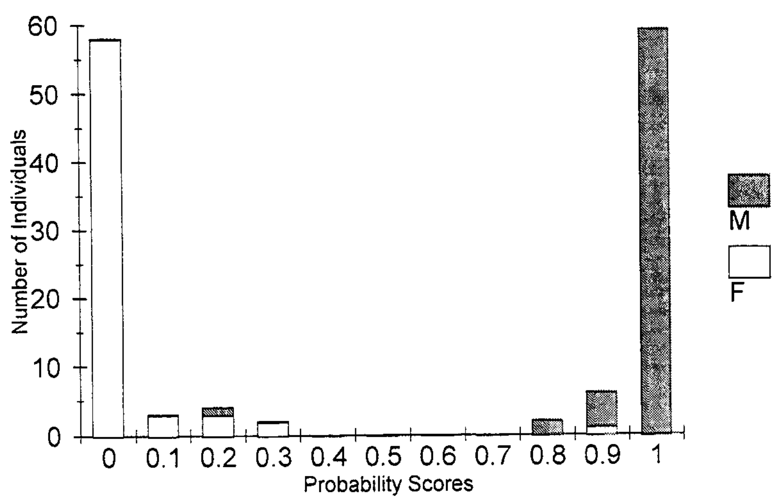


Figure 6.4: Calculated probabilities (P) for the hold-out test sample.

Table 6.3: Assessment of the fit of logistic regression models*

			Tested on Model Sample					Tested on Hold-Out Sample (n = 67 males and 67 females)				
			Female Correct‡		Male Correct‡			Female Correct		Male Correct		
Model†	n	-2LL	n	%	n	%	Total	n	%	n	%	Total
1	401	41	203	98.1	190	97.9	98.0	66	98.5	66	98.5	98.5
2	401	85	198	95.7	184	94.9	95.3	65	97.0	65	97.0	97.0
3	414	75	211	97.7	188	95.0	96.4	64	95.5	66	98.5	97.0
4	404	91	201	96.2	187	95.9	96.0	63	94.0	66	98.5	96.3
5	418	119	208	95.0	186	93.5	94.3	65	97.0	66	98.5	97.8
8	402	46	203	97.6	188	96.9	97.3	65	97.0	66	98.5	97.8
10	402	56	201	97.1	190	97.4	97.3	64	95.5	65	97.0	96.3
11	415	86	208	96.3	186	93.5	95.0	63	94.0	65	97.0	95.5
12	416	81	210	96.8	190	95.5	96.2	62	92.5	65	97.0	94.8
13	402	107	199	96.1	185	94.9	95.5	62	92.5	64	95.5	94.0
14	403	95	195	93.8	183	93.9	93.8	62	92.5	62	92.5	92.5
15	416	98	207	95.4	190	95.5	95.4	62	92.5	64	95.5	94.0
17§	419	111	207	95.0	187	93.0	94.0	60	89.5	63	94.0	91.8
19§	419	155	204	93.6	187	93.0	93.3	63	94.0	61	91.0	92.5
20	401	117	195	94.2	182	93.8	94.0	65	97.0	64	95.5	96.3
22	418	134	209	95.4	183	92.0	93.8	63	94.0	62	92.5	93.3
26	402	122	194	93.7	180	92.3	93.0	64	95.5	61	91.0	93.3

*Only models with an overall allocation accuracy of 90% or higher, and a difference in allocation accuracy between males and females of less than 5% in both the model sample and the test sample are presented. Hence, there are gaps in the model numbers.
†See Table 6.4 for the coefficient for each model.
‡Sample size varies with each model.
§Enter method instead of Forward: LR was used to select independent predictor variable.

Table 6.4: Coefficients for logistic regression models*.

Model†	Innominate Height	Iliac Breadth	SPRL	AIL	Max Di. Of Femur Head	Epicondylar breadth	Constant
1	0.5950	-0.5192	-1.1104		1.1696	0.5893	-61.5345
2		-0.1600	-0.5951	0.2920	1.0365	0.3901	-30.5291
3	0.2572		-0.9852		0.7303	0.3177	-40.5313
4	0.4323	-0.2217	-0.7404	0.3412			-30.3590
5	0.3084		-0.8092	0.2657			-28.3111
8	0.4868	-0.4903	-1.0597	0.2901	1.6241		-45.2528
10	0.5267	-0.3785	-0.8156			0.7758	-52.8262
11	0.2896		-0.8794			0.5783	-42.6362
12	0.2694		-0.9850		1.0484		-32.7486
13		-0.0963	-0.5042	0.4154		0.6453	-32.9320
14		-0.1428	-0.6382	0.3447	1.3968		-22.6111
15			-0.6964	0.2171	0.8489	0.3285	-31.3620
17			-0.6285		0.9906	0.4058	-30.9086
19			-0.4776			0.8032	-28.6274
20	0.2007	-0.4445		0.1734	0.5697	0.3915	-41.9071
22	0.3943		-0.8007				-25.4936
26	0.2326	-0.4321			0.6235	0.4085	-40.2291

*See Table 6.3 for an assessment of the fit of each model.

†See text for an example of the application of a logistic regression model to determine sex.

In various situations, only fragmentary remains are available for analysis. For example, if only the hipbone is available for analysis, then Model 4 would be used to assess sex. With this model, allocation accuracy for the model sample is 96% for each sex. When applied to the test sample, allocation accuracy is 96%: 94% for females and 98.5% for males. The pubis is susceptible to post-mortem damage, and therefore, two models that do not require the SPRL (Models 20 and 26) have also been developed and tested. Even with a test sample that includes very small individuals from the Coimbra Collection and larger individuals from the Terry collection, allocation accuracy is 97% for females and 95.5% for males (96% overall) for Model 20, and 95.5% for females and 91% for males (93% overall) for Model 26. Model 26 does not require either the SPRL or the AIL, and therefore, can be further tested using data from the Forensic Anthropology Data Bank (FDB). The FDB sample (n = 213) used for testing was selected to include only data from positively identified individuals 19 years of age and older and no data from the Terry Collection. Allocation accuracy for Model 26 when applied to the FDB data was 94% overall (89% for females and 96% for males). This approach to testing is equivalent to applying Model 26 in 213 forensic cases and suggests that older collections, such as the Terry Collection, can still be very useful for developing (and not just testing) forensic methods provided that the reference samples are carefully constructed.

The combination of high accuracy and small differences between male and female accuracy in a diverse test sample is essential for any useful sex determination method (Saunders and Hoppa 1997). Therefore, in order to maximize the reliability of the models in predicting sex in various situations, several criteria had to be met for models to be included in Tables 6.3 and 6.4. First, in both the model sample and the test sample, overall allocation accuracy had to be 90% or higher. Second, the difference in allocation accuracy between males and females had to be less than 5%. Both criteria are somewhat arbitrary yet more strict than other published guidelines (Saunders and Hoppa 1997) in order to establish confidence in allocation when the methods are applied in actual cases. A model developed in this study (Model 6) that is analogous to Washburn's (1948) ischium-pubis index or Novotny's ischio-pubic index (Ferembach et al. 1980) does not meet either the 90% accuracy or the 5% sex difference criteria. Model 6 performed acceptably on the model sample: 91% of females and 90% of males were allocated correctly. However, when applied to the test sample, the allocation accuracy for Model 6 is only 88%: 85% and 91% for females and males, respectively. The pubis and the ischium are important sources of information when determining sex but relying only on measurements of these two bones may produce misleading results.

When using sex determination methods it is as important to know when a method may fail as when it may provide useful information. An incorrect assessment of sex can be very misleading in both archaeological contexts and forensic investigations. A case by case review of all the individuals used in the model sample that are allocated incorrectly using Model 1 revealed a pattern (n = 8 or 2% of 401). Females with unusually large joints relative to the size of their pubis (SPRL) are allocated incorrectly. All the males with extremely small joints relative to their pubis (SPRL) are allocated incorrectly. In contrast to many other univariate and multivariate metric methods for determining sex, it is not simply the shorter, less robust males and the taller, more robust females that are allocated incorrectly (for example, see Dibennerdo and Taylor 1982). Very short males from the Coimbra Collection and taller females from the Terry Collection were consistently classified correctly even though the Coimbra males are so much shorter than the Terry Collection females that there is no significant difference in femur length ($t = -1.531$, $p = 0.128$, $n = 64$ females, $n = 58$ males) between males from one collection and females from another when birth cohort is held constant (1875-1899). Furthermore, there are no other patterns in mis-allocation related to age (including 19-21 year old individuals who had not completed growth), sex, or "race." Just over half of the individuals in the sample from the Terry Collection used in this study were described as "Negro" at the time they were included in the collection. The high allocation for all models presented and no difference in allocation accuracy by "race" strongly suggest that race-specific sex determination methods are not necessary.

The following example illustrates how the method is used and its reliability. Model 1 is used to determine the sex of an individual from the test sample from the Coimbra Collection. The data for this individual are as follows: hipbone height is 204 mm, iliac breadth is 151 mm, SPRL is 66 mm, maximum femur head diameter is 42 mm, and epicondylar breadth of the femur is 75 mm. In this case, the femur head is more than one standard deviation smaller than the Coimbra male mean, and equivalent to the Terry female mean. The epicondylar breadth is more than one standard deviation smaller than the Coimbra male mean, less than 1 mm larger than the Terry female mean, and about one standard deviation larger than the Coimbra female mean. Furthermore, this individual had a ventral arc very similar to the pattern described by Sutherland and Suchey (1991) in their Figure 7.

Using Model 1,

$$P = \frac{1}{(1 + e^{-(-61.3545 + 0.5950(204) - 0.5192(151) - 1.1104\,(66) + 1.1696(42) + 0.5893\,(75))})}$$

$P = 0.8147$

Therefore, there is an 81.5% probability that the individual is male despite the small size and the presence of a ventral arc. The documented sex of the individual is male. There is no doubt that the documentary data are correct for this individual and there has been no mixing of identity cards and skeletons. The cause of the death transcribed from hospital records is suicide by gunshot to the head. The cranium of this individual has a clear entrance wound in the palate, clear exit wound in the frontal, and damage to the eye orbits that is consistent with the peri-mortem effects of such a gunshot wound. See Appendix B for more detailed step-by-step instructions for calculating P.

This case can also be used to illustrate the importance of having highly reproducible measurements of the pubis, particulary for borderline cases. Table 6.5 shows how the calculated probability changes depending on the level of intra-observer error. The magnitude and the direction of the error are both critical factors that must be considered. Row one shows the actual data for this individual. The SPRL is 66 mm and all the other measurements -which are not shown in the table- are used to calculate P and determine the sex of the individual. In row two all the data are the same except that the measurement error of 0.6% is added to the SPRL to simulate a positive error in measurement. In row 3, again all things are equal except that a negative measurement error of 0.6% is subtracted from the SPRL to simulate a negative error in measurement. In row 4, the mean error found in the traditional pubis length is added to the SPRL to simulate a larger positive measurement error scenario. In row 5, the mean error found in the traditional pubis length is subtracted from the SPRL to simulate a larger negative measurement error. The critical scenario appears in row 4. In this scenario there is a hypothetical positive measurement error in the SPRL that is equivalent to the mean error of the traditional pubis length. The individual that is definitely male is actually classified as female. A positive measurement error of greater than 2% is the difference between a correct and incorrect allocation. This problem is avoided with the SPRL because in almost 95% of cases, measurement error for the SPRL is less than 2% (see Figure 6.3).

Table 6.5: Changes in calculated probability (P) with different intra-observer errors*.

Scenario†	SPRL (mm)	Calculated Probability (P)	Predicted Sex
Actual Data	66	0.8147	Male
SPRL + 0.6%	66.396	0.7390	Male
SPRL - 0.6%	65.604	0.8722	Male
SPRL + 2.7%	67.782	0.3780‡	Female
SPRL - 2.7%	64.218	0.9695	Male

*Model 1 is used to illustrate how P changes with various hypothetical errors for the SPRL.
†All independent variables are kept constant except for the SPRL which is modified by the amount indicated.
‡An error equivalent to the mean error for the traditional pubis length is the difference between a correct and incorrect allocation.

DISCUSSION

Previous research has shown that the applicability of sex determination methods can be restricted by the reference sample used to develop the method (Albanese 1997a, Albanese 1997b, MacLaughlin and Bruce 1990, Saunders and Hoppa 1997). Regardless of the statistical approach, single measurement sex determination methods can be particularly susceptible to these problems because the methods are dependent on absolute size differences in means of males and females for any given measurement. However, there is evidence that some non-metric pelvic methods may also have similar constraints related to the limits imposed by the reference sample. The high allocation accuracy (96%) for the Phenice method (1969) for a sample from the Terry Collection was not duplicated by MacLaughlin and Bruce (1990) on three separate samples from England (82%), the Netherlands (68%), and Scotland (59%), by Lovell (1989) on a sample of medical school cadavers from British Columbia (83%), nor by Rogers and Saunders (1994) on a 19th century cemetery sample from Ontario (88%). Similarly, using only the ventral arc, Sutherland and Suchey (1991) had an overall accuracy of 96% for a large sample of individuals autopsied in the County of Los Angeles, while Rogers and Saunders (1994) had an accuracy of 87% on the cemetery sample from Ontario. These are differences in allocation accuracies that are comparable to those seen when some single measurement sex determination methods are applied across diverse samples (see Saunders and Hoppa 1997). Ubelaker and Volk (2002) suggest that this difference in accuracy when using the Phenice method may be due to the experience of the investigator; however, their methodology makes it difficult to separate the effects of experience from other issues. Ubelaker and Volk's results follow the pattern described by Rogers and Saunders (1994) and the results from this current study: accuracy when using only data from the pubis and ischium (Phenice method or SPRL and AIL) is between 85% and 90%, but accuracy for a combination of data from the pubis and other parts of the pelvis is over 95%. Lovell (1989) did not find any significant differences in allocation accuracy based on experience. Maclaughlin and Bruce (1990) did find that the difference in allocation between experienced and inexperienced was significant but the allocation accuracy for experienced investigators was still only 78.9%.

Despite some of the applicability problems reported for some metric methods, the benefits of metric methods are usually the ease of measurement as opposed to scoring presence, absence, or pronouncement of a trait, and the ease of statistical analysis particularly when large samples are involved (see also Giles 1970). Actually, metric and non-metric approaches are often different ways of assessing the same variation. Two independent studies (Anderson 1990, Budinoff and Tague 1990) have shown that differential growth in females at the symphyseal end of the pubis is responsible for the presence of a ventral arc in most females and its absence in most males. Measurement of the pubis or the scoring of the ventral arc should assess the same sexual dimorphism in the pubis, and it is recommended that the traditional pubis length and the ventral arc should not be considered two independent sources of information for assessing sex (Budinoff and Tague 1990). The example above used to demonstrate the logistic model suggests that there are some exceptions. Discrepancies between methods may be due to a combination of factors including (1) measurement error for metric approaches (a good example is the traditional pubis length in this current study); (2) lack of well defined standards in scoring non-metric traits (MacLaughlin and Bruce 1990); and (3) idiosyncracies of individuals in specific cases who have a combination of what are considered typical male or typical female traits (such as having a relatively short pubis and a ventral arc).

The very high allocation accuracies for the model sample, the test sample, and supplementary testing on forensic data from such different collections as the Terry Collection, the Coimbra Collection, and the FDB suggest that the metric methods presented here can be applied in a wide range of cases. The high accuracy of the methods, particularly Model 1, can be attributed to several factors. First, the very low intra-observer error associated with the SPRL measurement is critical in determining sex, particulary with borderline cases when other methods may fail.

Second, a wide range of human variation was sampled and included in the reference sample. This non-random approach to sampling (1) considered the historical context of the collections in order to minimize bias and maximize representativeness, (2) considered other sources of variation such as year of birth and age at death even if these variables were not used as predictor variables, (3) included data from two very different collections, and (4) did not divide the reference sample into racial categories.

Third, the combination of long bone data *and* pelvic data contributes a great deal to the accuracy of the method. It is recommended in many sources (for example, Krogman and İşcan 1986) that all available data be used from the entire skeleton when attempting to determine sex. However, because of the emphasis placed on the pelvis when determining sex, it is implied that if the pelvis is complete, other data from the rest of the skeleton, with the exception of the cranium, do not have to be considered since the pelvis is the most sexually dimorphic part of the skeleton. Others bluntly state, "in many cases, particularly those involving the cranium and pelvis, qualitative morphological observations are sufficient for accurate sex attribution," (Holman and Bennett 1991:421). Qualitative morphological data from only the pubis would have resulted in an incorrect assessment of sex in the example from the Coimbra Collection presented in this paper. Data from long bones only tend to be considered if the pelvis is not recovered or is too damaged for assessment. Most published descriptions of sex determination methods that use bones other than the pelvis or cranium (for example, Berrizbeitia 1989, Black 1978, France 1983, Holman and Bennett 1991, Marino 1995) begin by stating that there is a need for "other" methods for determining sex in cases where both the cranium and the pelvis are not recovered or are too damaged for analysis. The non-pelvic/non-cranial methods are never tested in conjunction with pelvic and/or cranial sex determination methods and it is never recommended that the method be used in conjunction with other data. Even Schulter-Ellis and colleagues (1983, 1985) who clearly describe the value of including the femur when determining sex, defered to the femur head measurement only when they suspected that their hipbone method was not correct. This current study shows that it makes a great deal of biological sense to draw information from both pelvic measurements and femur measurements when determining sex.

The femur in general, and the femur joints in particular, are highly sexually dimorphic, and can be used to estimate the overall size of a person. The size of the pubis *relative* to the hipbone and the femur is a highly effective way of maximizing accuracy when determining sex. Furthermore, by including the maximum femur head diameter and the SPRL, it is possible to assess the relative length of the symphyseal end of the pubis -where there is differential growth between sexes- with the acetabular end of the pubis, which is approximated by the maximum femur head diameter. Paradoxically, leaving out the acetabular portion of the pubis when using the SPRL measurement results in more information, provided that the size of the acetabulum is considered in some way. The benefits of including femur data whenever it is available is reflected in the increase in allocation accuracy by about 2% from Model 4 (hipbone only) to Model 1 (hipbone and femur) and by the total elimination of the difference in allocation accuracy between males and females. See Table 6.3 for details. Even when the pubis is damaged and the SPRL measurement cannot be collected, allocation accuracy is still very high for Model 22 and Model 26 because data from both the hipbone and the femur are used to determine sex.

CONCLUSION

This study demonstrates that allocation accuracy of better than 98% is possible for very different samples from Europe and North America from the 19^{th} and 20^{th} centuries using a metric sex determination method. This high accuracy for such diverse samples can be attributed to several factors. First, the superior pubis ramus length measurement, a new, highly reproducible alternative of the traditional pubis length measurement is used. Second, historical and demographic information -including year of birth, and age at death- was used to construct reference samples that included a substantial amount of normal human skeletal variation. Third, data from the pelvis and a long bone are considered together to determine sex. Fourth, race-specific methods were not developed.

ACKNOWLEDGMENTS:

I would like to thank the administration at the Museum of Anthropology at the University of Coimbra in Portugal for access to the Coimbra Collection and the curators of the FDB at the University of Tennessee, Knoxville for data. Thanks to Dr. Shelley Saunders for assistance throughout the research, Hugo Cardoso for the illustration in Figure 6.2, and Sandra Albanese for her assistance with database management and editorial suggestions.

REFERENCES CITED

Albanese J. 1997a. A comparison of the Terry Collection and the Grant Collection using the head of the femur and head of the humerus: Implications for determining sex. Can Soc Forens Sci J 30:167.

Albanese J. 1997b. Skeletal variability in recent North Americans: Implication for the development of new sex determination methods. Paper presented at the 25th Annual Meeting of the Canadian Association for Physical Anthropology.

Albanese J. 2001a. Measurement error and metric methods for sex determination. Paper presented at the 29th Annual Meeting of the Canadian Association for Physical Anthropology.

Albanese J. 2001b. A New Metric Method for Sex Determination Using the Pelvis and Femur. Am J Phys Anthropol Suppl 32:31.

Albanese J. 2002. The use of skeletal data for the study of secular change: Methodological implications of combining data from different sources. Am J Phys Anthropol Suppl 34:36.

Angel L. 1976. Colonial to modern skeletal changes in the U.S.A. Am J Phys Anthropol 45:723-736.

Anderson BE. 1990. Ventral arc of the os pubis: Anatomical and developmental considerations. Am J Phys Anthropol 83:449-458.
Angel L. 1976. Colonial to modern skeletal change in the U.S.A. Am J Phys Anthropol 45:723- 736.

Bass WM. 1987. Human Osteology: A Laboratory and Field Manual. Third Edition. Columbia: Missouri Archaeological Society.

Berrizbeitia EL. 1989. Sex determination with the head of the radius. J Forensic Sci 34:1206-1213.

Black K. 1978. A new method for assessing the sex of fragmentary skeletal remains: femoral shaft circumference. Am J Phys Anthropol 48:227-232.

Bruzek J. 1992. Fiabilité des fonctions discrimininantes dans la détermination sexuelle de l'os coxal critiques et propositions. Bull et Mém de la Société d'Anthropologie de Paris 4:67-104.

Bruzek J. 2002. A method for visual determination of sex, using the human hip bone. Am J Phys Anthropol 117:157-168.

Budinoff LC, Tague RG. 1990. Anatomical and developmental bases for the ventral arc of the Human Pubis. Am J Phys Anthropol 82:73-79.

Buikstra JE, Ubelaker DH. 1994. Standards for Data Collection from Human Skeletal Remains. New York: Wiley-Liss.

Cunha E. 1995. Testing identification records: Evidence from the Coimbra Identified Skeletal Collections (Nineteenth and Twentieth centuries). In SR Saunders and A Herring (eds.), Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, pp. 179-198.

Dibennerdo R, Taylor JV. 1982. Classification and misclassification in sexing the black femur by discriminant function analysis. Am J Phys Anthropol 58:145-151.

Dudar JC, Dupras TL. 1996. Sex Estimation from the adult human rib: What to use? Discriminant function or logistic regression? Am J Phys Anthropol Suppl 22:99-100.

Ericksen MF. 1982. How "representative" is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 59:345-350.

Ferembach D, Schwidetzky I, Stoulkal M. 1980. Recommendations for age and sex diagnoses of skeletons. J Hum Evol 9:517-594.

France DL. 1983. Sexual Dimorphism in the Human Humerus, University of Colorado, Boulder.

Fuller K. 1998. Adult females and pubic bone growth. Am J Phys Anthropol 106:323-328.

Giles E. 1964. Sex determination by discriminant function analysis of the mandible. Am J Phys Anthropol 22:129-136.

Giles E. 1970. Discriminant function sexing of the human skeleton. In TD Stewart (ed.), Personal Identification in Mass Disasters. Washington, DC: National Museum of Natural History, pp. 99-109.

Holman DJ, Bennett KA. 1991. Determination of sex from arm bones. Am J Phys Anthropol 84:421-426.

İşcan MY. 1990. A comparison of techniques on the determination of race, sex and stature from the Terry and Hamann-Todd Collections. In GW Gill and S Rhine (eds.), Skeletal Attribution of Race. Albuquerque: Maxwell Museum of Anthropology, pp. 73-81.

İşcan MY, Miller-Shavitz P. 1984. Discriminant function sexing of the tibia. J Forensic Sci 29:1087-1093.

İşcan MY, Yoshimo M, Kato S. 1994. Sex determination from the tibia: Standards for contemporary Japan. J Forensic Sci 39:785-792.

Kieser JA, Moggi-Cecchi J, and Groeneveld HT. 1992. Sex allocation of skeletal material by analysis of the proximal tibia. Forensic Sci Int 56:29-36.

Krogman WM, İşcan MY. 1986. The Human Skeleton in Forensic Medicine. Second Edition. Springfield: CC Thomas.

Lovell NC. 1989. Test of Phenice's technique for determining sex from the os pubis. Am J Phys Anthropol 79:117-120.

MacLaughlin SM, Bruce MF. 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. J Forensic Sci 35:1384-1392.

Marino EA. 1995. Sex estimation using the first cervical vertebra. Am J Phys Anthropol 97:127- 133.

Meindl RS, Lovejoy CO, Mensforth RP, Carlos LD. 1985. Accuracy and direction of error in the sexing of the skeleton: Implications for Paleodemography. Am J Phys Anthropol 68:79-85.

Moore-Jansen PM, Ousley SD, Jantz RL. 1994. Data Collection Procedures for Forensic Skeletal Material. Third Edition. Knoxville: University of Tennessee Department of Anatomy.

Norušis MJ. 1990. SPSS Advanced Statistics Student Guide. Chicago: SPSS Inc.

Olivier G. 1969. Practical Anthropology. Springfield: CC Thomas.

Ousley SD, Jantz RL. 1996. FORDISC 2.0: Personal Computer Forensic Discriminant Functions. Knoxville: The University of Tennessee.

Ousley SD, Jantz RL. 1998. The forensic data bank: documenting skeletal trends in the United States. In K Reichs (ed.), Forensic Anthropology. Second Edition: Springfield: CC Thomas, pp. 441-458.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297-301.

Rocha MA. 1995. Les collections ostéologiques humaines identifiées du Musée Anthropologigue de l'Université de Coimbra. Anthropol Port 13:7-38.

Rogers T, Saunders S. 1994. Accuracy of Sex determination using morphological traits of the human pelvis. J Forensic Sci 39:1047-1056.

Santos AL. 2000. A Skeletal Picture of Tuberculosis. Doctoral thesis, University of Coimbra, Coimbra.

Santos AL, Roberts CA. 2001. A picture of tuberculosis in young Portuguese people in the early 20th century: a multidisciplinary study of the skeletal evidence. Am J Phys Anthropol 115:38-49.

Saunders SR, Hoppa RD. 1997. Sex allocation from long bone measurements using logistic regression. Can Soc Forens Sci J 30:49-60.

Schulter-Ellis FP, Hayek LA, Schmidt DJ. 1985. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part II. J Forensic Sci 30:178-85.

Schulter-Ellis FP, Schmidt DJ, Hayek LA, Craig J. 1983. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I. J Forensic Sci 28:169-80.

Schultz AH. 1930. The skeleton of the trunk and limbs of higher primates. Am J Phys Anthropol 2:303-438.

Smith SL. 1997. Attribution of foot bones to sex and population groups. J Forensic Sci 42:186- 95.

SPSS. 1998. SPSS 9.0 for Windows. Chicago: SPSS Inc.

Steele D, Bramblett CA. 1988. The Anatomy and Biology of the Human Skeleton. College Station: Texas A&M University Press.

Stewart TD. 1979. Essentials of Forensic Anthropology. Springfield: CC Thomas.

Sutherland LD, Suchey JM. 1991. Use of the ventral arc in sex determination. J Forensic Sci 36:501-511.

Tague RG. 1994. Maternal mortality or prolonged growth: Age at death and pelvic size in three prehistoric Amerindian Populations. Am J Phys Anthropol 95:27-40.

Terry RJ. 1940. On measuring and photographing the cadaver. Am J Phys Anthropol 26:433-447.

Thieme FP. 1957. Sex in the Negro Skeleton. J Forensic Med 4:72-81.

Trotter M. 1981. Robert James Terry, 1871-1966. Am J Phys Anthropol 56:503-508.

Ubelaker DH. 2000. Methodological considerations in the forensic applications of human skeletal biology. In MA Katzenberg and SR Saunders (eds.), Biological Anthropology and the Human Skeleton. New York: Wiley-Liss, pp. 41-67.

Ubelaker DH, Volk GC. 2002. A test of the Phenice method for the estimation of sex. J Forensic Sci 47:19-24.

Walker PL. 1995. Problems of Preservation and Sexism in Sexing: Some Lessons from Historical Collections for Palaeodemographers. In SR Saunders and A Herring (eds.), Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars press, pp. 31-47.

Washburn SL. 1948. Sex differences in the pubic bone. Am J Phys Anthropol 6:199-207.

CHAPTER 7

Discussion and Conclusion

NEWER REFERENCE COLLECTIONS: ARE THEY FREE OF BIAS?

In 1981, the American Association of Physical Anthropologists (AAPA) celebrated its 50th annual meeting with a special session on the history of physical anthropology in the United States[1]. This session was published later the same year in a special issue of Volume 56 of the *American Journal of Physical Anthropology*[2]. Some of the papers discussed historical changes in the discipline while others focused on specific physical anthropologists and their contributions to the field including mini-biographies of Robert J. Terry, and T. Wingate Todd (Trotter 1981, Cobb 1981, respectively). These pioneers of human skeletal research were responsible for the establishment of two collections that include thousands of identified human skeletons that have been used extensively for the study of human variation, the development of skeletal identification methods and as reference samples for the study of human evolution.

---

1 The meeting in 1981 marked the 50th Anniversary of the first independent meeting of the AAPA. The first meeting of the AAPA was held in 1930 in conjunction with the American Association of Anatomists in Charlottesville, Virginia largely for "pragmatic and political reasons" since most of the members of the new association were anatomists (Spencer 1981:360).

2 Some of the papers that appear in this issue of Volume 56 were first presented at a special symposium on December 1-12, 1980 in Charlottesville, Virginia.

While some physical anthropologists at those meetings in 1981 in Detroit, Michigan, were celebrating the past, others were discussing what appeared to be a dim future for skeletal biology in the United States with the lack of more recent identified skeletal reference collections that could rival the size of the Terry or Hamann-Todd Collections (Ericksen 1982, Ousley and Jantz 1998). The source of a large number of identified skeletons, through an association with an anatomy department, had been outside the reach of skeletal biologists for decades, and there was little hope for any change in the foreseeable future. Todd's premature death in 1938 (vacating the chair of the Anatomy Department at Western Reserve University) and Terry's retirement less than three years later (vacating the chair of the Department of Anatomy at Washington University) were two critical steps in the separation of physical anthropology and anatomy in North America. Mildred Trotter's retirement from the Anatomy Department at Washington University in 1967, and W. Montague Cobb's retirement from the Laboratory of Anatomy and Physical Anthropology at Howard University in 1969 (Rankin-Hill and Blakey 1994) completely severed the ties between the two disciplines. Through Trotter's efforts, the Terry Collection was transfered to the Smithsonian Institution in 1967 and avoided the fate of the Hamann-Todd Collection (Trotter 1981). For a period in the 1940s after the Hamann Museum at Case Western Reserve University was closed, the Hamann-Todd Collection was placed in crates and stored in a number of locations including basements, closets and attics (Jones-Kern and Latimer 1996). At one point some consideration was given to burying the entire collection, but instead, the collection was slowly moved to the Cleveland Museum of Natural History over a 20 year period beginning in 1951 (Jones-Kern and Latimer 1996).

If it seemed in the late 1970s and early 1980s that new, large and well documented skeletal collections were not possible, then the apparent void in reference data, particularly for forensic identification methods, might be filled in other ways. In 1982, a four-member committee convened at the 34th Annual Meeting of the American Academy of Forensic Sciences and began a process that led to the establishment of the Forensic Anthropology Data Bank (FDB) at the University of Tennessee, Knoxville (Ousely and Jantz 1998). The rationale behind the establishment of the FDB was, and is, that the main reference collections in the U.S.A. (Terry and Hamann-Todd Collections) that have been used to develop many forensic identification methods, are not representative of the greater population in the U.S.A. because of the source of the skeletons: anatomy school cadavers. Furthermore, the age of the collections (with the majority of the individuals in the collections having birth years in the 19th century) does not reflect the secular changes in stature that have occurred in the 20th century (Ousley and Jantz 1998). Some testing of race and sex determination methods seemed to support these assessments and it was concluded that the methods derived from the Terry Collection or the Hamann-Todd Collection are not applicable to forensic cases (Ousley and Jantz 1998).

The FDB was to serve as a more recent source of data that could be used to develop new identification methods that are applicable to more recent forensic investigations. Examples of these new methods include two versions of the computer application FORDISC (Ousley and Jantz 1996). In order to correct the problem with secular changes, individuals included in the database were born around or after 1900 and most were born in the 20^{th} century (Ousley and Jantz 1998). A "significant part" of the FDB including a "a good number" of the "Black" males in the reference sample used in FORDISC are individuals from the Terry Collection that meet this YOB criterion (Ousley and Jantz 1996:26, 28)[3].

The FDB does serve the purpose of preserving data that would have been lost. Management of the data at a central location with standardized procedures means that the data can be used for many different research projects, but a virtual collection cannot replace the research potential of a skeletal reference collection (Moore-Jansen et al 1994, Ousley and Jantz 1998). However, a new collection does not necessarily result in a less biased or more representative collection. The FDB is not free of bias and does not necessarily eliminate or control for any of the biases in the older skeletal collections that it was established to replace. As with the older collections, problems arise from how the collection is amassed and not only the source and age of the data or skeletons. The data that are included in the FDB are no more of a "skeletal census" of the U.S.A. than any of the anatomical reference collections. Aside from the data collected from the Terry Collection, skeletal data must follow a specific set of steps in order to be included in the FDB. First, the individual must die in a manner that is of forensic interest (homicide, suicide, mysterious disappearance, related to insurance settlements, etc.). Second, the

---

3 Two separate reference databases can be selected when using FORDISC 2.0: a sub-sample of positively identified individuals with complete data from the FDB and Howells' (1973, 1989) data (Ousley and Jantz 1996). A small number of individuals from the Hamann-Todd Collection are included in the FDB (Ousley and Jantz 1998) but are not mentioned in the description of the FORDISC 2.0 sample (Ousley and Jantz 1996).

remains must be in a state of decomposition where the regular authorities (coroner, forensic pathologist, forensic examiner, etc. depending on the jurisdiction) deem it necessary to seek the assistance of a forensic anthropologist.

Jantz and Moore-Jansen (1988) state that the FDB is broadly representative of the contemporary population in the United States as seen by forensic anthropologists. In other words, since the FDB was drawn mostly from forensic cases, it is a good reference collection to develop new sex and race determination methods specifically for forensic cases. The main problem with this assessment is that several criteria must be met before data from any given individual can be useful for research purposes, such as the development of forensic identification methods: a forensic anthropologist must collect the data and submit them to the FDB; the individual must be positively identified; the skeleton must be relatively complete and in good enough condition for 10-20 or more measurements to be collected from different skeletal elements. The *usable* data in the FDB are no more a random sample of forensic cases in the U.S.A. than either the Terry Collection or the Hamann-Todd Collection is of the population in the U.S.A.

There are important differences between the anatomical reference collections and the FDB. The collection process for the FDB is different than the collection process used to amass the Terry and Hamann-Todd Collections. The FDB also samples a different segment of the American population than do either the Terry or Hamann-Todd Collections. The demographic profile of the FDB is clearly different from that of the Terry Collection (Ousley and Jantz 1998). However, while there are many more older adults in the Terry Collection, because of the size of the collection it would be a grave error to assume that there are few younger adults in the Terry Collection. There are over 250 males (26.5% of males) and over 140 females (21.5% of females) aged 18-40[4] in the collection that can be sampled if younger adults are required for any given research project. Furthermore, the tendency towards younger adults in the FDB is still a bias that has resulted from the collection process and the source of the data.

The age bias of the usable portion of the FDB has been magnified by a combination of factors related to the year of birth (YOB) criteria for inclusion in the FDB and the source of a significant portion of identified and usable data. One hundred and thirty five individuals of the positively identified cases in the FDB are Terry Collection individuals who were born after 1898. Since the skeletons were added to the Terry Collection in the mid-1960s, most of the individuals included are also younger adults. Controlling for YOB when sampling the Terry Collection magnified the bias towards younger individuals in the FDB. Depending on the manner of sampling of the FDB, the bias towards younger individuals can have an impact not only on the assessment of age, should the FDB ever be used in this manner, but also on the development of sex determination methods as described in Chapters 2 and 6.

The approach to sampling the Terry Collection for inclusion in the FDB has resulted in another major bias in the database. The proportion of Terry Collection

---

4 The lower parameter is the age when individual are socially and legally considered adults, even if skeletal growth is not complete. The upper parameter of 40 years is arbitrary and used only to illustrate the point. Percentages were calculated by dividing by the number of individuals of the respective sex with confirmed ages.

individuals of the usable data in the FDB increases sharply as more variables are selected for analysis from the FDB. This increased proportion of Terry Collection data is not in itself necessarily a problem. However, the interpretation of the sources of the skeletal variation is problematic. The proportions of usable Terry Collection data vary by "race" and sex; and specific variables and the number of variables selected for analysis. Table 7.1 is a summary of the overall proportion of Terry "Blacks" and "Whites" in the FDB and variation in sample proportion based on variables selected for "race" and sex groups. Note how the proportion of Terry Collection individuals in the *analyzable* sample varies by "race," sex, and whether cranial or infra-cranial measurements are analyzed. For example, a comparison of infra-cranial dimensions (maximum femur length and maximum femur head diameter) of "Black" and "White" females may produce misleading "racial differences" that stem from the fact that 51% of the "Black" sample consists of Terry Collection females whereas only 18% of the "White" sample consists of Terry Collection females. Regardless of the bias in the Terry Collection or the FDB, the results presented in Chapter 5 suggest the Terry Collection and the non-Terry portion of the FDB are sampling very different populations. At least some of the variation attributed to "race," and possibly sex as well, results from variation in the proportion of the Terry Collection in the analyzable FDB sample. Given these varying proportions of Terry Collection individuals in the FDB, the representativeness of the FDB will vary depending on which variables are selected to develop forensic identification methods. The source and direction of the bias is different for the FDB and the anatomical reference collections, but there is no evidence that suggests that the bias introduced through the FDB

Table 7.1: Variation in the Proportion of Terry Collection individuals in the FDB by "race" and sex depending on variables selected for analysis*.

	"Blacks"					"Whites"				
	Both Sexes	Female		Male		Both Sexes	Female		Male	
	ID[1]	Cranial[2]	Infra[3]	Cranial[2]	Infra[3]	ID[1]	Cranial[2]	Infra[3]	Cranial[2]	Infra[3]
Terry Collection in FDB	85	42	33	43	34	46	31	22	14	6
Total in FDB	325	145	65	172	90	566	236	117	289	161
% of Terry Collection in FDB	26%	29%	51%	25%	38%	8%	13%	18%	5%	4%

*Note how the proportion of Terry Collection individuals in the FDB varies by "race", sex, and whether cranial or infra-cranial measurements are analyzed.

[1] Positively identified individuals of both sexes.

[2] Sample sizes if two commonly collected cranial measurements are analyzed: maximum cranial width and maximum cranial breadth.

[3] Sample sizes if two commonly collected infra-cranial measurements are analyzed: maximum femur length and maximum femur head diameter. Data from left side presented.

collection process is having less of an effect on variation in the database than the bias introduced when the Terry or Hamann-Todd Collections were amassed. The bias in, for example, the Terry Collection is much more transparent. As illustrated in Chapter 3, it is possible to identify sources of bias in the Terry Collection because of the emphasis on record keeping and cataloguing. Although the bias introduced in the FDB through the sampling of the Terry Collection can be assessed, the effects of other sources of bias in the FDB related to the collection process are less clear: Why do certain physical/forensic anthropologists contribute data while others do not? Which cases are contributed and why? As with all reference collections, it may be problematic to assume that the FDB is free of bias. As much caution is required when sampling the FDB as when sampling other older reference collections.

ANATOMICAL REFERENCE COLLECTIONS: LEVELS OF REPRESENTATIVENESS

For decades, the Terry Collection and the Hamann-Todd Collection have been described as not representative of the populations in the U.S.A. (Giles 1964). More recently, others have suggested that these collections may no longer be useful for the development of forensic identification methods (Ousley and Jantz 1998). However, with one exception (Ericksen 1982), the impact of bias on skeletal variation in the collections has never been assessed. Two major confounding factors with Ericksen's study are secular changes and age-related changes, and Ericksen concludes, "that differences between groups could be artifacts of imperfect age matching and small sample size..."

(1982:350).

Ericksen (1982) compared age-related changes in the proximal femurs of "White" Terry Collection females that were unclaimed remains ("Terry regulars") with "White" females from the collection that were bequeathed for scientific research ("Terry willed"). Her underlying assumption is that the individuals who bequeathed their remains are of higher socio-economic status than individuals from the rest of the collection; however, "the flaw in this assumption, is of course, that the status achieved by the time of signing a donor form does not necessarily reflect lifelong status" (Ericksen 1982:349).

Some of the mechanisms and effects of how this change in status through life may affect variation in the collection are obvious. For example, femur length only reflects living conditions during the growth period. Even if the growth period is extended for catch-up growth, once growth is completed, no improvement in living conditions of any magnitude will result in more growth in long bone length. In another more complex example, external and internal robustness of the shaft of the femur will reflect the competing effects of activity throughout life and the loss of robusticity due to illness or old age which reflect behaviours that are affected by socio-economic status. Matters are confounded because socio-economic status (living conditions, behaviour, activity, occupation, etc.) at the time of death is a contributing factor as to whether the individual is included in the collection.

As discussed in Chapters 3 and 4 and following Ericksen's approach, it should not necessarily be assumed that all the Terry regulars were of a lower socio-economic status throughout their lives. The biological costs of poverty to those who were eventually included in the Terry or Hamann-Todd Collections cannot be dismissed (Terry 1940). However, it is an oversimplification to dismiss both collections based on this indicator. Additionally, the effects of low socio-economic status varied in different communities. As illustrated in Chapter 5, Coimbra Collection females are significantly shorter than Terry Collection females when birth cohort is held constant and Coimbra Collection males are not significantly taller than Terry Collection females in some birth cohorts. At least some of the differences in mean femur length between the Coimbra and the Terry Collections can likely be attributed to living conditions for two groups that were of low socio-economic status in their respective communities. This conclusion is supported by the difference in the pattern of completion of skeletal growth in the two collections. There is a clear pattern of delayed completion of growth in the Coimbra Collection. Unlike the Terry Collection, none of the individuals under 20 years of age in the Coimbra Collection were included in the skeletal sample because they did not meet growth criteria: at least one epiphysis is completely separate on at least one bone besides the sternal end of the clavicle. This pattern, associated with an extended catch-up growth period, is consistent with the smaller mean stature for both males and females in the Coimbra Collection and poorer living conditions.

Ericksen investigated the representativeness of the Terry Collection through some internal testing of variation. Ousley and Jantz (1998) have compared the older anatomical collections and the FDB through direct comparisons and by testing older forensic identification methods using FDB data. They conclude that secular changes in the U.S.A. have made obsolete the methods developed from these anatomical reference collections.

Although sex and race are often considered together in identification methods (reviewed in Chapter 2), the problems with testing these methods arise from different issues and will be discussed separately.

There are two factors that need to be considered with metric sex determination methods. First, with a few exceptions (such as Boulinier 1969, Bruzek 1992, Saunders and Hoppa 1997), published metric sex determination methods are almost never extensively tested on large and/or independent samples. A decrease (or at least a difference) in allocation accuracy of any given model would be expected when the method is applied to a truly independent sample than the one used to develop the method.

Second, the problems with many metric sex determination methods relate to the criteria (age, sex and "race") used to select a reference sample and statements of applicability, not necessarily the reference collection that is used. A series of models for sex determination with high allocation accuracies (presented in Chapter 6) follow a new methodology for sampling the Terry Collection and the Coimbra Collection (described in Chapter 4). When tested on FDB data, the allocation accuracy remains high (94%). This approach to testing using FDB data is equivalent to applying the Model 26 in 213 forensic cases and suggests that older collections, such as the Terry Collection, can still be very useful for developing (and not just testing) forensic methods, provided that the reference samples are carefully constructed.

The disastrous performance of race determination methods (Giles and Elliot 1962, İşcan and Cotton 1990) tested using the FDB and other data described by Ousley and Jantz (1998) is due to an entirely different issue than the problems with some metric sex determination methods. Ousley and Jantz attribute the poor performance of race determination methods to secular changes,

> given the secular changes in BBH (skull height) of whites, the relationships of other American Indians and whites has changed: whites now have higher mean skull heights than American Indians. Thus, the Giles and Elliot [1962] white-Indian function weighs skull height in the wrong direction for modern American Indians and whites (Ousley and Jantz 1998:447-448).

Additionally in the "User's Guide" that accompanies the computer application FORDISC 2.0, the same authors state, "using 8-10 craniometrics... American white males born 1840-1890 can be separated from American white males born 1930-1980 with 88-96% accuracy" (Ousley and Jantz 1996:22). The results present in Chapter 5 strongly suggest that the patterns of secular changes they identify are the result of how they constructed their samples. It is not surprising that they found that the greatest "secular changes" have occurred in "White" males. Differences in the apparent magnitude of secular change between sex and "race" sub-samples are expected simply because the proportion of the individuals from the Terry Collection in the FDB varies by sub-sample and is smallest for "White" males (Table 7.1).

There is an alternative interpretation of Ousley and Jantz's (1996) ability to discriminate between "White" males born in the 20th century versus the 19th century: craniometric morphology does not correspond with racial categories or continental origin. Some of the evidence that supports this alternative conclusion includes 1) the review of the methodological issues for the study of secular change discussed in Chapter 5; 2) the ability to predict the "race" of females in the Terry Collection with an allocation accuracy of 69.5% using only age at death and YOB[5] (Chapters 3 and 6); 3) the high allocation accuracy for the metric sex determination method independent of "race" (Chapter 6); 4) the poor performance of various race determination methods (Ousley and Jantz 1998, Goodman 2000); 5) Relethord's (1994, 2001, 2002) results using the same Howells' craniometric data that is used in FORDISC.

This lack of concordance between a suite of traits and a given race is not a new idea. Using anthropometric data from the Hamann-Todd Collection and data from live subjects, Cobb (1936) clearly demonstrated the same non-concordance between race and morphology 60 years ago when secular change was not a confounding factor in reference collections. More recently, when FORDISC 2.0 was extensively tested using large independent samples, allocation accuracies when determining race are very poor regardless of whether Howell's data are selected as a reference sample or the FDB reference sample is used (Ginter 2001, Belcher et al 2002, Leathers et al 2002; Ginter 2002). The poor performance of these methods is due to the non-concordance between discrete categories based on folk taxonomies and actual continuous human variation.

In many cases the poor performance of various forensic identification methods can be attributed to problems with the methods or underlying assumptions, rather than the

---

5 Similarly, it is possible to predict "race" with an allocation accuracy of 66% using only age at death and YOB as independent variables using a logistic regression model for a sample of all the males over 18 years of age in the Terry Collection (n = 987). Considering Trotter's efforts on collecting younger "White" females, a lower allocation accuracy is expected for males using only age and YOB as predictor variables. However, an allocation accuracy greater than chance suggests a similar bias related to age and YOB in males but to a lesser extent than in females.

reference collections. Research bias, bias in the research field, and greater cultural biases have had, and continue to have, an impact on the composition of reference collections and the research questions addressed with those collections. Recognizing those biases in their social and historical context is the first critical step toward addressing the biases. Several examples presented in this thesis illustrate some of the problems associated with ignoring these biases in older reference collections (and methodologies), and some of the benefits of identifying and controlling the biases through careful sampling. It is important to note that random sampling does not correct for or control any of these biases. For example, in the case of age and YOB effects, a randomly selected sample from the FDB will result in a high proportion of young adults and the age-related variation associated with older adults (particularly older females) will be under-represented in the reference sample. Conversely, a random sample of the Terry Collection will result in a high proportion of older adults in the sample and the age-related variation associated with younger adults (particularly younger males) will be under-represented in the reference sample. The Terry Collection contains disproportionately more older adults, but it is possible to sample around this bias. Because of the overall size of the collection, there still are a large number of younger individuals that can be included in a sample, provided that documentary data besides sex are considered.

CHANGE IN CONCEPT OR A CHANGE IN TERMINOLOGY?

The debate over the relevance of race determination methods is a separate issue from the discussion of the use of racial groups or continental origin for the analysis of human variation. Goodman argues that the continued typological approach to race determination using morphological or metric approaches in forensic anthropology is very misleading,

> What is the take home message? From our reading, and I believe even more so from the naive student's reading and use of this material [casts by France Casting], human variation is reduced to how well it fits ideal types... Students do not learn about the continuous and nonconcordant nature of human variation. They learn nothing of the complexities of biology. Rather they are sold the comfortingly simple story: there are old and static ideal types and with a minimum of training one can play the game of fitting the crania to ideal types (1995:224).

Some forensic anthropologists disagree with arguments such as Goodman's because race determination is considered a necessity for identifying unknown individuals in countries where socially defined race is a part of ante-mortem identification (Kennedy 1995). Others recognize that race assessment in a forensic context, at times, amounts to using statistical associations to "test relationships between these cultural categories and metric variation" (Ousley and Jantz 1996:20). When law enforcement officials in the U.S.A. ask forensic anthropologists for assistance, they want to know if the person was Black, White, etc to help with identification. If the forensic anthropologist replies with a lecture on patterns of human variation, she/he will not be called on again for assistance (Kennedy 1995). In light of the poor allocation accuracies of various race determination methods (Ousley and Jantz 1998, Ginter 2001, Belcher et al 2002, Leathers et al 2002; Ginter 2002), a critical approach to the race concept is still essential for assessing the limitations of the race assessment methods. As Armelagos and Van Gerven (2003) suggest, a different approach to considering the race concept in forensic contexts may be more useful than new modifications to old methods.

The greater issue is the use of racial terminology or equivalents for investigating human variation. In the context of a discussion on race determination methods, questions can be raised regarding the concept of race. In the context of a discussion of sex, stature, or age determination methods, race-specific methods imply to students of anthropology, professional anthropologists and the greater public that some variation is explained by "race." The differences in sexual dimorphism between "Blacks" and "Whites" from the Terry Collection are not so great that Schulter-Ellis and colleagues (1983, 1985) should have devoted two separate papers to developing race-specific sex determination methods. In fact, the opposite is true. The sex determination method presented in Chapter 6, which in part, draws on some of the (non-racial) issues raised by these authors, is highly successful without considering "race." It is likely that there are limits to the applicability of all sex, age, and stature determination methods. Defining these limits on the basis of race when there is clear evidence that genetic and morphometric variation does not correspond with racial groups (Keita and Kittles 1997, Cartmill 1998, Templeton 1998, Armelagos and Goodman 1998, Lieberman 2001, Relethford 2002, Molnar 2002) only provides false parameters for the applicability of those methods. As illustrated in this thesis, the differences in size and sexual dimorphism are between the Terry and Coimbra Collections and not between the racial groups within the Terry Collection when age, year of birth are controlled, and circumstances that influenced the composition of the collection are considered. Rather than define the parameters of applicability, race-specific methods make forensic identification methods more difficult to apply. Students and professionals are left asking which race-specific method should be used if "race" is not known and cannot be determined, and not, why is it necessary to developing race-specific methods in the first place?

In addition to the forensic identification contexts, racial groups (or continental origin) continue to be used to investigate human variation. The review of the methodology for assessing secular change (Chapter 5) is consistent with Relethford's (2002) conclusions based on craniometric evidence and Armelagos and Goodman's (1998) conclusions based on historical evidence. Furthermore, this review strongly suggests that racial categories are not "coarse" but meaningful groups that may reveal at least a vague pattern of human variation. In some cases, matters may be much more complex than Relethford (2002) suggests when he concludes that skin colour is not a proxy for variation. Controlling for race does not control for genetic variation and, as illustrated in the review of the methodology for assessing secular change, the use of racial categories may obscure actual patterns of variation.

The racial terms used in various fields of research and in the greater society have changed throughout the 20th century. For example, Terry (1932:352-353), Giles (1964:129), and Ousley and Jantz (1996:19)[6] refer to samples that were used in their respective analyses that include individuals from the Terry Collection. The terminology changes (population, stock, continental origin) but the actual groupings and the

---

6 Direct quotes from these sources are presented on pages 44-46 and are not repeated here.

underlying concepts regarding these groupings remain the same: hybridization of non-"Whites" but never "Whites;" existence of original pure races; and "White" and "Black" genes or stock. Despite a clear departure from racial determinism in the post-World War II period, changes in terminology alone will not address any of the underlying issues if topological concepts of variation persist and "21st century technology is applied to 19th Century biology" (Alan Swedlund quoted in Aremalgos and Van Gerven 2003:60).

Some of the arguments that Washburn (1951, 1963) made about the persistence of an antiquated approach are echoed in a number of recent publications by Goodman (1995), and Armelagos and Van Gerven (2003). Despite some indications that there has been a change in how the race concept is applied (Armealgos and Van Gerven 2003, Lieberman et al 2003), problems persist because this concept continues to be used. The results presented in this thesis are consistent with the overwhelming evidence that has accumulated since the emergence of the new physical anthropology: human variation is real but it does not fit into three to five groups regardless of what terms are used to describe those categories. The substitution of various terms (ethnicity, ancestry, geographic/continental origin, or population) for the investigation of human variation, while maintaining the same assumptions (closed homogeneous isolated breeding group), does not solve any of the problems associated with typological approaches to research (Caspari 2003)[7].

---

[7] Gill (1990), a strong proponent of the race concept as biologically meaningful, argues against the use of terms such as "ethnic group" and argues that if the discussion is about the race concept, then the term race should be used.

Aside from the social and political issues (Armelagos and Goodman 1998. Blakey 1998), the persistence of racial categorizations may be due to the problems with establishing a simple, universal classification system with discrete, meaningful biological groups for the study of continuous human variation. Defining group boundaries is particularly difficult in complex industrialized societies with multiple socio-economic strata and overlapping cultures (Saunders and Herring 1995b). Though they may be referred to as ethnicity, ancestry, or regional origin, racial categorizations continue to be used because there are no easy replacements for the simple, yet biologically meaningless, racial approach.

LIMITATIONS OF CURRENT RESEARCH

There are limitations to some of the research presented in this thesis. Chapter 5 illustrates some potential problems when assessing secular change using skeletal data; it does not provide an adequate measure of the effects of secular change on the skeletal variation in the collections. Some of the results suggest that there are no significant changes in stature over time in any of the collections included in the analysis. However, the analysis presented in Chapter 5 was designed to test how secular changes are investigated using reference collections and not to assess secular changes in any one collection or the effects of secular change on sexual dimorphism. Further analysis using alternative parameters for cohorts and additional measurements is required to assess the magnitude and nature of secular changes in the Terry and Coimbra Collections.

More broadly, the secular change analysis illustrates some of the possible problems of representativeness of the reference collections when making inferences about the populations from which they were drawn. With few exceptions, reference collections have not been used for this purpose. These results from this study do not necessarily suggest that reference collections should never be used to make inferences about the living populations from which they were derived but that caution should be exercised.

There are also some limitations to the sex determination method presented in Chapter 6. First, assessment of the magnitude of inter-observer measurement error of the SPRL is required to complement the results of the intra-observer error tests presented. Second, at first glance, the computational skills required for applying a logistic regression model to actual data may seem complex. In actuality, logistic regression requires only a few more calculator button strokes than a discriminant function with the same number of independent variables. The benefit of logistic regression is that the probability of the allocation is calculated without secondary posterior probability and/or typicality statistics, which are required for a comparable approach with discriminant function analysis. A detailed step by step solution to one case is provided in Appendix B to illustrate the ease of determining sex using logistic regression. Still, why do any calculations at all when several morphological methods that have been used for decades provide good results? A metric method that uses logistic regression is superior to a morphological method because it is simple to calculate a level of certainty when assessing the sex of any one case. Furthermore, with a metric approach it is possible to include pelvic data and long bone data together in one method.

The third limitation relates to the choice of predictor/independent variables. The pelvis is widely recognized as the best source of information for assessing sex but it is easily damaged by taphonomical processes or during excavation. Complete or nearly complete specimens are required to take accurate measurements. In order to maximize the applicability of the methods, multiple models are presented, including a model that does not require any measurement of the pubis and a model that does not require measurements of either the pubis or the ischium.

CONCLUSION

Reflexivity and an awareness of researcher bias do not necessarily impede a research approach within the scientific method (Singer 1998, Lieberman 2001). As Lieberman notes, "knowledge can represent nature with increasingly greater accuracy when we are aware of our methodological errors, our accumulated knowledge, and the influence of our social and historical context" (2001:74). There are some benefits to considering the theoretical issues discussed in *Grave Reflections* and *Building a New Biocultural Synthesis*, and combined in the comprehensive model presented in this thesis. Walker (1995) shows conclusively that these issues must be considered in the most fundamental aspects of data collection in skeletal biological research on sexual dimorphism: what are real morphometric traits associated with each sex and how are our perceptions and interpretations of these traits influenced by our culturally constructed views of gender? Many other examples in *Grave Reflections* also clearly demonstrate the importance of many of these issues related to bias in skeletal collections.

Goodman and Leatherman (1998b) summarize the two-fold goal of a political economic approach to biocultural research as, 1) more thorough and rigorous research on variability by considering all types of factors, and, 2) on a political level, because of the nature of the topics of research, stimulating praxis in some form is essential. Armelagos and Goodman's (1998) review of the use of the race concept in physical anthropology has very obvious political implications. However, ignoring their conclusions can only lead to misleading results in the study of human variation. Assessing bias and levels of representativeness and reflexivity in the research questions, are issues related to the rigor of methodology and do not necessarily have to result in praxis as is advocated in *Building a New Biocultural Synthesis*. Walker's (1995) research will likely not bring an end to sexist stereotypes in North America, but considering some of the issues that he raises will certainly lead to a more rigorous methodology in the study of sexual dimorphism and the development of sex determination methods.

Several specific conclusions for each independent chapter and more general conclusions may be drawn. Based on the historical review and analysis in Chapter 3, it can be concluded that the source of the collection and the collection process have resulted in some major biases in the demographic structure of the Terry Collection and patterns of skeletal variation within the collection. Additionally, the quality and accuracy of the documentation of the collection, particulary regarding age, is very good. Using this documentary evidence and considering the historical context for amassing the collection, it is possible to assess and control for some of the biases. The Terry Collection continues to be an important resource for skeletal biological research because of Terry's and Trotter's efforts for curating skeletal material *and* documentary information.

There are several conclusions that can be drawn following a review of the methodology for assessing secular changes using skeletal data (Chapter 5). Pooling samples from different sources may produce very misleading patterns of secular change. Simply, significant changes in femur length correspond with a shift in the data source which indicates that the analysis is a comparison of different samples rather than a measure of secular change. Controlling for "race" in these samples does not control for differences between data sources, but rather, further obscures patterns of variation. Males and females must be considered separately when investigating secular change because variation in the sex ratio alone may result in misleading patterns of variation over time.

In Chapter 6, a series of logistic regression sex determination models are presented with allocation accuracies of 90-98%. This high allocation accuracy can be attributed to the combined use of pelvic and femur measurements, and the use of a reference sample drawn from the Terry and Coimbra Collections using year of birth and age at death data. An alternative pubis measurement, known as the superior pubis ramus length (SPRL) also contributes to the high accuracy by reducing misclassification due to measurement error. Measurement error greater than 2% in the measurement of the pubis may be a contributing factor to a mis-allocation of sex. The SPRL can be collected more reliably with less intra-observer error than the traditional measurement of the pubis. Racial categorization was irrelevant to the accuracy of the method. Tests of various models on large independent samples from the Terry Collection, the Coimbra Collection, and the FDB suggest that the method is widely applicable in archaeological and forensic cases.

There are several general conclusions that can be made with respect to the hypotheses stated in Chapter 1. First, racial categories, or equivalent terms such as continental origin, are methodological impediments for the study of humans. Others have demonstrated that racial groupings or continental origin are not proxies for genetic or morphometric variation. The results in this thesis confirm these results and go one step further: racial groups are not only poor parameters for describing the unit of analysis, they can obscure the actual pattern of variation being investigated, whether it is secular change or sexual dimorphism. Second, biases introduced in the collection process may result in the mis-interpretation of human variation when using samples from reference collections. Specifically age, year of birth and other sources of variation may be mis-interpreted as racial differences in the Terry Collection and the FDB if biases at different levels are not controlled. Third, using the comprehensive model proposed in this thesis, it is possible to identify and control for different levels of bias in the Terry Collection and other reference collections. This model represents the successful expansion of the cemetery studies theory to non-archaeological contexts and application of several NBS concepts to forensic anthropology with constructive, tangible results. By considering various levels and sources of bias, levels of representativeness can be assessed, and reference samples can be constructed from the reference collections to address numerous research questions. The series of models for sex determination presented in this thesis illustrates that the comprehensive model is useful for developing new forensic identification methods and for exploring alternative methodologies for developing those methods.

FUTURE RESEARCH

The Terry Collection and other reference collections are still very useful for addressing a number of different research questions including developing new forensic identification methods. Real, as opposed to virtual, reference collections have obvious advantages. The development and testing of the SPRL measurement, or similar research, could not have been possible with virtual collections. The reality is that human remains have to be re-buried, and all that is left in some cases are virtual collections. Virtual collections have a great deal of research potential but cannot replace real collections regardless of their age.

Several different studies are planned for the future to build on the current research and to address some of the limitations described above. Tests of inter-observer measurement error of the SPRL and other standard measurements are essential to confirm the results presented in Chapter 6 and to further assess how measurement error may affect the accuracy of metric approaches for sex determination. Additional tests of the sex determination methods presented in Chapter 6 using totally independent samples from the Grant Collection are also planned. The results from various preliminary analyses are very promising for developing additional sex determination methods using various combinations of measurements from the combined Coimbra-Terry reference sample (Albanese 2002). Lastly, the application of the comprehensive model to research involving the Grant Collection is planned. The goal is to test and further develop the entire comprehensive model.

The development of sex determination methods has been the focus of this thesis;

however, the issues discussed in the comprehensive model may be useful for other research involving reference collections. The main purpose of the comprehensive model is to construct useful reference samples through the assessment of bias in the reference collections. Thus, the model as a whole, or various aspects of it, can be used to construct reference samples for developing additional sex determination methods, as well as age and stature estimation methods. Additionally, the approach may also be applied in paleoanthropological investigations when fossil specimens from various time periods are compared to modern human variation.

APPENDIX A

A Brief History of Metric Sex Determination Methods Developed in The United States

---

Note: All footnotes are listed at the end of the table (pages 252-253).

Reference[1]	Bones	Collection	Sample Size[2]	"Race"[3]	Other data[4]	Accuracy by "Race"[5]	DF
Dwight 1894 (in Stewart 1979)	sternum scapula humerus femur	Harvard Medical School cadavers	123-228 (varies by bone; about equal numbers of WM and WF)	Y	N (data collected from anatomy school cadavers)	unknown	N
Dorsey 1897 (in Stewart 1979)	humerus femur	First Nations archaeological	125	Y	unknown	Y	N
Dwight 1905	humerus femur	Harvard Medical School cadavers	400 (200 WF, 200 WM)	Y	N (data collected from anatomy school cadavers)	Y	N
Letterman 1941	sciatic notch	Terry	426 (Just over 100 each of BF, BM, WM, WF)	Y	Y (Age: sample presented by age)++	Y	N
Washburn 1948	ischium pubis	Hamann-Todd	300 (50 BF, 50 BM, 100 WM, 100 WF)	Y	N	Y	N
Thieme and Schull 1957	femur humerus clavicle pubis ischium ilium	Terry	200 (101 BF, 99 BM)	Y	N	Y	Y**
Giles and Elliot 1963	cranium	Terry Hamann-Todd	300 (75 each of BM, BF, WM, WF)	Y	Y (age: 21-73 years)$	Y	Y

Reference[1]	Bones	Collection	Sample Size[2]	"Race"[3]	Other data[4]	Accuracy by "Race"[5]	DF
Giles 1964	mandible	Terry	265 (75 BM, 75 BF, 31 WM, 30 WF)	Y	Y (age: 21-73 years)$	Y	Y
Steele 1976	talus calcaneus	Terry	120 (30 each of BM, BF, WM, WF)	Y	Y (used criteria, including age, to select sample for stature research; sex determination was secondary)	Y	Y
Stewart 1979	scapula humerus sacrum hip bone femur	Terry	5-40 per sex and "race" group	Y*	N (exception is the sacrum; age considered)	Y*	N
DiBennardo and Taylor 1979	femur	Identified dissecting room skeletons, American Museum of Natural History, NY	85 (35 WF [all "white" females from collection] and 50 WM [randomly selected from collection])	Y	Y (Age: tested if age is a significant independent [predictor] variable)	Y	Y +
DiBennardo and Taylor 1982	femur	Terry	130 (65 BF, 65 BM)	Y	N	Y	Y
Schulter-Ellis et al 1983, 1985	hip bone femur	Terry	200 (50 each of BM, BF, WM, WF)	Y	N	Y	Y

Reference[1]	Bones	Collection	Sample Size[2]	"Race"[3]	Other data[4]	Accuracy by "Race"[5]	DF
France 1983	humerus	Terry various archaeological samples	746 (variably divided by source, sex and "race")	Y	N	Y	Y
Iscan and Miller-Shaivitz 1984	tibia	Terry	159 (40 each of BM, BF, WM, 39 WF)	Y	N	Y	Y
Holland 1986	cranial (base)	Terry	100 (25 each of BM, BF, WM, WF)	Y	Y (Age: limited sample to 20-50 years "to eliminate the age bias")+++	Y	N&
Holland 1991	tibia	Hamann-Todd	100 (50 F, 50 M)	Y	Y (Age: data presented)++	N	Y
Holman and Bennett 1991	humerus radius ulna	Terry	302 (about 75 each of BM, BF, WM, WF)	Y	N	Y	Y
Marino 1995	vertebra	Terry (develop method) Hamann-Todd (test method)	100 (50 each of BM, BF, WM, WF)	Y	N	Y (develops general and "race" specific equations)	Y

Reference[1]	Bones	Collection	Sample Size[2]	"Race"[3]	Other data[4]	Accuracy by "Race"[5]	DF
Ousley and Jantz 1996 (FORDISC 2.0)	Skull all major long bones pelvis talus calcaneus	FDB (includes large Terry sample)	sample varies greatly depending on procedure and variables selected (see Table 7.1)	Y	N	Y (DF discriminates between various racial groups and/or by sex)	Y
Smith 1997	foot	Terry Huntington	160 (40 each of BM, BF, WM, WF)	Y	Y (Age: sample based on availability of foot bones and age degeneration)	Y (DF discriminates between WM, BM, BF, and WF)	Y
Seidemann et al 1998	femur	Hamann-Todd	203 (about 50 each of BM, BF, WM, WF)	Y	Y (Age-effects assessed using ANCOVA)	Y	N

NOTES:
1.Chronological order by publication date.
2. If more than one bone or one measurement was used, sample size range is provided for all measurements.
3. "Race" used to define the sample or samples divided by "race."
4. Other data considered when sample was selected besides sex and "race."
5. Limits of applicability of method defined using racial terms.

DF = Discriminant function, BF = "Black" female, BM = "Black" male, WF = "White" females, WM = "White" male.

*Stewart selected the sample based on "race" and presents the data graphically by race for all measurements except the sacrum, but he also presented the results (sectioning points) for the femur and the humerus by sex only.

**Also used indices and absolute measurements.

$Age was used to set parameters of the sample, but there is no indication that any attempt was made to select a sample that was distributed relatively evenly over a range of ages.

&Holland (1986:205) used "multiple linear regression, a procedure functionally equivalent to simple linear discriminant-function".

+Used DF and other metric approaches and concluded that DF analyses were an unnecessary complication.

++Sample was arranged by age group or mean and range of age data presented; however, there is no indication that any attempt was made to select a sample that was distributed over the range of ages. While a mean may be a useful source of information in some contexts, providing a mean age at death or even a range for the sample does not provide any information about the sample. "...the watch averaged well, but nothing more. For half a day it would go like the very mischief, and keep up such a barking and wheezing and whooping and sneezing and snorting, that I could not hear myself think for the disturbance; and as long as it held out, there was not a watch in the land that stood any chance against it. But the rest of the day it would keep on slowing down and fooling along until all the clocks it had left behind caught up again. So at last, at the end of twenty-four hours, it would trot up to the judges' stand all right and just on time. It would show a fair and square average, and no man could say it had done more or less than its duty. But a correct average is only a mild virtue in a watch..." Mark Twain. December 1870.

+++Does not specify what is meant by "age bias."

APPENDIX B

Step-by-Step Calculation Using Model 1 (Chapter 6) to Determine Sex

The logistic model is an S-shaped function of the form

$$P = \frac{1}{(1 + e^{-Z})}$$

where P is the probability of the event (male or not male in this case) and Z is a linear combination of the independent variables such as,

$$Z = \beta_o + \beta_1 X_1 + \beta_2 X_2$$

Calculated probabilities are always between 0 and 1. For example, if P = 0.84, the individual is classified as male since P is greater than 0.5 and the estimated probability is 84% that this combination of measurements would be found in a male. Alternatively, there is only a 16% probability (1-P) that this combination of measurements would be found in a female.

Sample data:

hipbone height (HBH) = 204

iliac breadth (IB) = 151

superior pubis ramus length (SPRL) = 66

maximum diameter femur head (MHD) = 42

maximum epicondylar breadth of femur (MEB) = 75

(see Chapter 6 for descriptions of measurements)

Using Model 1,

$$P = \frac{1}{(1 + e^{-(0.5950\,(HBH) - 0.5192\,(IB) - 1.1104\,(SPRL) + 1.1696\,(MHD) + 0.5893\,(MEB) - 61.5345)})}$$

$$P = \frac{1}{(1 + e^{-(0.5950(204) - 0.5192(151) - 1.1104\,(66) + 1.1696(42) + 0.5893(75) - 61.3545)})}$$

First calculate Z as with a linear function or discriminant function equation:

$Z = 0.5950(204) - 0.5192(151) - 1.1104\,(66) + 1.1696(42) + 0.5893(75) - 61.3545$

$Z = 1.4806$

Now calculate P:

$$P = \frac{1}{(1 + e^{-Z})}$$

$$P = \frac{1}{(1 + e^{-1.4806})}$$

e is the mathematical constant which is available on any scientific calculator. On scientific calculators it appears as a button such as "e^{x}".

e raised to power of - 1.4806 = 0.227501

$$P = \frac{1}{(1 + 0.227501)}$$

P = 0.8147

Since P = 0.8147, this person is male and there is an 81.47% probability that this combination of measurements would be found in a male. Despite the small size, the documentation for this person is clear and this person is male (see Chapter 6 for details).

REFERENCES CITED

Albanese J. 1997a. A comparison of the Terry Collection and the Grant Collection using the head of the femur and head of the humerus: implications for determining sex. Can Soc Forens Sci J 30:167.

Albanese J. 1997b. Similarities and differences between the Terry Collection and the Grant Collection: the implications of collection and sample selection when developing sex determination methods. Paper presented at the 25th Annual Meeting of the Canadian Association for Physical Anthropology, London, Ontario.

Albanese J. 1997c. Skeletal variability in recent North Americans: implication for the development of new sex determination methods. Paper presented at the 25th Annual Meeting of the Canadian Association for Physical Anthropology, London, Ontario.

Albanese J. 2000. Potential methodological problems when combining skeletal data from different sources for the study of secular change. Paper presented at the 28th Annual Meeting of the Canadian Association for Physical Anthropology, Burlington, Ontario.

Albanese J. 2002. A new metric sex determination method using the clavicle and humerus. Paper presented at the 30th Annual Meeting of the Canadian Association for Physical Anthropology in Ottawa, Ontario, October 24-26.

Angel L. 1976. Colonial to modern skeletal change in the U.S.A. Am J Phys Anthropol 45:723-736.

Armelagos GJ, Van Gerven DP. 1982. The theoretical foundations and development of skeletal biology. In Spencer F, editor. A History of American Physical Anthropology 1930-1980. New York: Academic Press, p 305-328.

Armelagos GJ, Goodman AH. 1998. Race, racism, and anthropology. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p. 359-378.

Armelagos GJ, Van Gerven DP. 2003. A century of skeletal biology and paleopathology: contrasts, contradictions, and conflicts. Am Anthropol 105:53-64.

Asala SA. 2001. Sex Determination from the head of the femur of South African whites and blacks. Forensic Sci Int 117:15-22.

Belcher R, Williams F, Armelagos GJ. 2002. Misidentification of Meroitic Nubians using FORDISC 2.0. Am J Phys Anthropol Suppl 34:42.

Blakey ML. 1987. Skull Doctors: Intrinsic social and political bias in the history of American physical anthropology with special reference to the work of Aleš Hrdlička Crit Anthropol 7: 7-35.

Blakey ML. 1998. Beyond European enlightenment: Towards a critical and humanistic human biology. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p 379-406.

Boas F. 1912. Changes in bodily form of descendants of immigrants. Am Anthropol 14:530-562.

Boas F. 1940. Race, Language and Culture. New York: Free Press.

Boulinier G. 1969. Variation avec l'âge du dimorphism sexuel des crânes humains adultes: influence sur les possibilités de discrimination statistique des sujets. Bull et Mém de la Société d'Anthropologie de Paris, 12th series, 3:301-316.

Brace CL. 1982. The roots of the race concept in American physical anthropology. In Spencer F, editor. A History of American Physical Anthropology 1930-1980. New York: Academic Press, p 11-29.

Breeden JO. 1975. Body snatchers and anatomy professors: Medical education in 19th Century Virginia. Va Mag Hist Biogr 83: 321-345.

Brooks ST. 1955. Skeletal age at death: the reliability of cranial and pubic age indicators. Am J Phys Anthropol 13:567-597.

Brooks ST, Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks Method. Hum Evol 5:227-238.

Brown RA, Armelagos GJ. 2001. Apportionment of racial diversity: A review. Evol Anthropol 10:34-40.

Bruzek J. 1992. Fiabilité des fonctions discrimininantes dans la détermination sexuelle de l'os coxal critiques et propositions. Bull et Mém de la Société d'Anthropologie de Paris 4:67-104.

Buikstra JE, King JL, Nystrom KC. 2003. Forensic anthropology and bioarchaeology in the American Anthropologist: Rare but exquisite gems. Am Anthropol 105:38-52.

Cannon A. 1995. Material culture and burial representativeness. In: Saunders SR, Herring A, editors. Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p 3-17.

Cardoso, HFV. 2000. Dimorfismo Sexual na Estatura, Dimensões e Proporções dos Ossos Longos dos Membros. Master's thesis. University of Coimbra.

Cartmill M. 1998. The status of the race concept in physical anthropology. Am Anthropol 100: 651-660.

Cartmill M, Brown K. 2003. Survey the race concept: A reply to Lieberman, Kirk, and Littlefield. Am Anthropol 105:114-115.

Caspari R. 2003. From types to populations: A century of race, physical anthropology, and the American Anthropological Association. Am Anthropol 105:65-76.

Cobb. WM. 1930. Human materials in American institutions available for anthropological study. Supplement to Am J Phys Anthropol 17.

Cobb. WM. 1936. Race and runners. J Health Phys Educ 7:1-9.

Cobb. WM. 1939. The Negro as a biological element in the American population. J Negro Educ 8:336-348.

Cobb WM. 1952. Skeleton. In Lansing AI, editor. Cowdry's Problem of Aging, Third Edition. Baltimore: Williams and Wilkins, p. 791-856.

Cobb WM. 1959. Thomas Wingate Todd, M.B., Ch. B. ERCS (Eng), 1855-1938. J Natl Med Assoc 51:233-246.

Cobb WM. 1981. Thomas Wingate Todd, 1885-1938. Am J Phys Anthropol 56:517-520.

Coon CS. 1962. The Origin of Races. New York: Afred A Knof.

Cunha E. 1995. Testing identification records: evidence from the Coimbra Identified Skeletal Collections (nineteenth and twentieth centuries). In: Saunders SR, Herring A, editors. Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 179-198.

Dibennardo R, Taylor JV. 1979. Sex assessment of the femur: A test of a new method. Am J Phys Anthropol 50:635-638.

Dibennardo R, Taylor JV. 1982. Classification and misclassification in sexing the black femur discriminant function analysis. Am J Phys Anthropol 58:145-151.

Dorsey GA. 1897. A sexual study of the size of the articular surfaces of the long bones in aboriginal skeletons. Boston Med Surg J 137:80-82.

Dwight T. 1894. The range and significance of variation in the human skeleton [Shattuck Lecture]. Boston Med Surg J 131:73-76, 97-101.

Dwight T. 1905. The size of the articular surfaces of the long bones as characteristic of sex; an anthropological study. Am J Anat 4:19-32.

Ericksen MF. 1982. How "representative" is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 59:345-350.

Fledesman MR, Fountain RL. 1996. "Race" specificity and the femur/stature ratio. Am J Phys Anthropol 100:207-224.

Fish, JM. 1999. Why psychologists should learn some anthropology. In: Sussman, RW, editor. The Biological Basis of Human Behavior: A Critical review. New Jersey: Prentice Hall. p 197-198.

France DL. 1983. Sexual Dimorphism in the Human Humerus. Ph. D. dissertation, Department of Anthropology, University of Colorado.

Gilbert BM, McKern TW. 1973. A method for aging the female *Os pubis*. Am J Phys Anthropol 38:31-38.

Giles E. 1964. Sex determination by discriminant function analysis of the mandible. Am J Phys Anthropol 22:129-136.

Giles E. 1970a. Sexing crania by discriminant function analysis: Effects of age and number of variables. Proceedings, 13th International Congress of Anthropological and Ethnological Sciences, Tokyo, 1:59-61.

Giles E. 1970b. Discriminant function sexing of the human skeleton. In Stewart TD, editor. Personal Identification in Mass Disasters. Washington, DC: National Museum of Natural History. p. 99-109.

Giles E, Elliot O. 1962. Race identification from cranial measurements. J Forensic Sci 7:247-257.

Giles E, Elliot O. 1963. Sex determination by discriminant function analysis of crania. Am J Phys Anthropol 21:53-68.

Gill GW. 1990. Introduction. In Gill GW, Rhine S, editors. Skeletal Attribution of Race. Maxwell Museum of Anthropology, Anthropological Papers No. 4, p vii-xii.

Ginter JK. 2001. Dealing with Unknowns in a Non-Population: The Skeletal Analysis of the Odd Fellow Series. Masters Thesis. London: University of Western Ontario.

Ginter JK. 2002. Issues of skeletal identification: ancestry determination of the Odd Fellows Series. Can Soc Forensic Sci J 35:41.

Goodman AH. 1995. The problematics of "race" in contemporary biological anthropology. In Boaz NT, Wolfe LD, editors. Biological Anthropology: The State of the Science. Bend: International Institute for Human Evolutionary Research.

Goodman AH. 2000. Why genes don't count (for racial differences in health). Am J Public Health. 90:1699-1702.

Goodman H, Leatherman TL. 1998a. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press.

Goodman H, Leatherman TL. 1998b. Traversing the chasm between biology and culture: An introduction. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p. 3-41.

Gravlee CC, Bernard HR, Leonard WR. 2003. Heredity, environment and cranial form: A reanalysis of Boas's immigrant data. Am Anthropol 105:125-138.

Harrington JM, Blakely RL. 1995. Rich man, poor man, beggar man, thief: Selectivity exercised by grave robbers in Georgia. In, Saunders SR, Herring A, editors, Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 153-178.

Harrington JM and Blakely RL. 1997. Bones in the Basement: Postmortem Racism in Nineteenth-Century Medical Training. Washington: Smithsonian Institution Press.

Hartmann J, Keister R, Houlihan B, Thompson L, Baldwin R. 1994. Diversity of ethnic and racial VNTR RFLP fixed-bin frequency distributions. Am J Hum Genet 55:1268-1278.

Hennessy RJ, Stringer CB. 2002. Geometric Morphometric Study of the Regional Variation of Modern Human Craniofacial Form. Am J Phys Anthropol 117:37-48

Holland TD. 1986. Sex determination of fragmentary crania by analysis of cranial base. Am J Phys Anthropol 70:203-208.

Holland TD. 1991. Sex assessment using the proximal tibia. Am J Phys Anthropol 85:221-227.

Holman DJ, Bennett KA. 1991. Determination of sex from arm bone measurements. Am J Phys Anthropol 84:421-426.

Holloway RL. 2002. Head to head with Boas: Did he err on the plasticity of head form? Proc Natl Acad Sci 99:14622-14623.

Hoppa RD. 1996. Representativeness and Bias in Cemetery Samples: Implications for Palaeodemographic Reconstruction of Past Populations. Ph. D. Thesis. McMaster University.

Howells WW. 1973. Cranial Variation in Man. Cambridge: Papers of the Peabody Museum, Volume 67.

Howells WW. 1989. Skull Shapes and the Map. Cambridge: Papers of the Peabody Museum, Volume 78.

Hrdlička A. 1918. Physical Anthropology: Its scopes and aims: Its history and present state in America. Am J Phys Anthropol 1:3-23.

Hrdlička A. 1925a. Relation of the size of the head and skull to capacity in the two sexes. Am J Phys Anthropol 8:249-250.

Hrdlička A. 1925b. The Old Americans. Baltimore: Williams and Wilkins.

Hrdlička A. 1927. Anthropology of the American Negro: Historical notes. Am J Phys Anthropol 10:205-235.

Hrdlička A. 1941. Observations and measurements on the members of the National Academy of Sciences. Mem Natl Acad Sci 23:1-108.

Humphrey DC. 1973. Dissection and discrimination: The social origins of cadavers in America, 1760-1915. Bull N Y Acad Med 49:819-827.

İşcan MY, Cotton TS. 1990. Osteometric assessment of racial affinity from multiple sites in the postcranial skeleton. In Gill GW, Rhine JS, editors. Skeletal Attribution of Race. Albuquerque: Maxwell Museum of Anthropology, p. 83-90.

İşcan MY, Loth SR, and Wright RK. 1984. Age estimation from the rib by phase analysis: White males. J Forensic Sci 29:1094-1104.

İşcan MY, Loth SR, and Wright RK. 1985. Age estimation from the rib by phase analysis: White females. J Forensic Sci 30:853-863.

İşcan MY, Loth SR, and Wright RK. 1987. Racial variation in the sternal extremity of the rib and its effect on age estimation. J Forensic Sci 32:452-466.

İşcan MY, Miller-Shaivitz P. 1984. Discriminant function sexing of the tibia. J Forensic Sci 29:1087-1093.

Jantz RL, Moore-Jansen PH. 1988. A Database for Forensic Anthropology: Structure, Content and Analysis. Report of Investigation No. 47, Department of Anthropology, University of Tennessee, Knoxville.

Jones-Kern K and Latimer B. 1996. Skeletons out of the closet. Explorer 38.

Keita SOY, Kittles RA. 1997. The persistence of racial thinking and the myth of racial divergence. Am Anthropol 99:534-544.

Kennedy KAR. 1995. But professor, why teach race identification if races don't exist? J Forensic Sci 40:97-800

Krogman WM, İşcan MY. 1986. The Human Skeleton in Forensic Medicine. Second Edition. Springfield: CC Thomas.

Lassek AM. 1958. Human Dissection: Its Drama and its Struggle. Springfield: C. C. Thomas.

Letterman GS. 1941. The greater sciatic notch in American Whites and Negroes. Am J Phys Anthropol 28:99-116.

Leathers A, Edwards J, Armelagos GJ. 2002. Assessment of classification of crania using FORDISC 2.0: Nubian X-Group test. Am J Phys Anthropol Suppl 34:99-100.

Levins R, Lewontin R. 1998. Forward. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p xi-xv.

Lewontin R. 1972. The apportionment of human diversity. Evol Biol 6:381-398.

Lieberman L. 2001. How "Caucasoids" got such big crania and why they shrank. Curr Anthropol 42:69-95.

Lieberman L, Kirk RC, Littlefield, A. 2003. Perishing paradigm: Race 1931-99. Am Anthropol 105:110-113.

Livingstone F. 1962. On the non-existence of human races. Curr Anthrop 3:279-181.

Lovejoy CO, Meindl RS, Mensforth RP, and Barton TJ. 1985a. Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. Am J Phys Anthropol 68:1-14.

Lovejoy CO, Meindl RS, Pryzbeck TR and Mensforth RP. 1985b. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. Am J Phys Anthropol 68:15-28.

Marino EA. 1995. Sex estimation using the first cervical vertebra. Am J Phys Anthropol 97:127-133.

Matthews W, Billings JS. 1891. The human bones of the Hemenway Collection in the United States Army Medical Museum in Washington. Mem Natl Acad Sci 6:139-286.

McKern TW, Stewart TD. 1957. Skeletal Age Changes in Young American Males. Analysis from the Standpoint of Age Identification. Environmental Protection Research and Development Division (Quartermaster Research and Development Center, U.S. Army, Natick, Massachusetts), Technical report EP-45.

Meadows Jantz L, Jantz RL. 1999. Secular change in long bone length and proportion in the United States, 1800-1970. Am J Phys Anthropol 110:57-67.

Meindl RS, Lovejoy CO, Mensforth RP, and Walker RA. 1985a. A revised method of age determination using the os pubis, with a review and tet of accuracy of other current methods of pubic symphyseal age. Am J Phys Anthropol 68:29-45.

Meindl RS, Lovejoy CO, Mensforth RP, Carlos LD. 1985b. Accuracy and direction of error in the sexing of the skeleton: Implications for Paleodemography. Am J Phys Anthropol 68:79-85.

Meindl RS, Russel KF, Lovejoy CO. 1990. Reliability of age at death in the Hamann-Todd Collection: Validity of subselection procedures used in blind tests of the summary age technique. Am J Phys Anthropol 83:349-357.

Molleson T. Rates of aging in the eighteenth century. In, Saunders SR, Herring A, editors, Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 199-222.

Molnar S. 2002. Human Variation: Races, Types, and Ethnic Groups. Fifth Edition. Upper Saddle River, N.J.: Prentice Hall.

Montagu A. 1964. The Concept of Race. London: Collier-McMillian.

Montagu A. 1974. Man's Most Dangerous Myth: The Fallacy of Race. Fifth Edition. London: Oxford University Press.

Moore-Jansen PM, Ousley SD, Jantz RL. 1994. Data Collection Procedures for Forensic Skeletal Material. Report of Investigation no. 48. The University of Tennessee, Knoxville, Department of Anthropology.

Nei M, Roychoudhury Ak. 1982. Evolutionary relationships and evolution of human races. Evol Biol 14:1-59.

Norušis MJ. 1990. SPSS Advanced Statistics Student Guide. Chicago: SPSS Inc.

Ousley SD, Jantz RL. 1996. FORDISC 2.0: Personal Computer Forensic Discriminant Functions. Knoxville: University of Tennessee.

Ousley SD, Jantz RL. 1998. The forensic data bank: documenting skeletal trends in the United States. In Reichs KJ, editor, Forensic Anthropology. Second Edition. Springfield: CC Thomas, p. 441-458.

Owsley SD, Jantz RL. 2001. Archaeological politics and public interest in paleoamerican studies: Lessons from Gordon Creek and Kennewick Man. Am Antiquity 66:565-575.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297-302.

Pons J. The sexual diagnosis of isolated bones on the skeleton. Hum Biol 27:12-21.

Rankin-Hill LM, Blakey ML. 1994. W. Montague Cobb (1904-1990): Physical anthropologist, anatomist, and activist. Am Anthropol 96:74-96.

Relethford JH. 1994. Craniometric variation among modern human populations. Am J Phys Anthropol 95:53-62.

Relethford JH. 2001. Global analysis of regional differences in craniometric diversity and population substructure. Hum Biol 73:629-636.

Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. Am J Phys Anthropol 118:393-398.

Richardson R. 1987. Death, Dissection, and the Destitute. New York: Routledge & Kegan Paul.

Rocha A. 1995. Les collections ostéologiques humaines identifiées du Musée Anthropologigue de l'Université de Coimbra. Anthropol Port 13:7-38.

Roseberry W. 1998. Political-economy and social fields. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p. 75-91.

Ryman N, Chakraborty R, Nei M. 1983. Differences in the relative distribution of human genetic diversity between electrophoretic and red and white cell antigens. Hum Heredity 33:93-102.

St. Hoyme LE. 1957. The earliest use of indices for sexing pelves. Am J Phys Anthropol 15:537-546.

St. Hoyme LE, İşcan MY. 1989. Determination of sex and race: accuracy and assumptions. In İşcan MY, Kennedy KA, editors, Reconstruction of Life from the Skeleton, p. 53-93.

Santos AL. 2000. A Skeletal Picture of Tuberculosis. Doctoral thesis. University of Coimbra.

Saunders SR, Herring A. 1995a. Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press.

Saunders SR, Herring A. 1995b. Preface. In: Saunders SR, Herring A, editors. Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 1-2.

Saunders SR, Herring A, Sawchuk LA, Boyce G. 1995. The nineteenth century cemetery at St. Thomas' Anglican Church, Belleville: Skeletal remains, parish records, and censuses. In, Saunders SR, Herring A, editors, Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 93-117.

Saunders SR, Hoppa RD. 1997. Sex allocation from long bone measurements using logistic regression. Can Soc Forens Sci J 30:49-60.

Schell LM. 1997. Culture as a stressor: A revised model of biocultural interaction. Am J Phys Anthropol 15 102:67-77.

Schell LM, Czerwinski SA. 1998. Environmental health, social inequality and biological differences. In Human Biology and Social Inequality, S. S. Strickjland and P. S. Shetty (Eds). Cambridge: Cambridge University Press, p. 114-131.

Scheuer JL, Bowman JE. 1995. Correlation of documentary and skeletal evidence in the St. Bride's Crypt population. In, Saunders SR, Herring A, editors, Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 49-70.

Schulter-Ellis F P, Schmidt DJ, Hayek LA and Craig J. 1983. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I. J Forensic Sci 28:169-80.

Schulter-Ellis FP, Hayek LA and Schmidt DJ. 1985. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part II. J Forensic Sci 30:178-85.

Schultz AH. 1930. The skeleton of the trunk and limbs of higher primates. Am J Phys Anthropol 2:303-438.

Seidemann RM, Stojanowski CM, and Doran GH. 1998. The use of the superior-inferior femoral neck diameter as a sex assessor. Am J Phys Anthropol 107:305-313.

Shapiro HL. 1939. Thomas Wingate Todd. Am Anthropol 41:458-464.

Singer M. 1998. The development of a critical medical anthropology: implications for biological anthropology. In: Goodman AH, Leatherman TL, editors. Building a New Biocultural Synthesis: Political-economic Perspectives on Human Biology. Ann Arbor: University of Michigan Press, p. 93-123.

Smith SL. 1997. Attribution of foot bones to sex and population groups. J Forensic Sci 42:186-95.

Sparks CS, Jantz RL. 2002. A reassessment of human cranial plasticity: Boas revisited. Proc Natl Acad Sci 99:14636-14639.

Spencer F. 1981. The rise of academic physical anthropology in the United States (1880-1980): A historical overview. Am J Phys Anthropol 56:353-364.

Spencer F. 1997a. American Association of Physical Anthropologists. In Spencer F, editor, History of Physical Anthropology. An Encyclopedia Vol 1. New York: Garland Publishing, p. 60-63.

Spencer F. 1997b. Aleš Hrdlička (1869-1943). In Spencer F, editor, History of Physical Anthropology. An Encyclopedia Vol 1. New York: Garland Publishing, p. 503-505.

Steele DG. 1976. The estimation of sex on the basis of the talus and calcaneus. Am J Phys Anthropol 45:581-588.

Stewart TD. 1979. Essentials of Forensic Anthropology. Springfield: CC Thomas.

Stewart TD. 1981. Aleš Hrdlička, 1869-1943. Am J Phys Anthropol 56:347-351.

Templeton AR. 1998. Human races. A genetic and evolutionary perspective. Am Anthropol 100:632-650.

Terry RJ. 1932. The clavicle of the American Negro. Am J Phys Anthropol 3:351-379.

Terry, RJ. 1940. On measuring and photographing the cadaver. Am J Phys Anthropol 26:433-447.

Thieme FP, Schull WJ. 1957. Sex determination from the skeleton. Hum Biol 29:242-273.

Todd TW. 1920. Age changes in the pubic bone: I. The male White pubis. Am J Phys Anthropol 3:285-334.

Todd TW. 1921. 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.

Todd TW, Lyon DW. 1924. I Endocranial suture closure in the adult males of White stock. Am J Phys Anthropol 7:325-384.

Todd TW, Lyon DW. 1925a. II Ectocranial suture closure in the adult males of White stock. Am J Phys Anthropol 8:23-45.

Todd TW, Lyon DW. 1925b. Endocranial suture closure in the adult males of Negro stock. Am J Phys Anthropol 8:47-71.

Todd TW, Lyon DW. 1925c. Ectocranial suture closure in the adult males of Negro stock. Am J Phys Anthropol 8:149-168.

Trotter M. 1981. Robert James Terry, 1871-1966. Am J Phys Anthropol 56:503-508.

Trotter M, Gleser GC. 1951. Trends in stature of American whites and Negroes born between 1840 and 1924. Am J Phys Anthropol 9:427-440.

Turner W. 1885. The index of the pelvic brim as a basis of classification J Anat And Physiol 20:125-143.

Ubelaker DH. 1997. Thomas Dale Stewart (1901-). In Spencer F, editor, History of Physical Anthropology. An Encyclopedia Vol 2. New York: Garland Publishing, p. 999-1001.

Walker PL. 1995. Problems of preservation and sexism in sexing: Some lessons from historical collections and palaeodemographers. In, Saunders SR, Herring A, editors, Grave Reflections: Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, p. 31-47.

Washburn SL. 1948. Sex differences in the pubic bone. Am J Phys Anthropol 6:199-207.

Washburn SL. 1951. The new physical anthropology. Trans N Y Acad Sci (Series II) 13:298-304.

Washburn SL. 1963. The study of race. Am Anthropol 65:521-531.

Wilf SR. 1989. Anatomy and punishment in late eighteenth-century New York. J Soc His 22:507-30.