SKELETAL SEX DETERMINATION AND AGE ESTIMATION

SEX DETERMINATION AND AGE ESTIMATION: SKELETAL EVIDENCE FROM ST. THOMAS' CEMETERY BELLEVILLE, ONTARIO

BY

TRACY L. ROGERS, B.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

November 1991

MASTER OF ARTS (1991)
(Anthropology)

McMaster University
Hamilton, Ontario

TITLE: SEX DETERMINATION AND AGE ESTIMATION: SKELETAL EVIDENCE FROM ST. THOMAS' CEMETERY BELLEVILLE, ONTARIO

AUTHOR: Tracy L. Rogers

SUPERVISOR: Professor Shelley R. Saunders

NUMBER OF PAGES: xii, 184

ABSTRACT

The precision and accuracy of 17 pelvic and 17 cranial features commonly used to assess adult skeletal sex and three systems of adult pelvic age estimation were tested on a sample of 278 individuals from the 19th century St. Thomas' Anglican Church Cemetery in Belleville, Ontario. Documentation in the form of parish burial, marriage and baptismal records exists for the entire cemetery, but only fifty-five individuals over 17 years of age are personally identified.

Both the pelvic and cranial methods of morphological sex determination produce sex ratios indistinguishable from the documented ratio. On an individual basis, the precision and accuracy of the pelvic approach is superior to that of the cranial approach. The most reliable and accurate pelvic results may be obtained from an examination of the following six criteria: ventral arc, obturator foramen, true pelvis shape, sacrum shape, subpubic concavity and pubis shape. The following eight cranial criteria produced the least amount of intraobserver error and the greatest degrees of accuracy: supraorbital ridges, mastoid size, malar size and shape, occipital markings, chin form, general size, zygomatic root extension and mandible shape.

Three features, posterior view of the sacrum, zygomatic root extension and occipital muscle markings, exhibit increased accuracy with age. There is no age-related affect on accuracy when the complete trait lists are employed. Pelvic assesments appear to be biased in favour of

males while cranial assessments seem biased in favour of females.

Of the Phenice (1969) criteria, the ventral arc is the single most useful trait exhibiting both high precision (0% intraobserver error) and high accuracy (86.9%). The subpubic concavity ranked second and the ischiopubic ramus last.

None of the age techniques examined in this investigation are precise and accurate enough to be employed exclusive of other systems. When both precision and accuracy are taken into account, the auricular surface technique of estimating skeletal age-at-death produces the best results for individual cases. One major problem with this method is its tendency to overage individuals less than 40 years and underage individuals greater than 40 years.

The accuracies of the Suchey methods are acceptable mainly because of the wide age ranges covered by each phase. The estimates provided by these methods are too broad to be useful in individual cases. When used to produce an age profile the Suchey systems (with the Jackes [1985] distribution method) are better than the auricular surface method. The pattern is slightly distorted by problems associated with phases III (2), IV (1) and IV.

ACKNOWLEDGEMENTS

If 250 adult skeletons are to be moved, examined for sex-related features, moved again, examined for age-related features, photographed, moved yet again, and the relevant surfaces must be cast, all within four months, and, if the results are to be organized in a readable manner, there will be many people to thank.

First I would like to express my appreciation to my thesis committee. I am grateful to Dr. Shelley R. Saunders, my thesis supervisor, for involving me in the St. Thomas' project, and I thank her for her interest, her support and for "knowing everything" and being willing to share the knowledge. Thank you Dr. Ann Herring for all of your help with computers and statistics, for your insightful comments and your willingness to answer "just one more quick question". I also wish to express my gratitude to Dr. Susan Pfeiffer for her (and consequently my) dedication to the scientific method, her valuable suggestions and for her ever-present enthusiasm for my project.

I am indebted to the parishioners and council members of St. Thomas' Anglican Church in Belleville, Ontario for deciding to have the cemetery archaeolgically excavated and the material examined. Similarly, I am grateful to Dr. Heather McKillop, the site supervisor, and Archaeological Services Incorporated for allowing me the opportunity to study the material.

This thesis could not have been completed without the assistance of a number of specialists. Gerry Boyce, the site historian, provided the invaluable list of identified individuals and was always pleased to help me with questions concerning the parish records. Dorothy Kealey and Laurel Parson at the General Synod Archives, Anglican Archives at Church House in Toronto were extremely helpful during my long visits with the microfilm reading machine. Dr. Peter MacDonald from the Mathematics and Statistics Department at McMaster University made statistics much easier to deal with.

Throughout the entire "thesis process" there have been many people who have given me their advice, their help, their cars and their computers and I would like thank each one.

From helping me carry boxes to renumbering pelves to making suggestions to loaning me his computer, Colin Varley has been a great friend and colleague. I want to officially thank him for everything he has done to help make the process easier and more enjoyable.

Always willing to manipulate files, give an opinion or whip up yet another graph, Carol De Vito has been enormously helpful and as a result, has my sincere gratitude.

Thank you Margit and Pop for your support (both emotional and financial) and the use of your car.

For refraining from asking me, "are you done yet?" and for reading the rough drafts and enthusiastically congratulating me for eighty-word sentences, I want to thank David Lane.

During the actual research period, I was able to exploit most of my friends and many of my fellow students. I want them to know how much they were appreciated. Thank you Andrea Lagan for taking photos, making casts and for all the other little things you were so willing to help me with. Thank you Karen for doing just about everything including moving skulls up and down the stairs. Thank you Dave Warnock, Gary Jessup and Kathryn Denning for filling up your spare time by finishing my casting. Thank you Shannon Rogers for saving me from the often mind-numbing task of entering data. Thank you Paul Rauscher for helping renumber my pelves.

Finally, I want to thank my friends and colleagues who also worked on the Belleville project, Anne Keenlyside, Sue Jimenez, and Becky Southern and my office mate Judy Clark for their willingness to give a much needed opinion or to listen to confused rambling and help me organize my thoughts. Thanks are also due to all of the people in the Anthropology Department at McMaster for their support and interest.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

PAGE

Tables

PAGE

Figures

CHAPTER ONE
1.0 INTRODUCTION

A fundamental part of any skeletal analysis, whether archaeological or forensic, is the estimation of age and determination of the sex of the individual or individuals in question. Numerous methods for obtaining these data currently exist and revisions to established techniques as well as the introduction of new procedures are commonplace in the anthropological literature (for sex determination see: Acsádi and Nemeskeri, 1970; Giles and Elliot, 1963; Keen, 1950; Kelley, 1979; Phenice, 1969; Richman et al., 1979; St. Hoyme, 1984; Singh and Potturi, 1978; Stewart, 1952; Washburn, 1948; Workshop of European Anthropologists, 1980; for age estimation see: Acsádi and Nemeskeri, 1970; Gilbert and McKern, 1972; Jackes, 1985; Katz and Suchey, 1989; Lovejoy, Meindl, Mensforth and Barton, 1985; Lovejoy, Meindl, Pryzbeck, and Mensforth, 1985; Stewart, 1957; Suchey, Brooks and Katz, 1988; Todd, 1920; Workshop of European Anthropologists, 1980). Equally important are studies which focus upon the accuracy, bias, subjectivity and consequently, the applicability of these improvements and new approaches to samples outside the reference population (Anderson, 1990; Bedford et al., 1989; Brooks and Suchey, 1990; Budinoff and Tague, 1990; Gilbert, 1972; Henke, 1977; Kajanoja, 1966; Kelley, 1978; Lanphear, 1989; Lovell, 1989; MacLaughlin and Bruce, 1990; Meindl, Lovejoy, Mensforth and Carlos, 1985; Murray and Murray, 1991; Sutherland and Suchey, 1991, 1987; Weiss, 1972). Unfortunately, research of this type requires large skeletal samples containing individuals of known sex and age. Such samples are rare.

In 1989, the consulting firm Northeastern Archaeological Associates, under the direction of Dr. Heather McKillop, was employed by St. Thomas' Anglican Church in Belleville, Ontario, to relocate the obsolete cemetery on the church premises. The number of burials was estimated to be somewhere between 60 and 80. Thus, Northeastern undertook what was expected to be a relatively small operation. Instead, they became involved in one of the largest historic period cemetery excavations in North America. In total, 579 graves were removed (McKillop et al., 1989).

Although most of the tombstones at the cemetery had been relocated or damaged over the past one hundred years, it was still possible to identify individuals in approximately 15% of the excavated burials due to legible coffin nameplates. It was then possible to link the nameplates to the St. Thomas' parish registers which contain records, kept by each rector, of the number of interments in the cemetery. The result of this process is a subsample of 55 adult individuals of known sex and/or age.

Dr. Shelley Saunders of McMaster University was contacted by Northeastern and was asked to conduct an analysis of the skeletal material. Under her supervision, various graduate students and researchers from McMaster and other universities worked together to recover as much information about the skeletal sample as was possible during the one year period in which the material remained available for study before reburial. I was asked to conduct sex determinations and age estimations through morphological examinations of all of the adult skeletons.

"Adult" was defined as individuals 17 years and older and adult status was skeletally determined through an examination of epiphyseal closure. In order to be classified adult, the ischium, ilium and pubis had to be completely fused and fusion of the epiphyses of the femur, tibia and fibula must have been initiated (Krogman and İşcan, 1986). Analysis of the adult material was conducted separately from subadult analysis because methods which apply to one frequently do not apply to the other.

The purpose of this investigation was to determine whether sex estimations obtained through morphological examinations of the pelvis and skull are significantly different from the actual sex of the individuals as recorded in the parish records. In addition, this analysis examined whether there is a significant difference between three methods used to estimate adult age at death and the actual documented age at death.

The techniques under consideration include the Suchey, Brooks and Katz (1988) and Suchey-Katz (1986; with 1987 refinements) pubic symphyseal face system, and the auricular surface method (Lovejoy et al., 1985). These techniques were chosen based on their alleged ease of application and likelihood of being employed by other researchers as well as for their professed accuracy.

Once it was established that the information in the parish registers was accurate, both general and specific comparisons were made between the results of the skeletal analyses and the parish burial records. The sex and age profiles generated from the parish register data were compared to the sex and age distributions produced from the skeletal material in an effort to identify significant patterns of differences which would indicate that the methodologies under consideration were inadequate. The sex and age estimations of the 55 personally identified adults were compared to the actual sexes and ages in order to evaluate the levels of inaccuracy and bias of each method.

The nature of the St. Thomas' skeletal sample sets this investigation apart from other studies which have focused on the accuracy and reliability of pelvic and skull macroscopic methods of determining sex or on the accuracy and bias of age estimation techniques. Much of the research in this area has been conducted on autopsy material (Sutherland

and Suchey, 1987), medical school cadavers (Lovell, 1989), undocumented archaeological collections (Kelley, 1978) and large dissection room samples such as the Terry (Murray and Murray, 1991), Grant (Bedford et al., 1989) and Hamann-Todd (Meindl et al., 1985; Lovejoy et al., 1985) collections. The problem with archaeological material is clear: the sex and age of the individuals are generally unknown. While such information is available for dissecting room collections, these samples are generally biased in favour of individuals with below average socioeconomic status and, consequently, inadequate diet and below average health (St. Hoyme and Īşcan, 1989).

Researchers who employ dissection room samples to develop new techniques could easily mistake the results of poverty for normal variation due to sex (ibid) and age. Testing these techniques upon similarly affected collections would simply re-enforce such misconceptions. In order to reliably test these methodologies, it is necessary to employ material free of these biases. Ideally, one should examine a large skeletal sample of formerly healthy individuals, for whom documented sex and age are available. Since most samples are plagued by some type of bias, the next best approach is to test the available techniques on a wide variety of samples from diverse backgrounds.

The St. Thomas' cemetery site is particularly important since it is the largest archaeologically excavated historic cemetery sample in North America. In addition, the existence of parish records, newspaper obituaries, census data and County Directories makes this sample invaluable (Saunders et al., 1991). To date, there are few examples of the use of historic cemetery material as a means to test skeletally derived sex and age (see MacLaughlin and Bruce, 1990; Lanphear, 1989). Thus the St. Thomas' cemetery site has provided the opportunity for a relatively unique approach to testing methods of determining sex and estimating age from the human skeleton.

CHAPTER TWO
2.0 THE SAMPLE AND ITS ASSOCIATED DOCUMENTATION

2.1 SITE LOCATION AND BACKGROUND

The St. Thomas' cemetery is located on the north side of the St. Thomas' Anglican Church, situated at the corner of Bridge and Church streets in Belleville, Ontario. The city itself is located midway between Toronto and Montreal on the north shore of the Bay of Quinte (Mika and Mika, 1986). Belleville's beginnings date back to the 1780's when Asa Wallbridge constructed a log cabin on the banks of the Moira River. By 1789 approximately 50 United Empire Loyalist families had arrived and formed the first permanent pioneer settlement at the present day site of Belleville. Native peoples, including Cayugas, Mississagas, and Mohawks had been established in this region for some time prior to the arrival of the Loyalists (Boyce, 1991).

Originally known as Thurlow Village, the town's name was officially changed to Belleville in 1816 following a visit by Lieutenant-Governor Gore and his wife Anna Bella. The population at this time was 150 individuals (ibid).

An important aspect of early Belleville was its religious life. St. Andrew's Presbyterian, St. Michael's Roman Catholic and St. Thomas' Anglican Church were all established during the 1820's, at which time the population had risen to nearly 1000 (ibid). The St. Thomas' Parish was actually founded in 1818; the church was completed in 1821; and the first burial took place in August of that same year (Bellestedt, 1969; Diocese of Ontario Archives [DOA] series I, part 35A; Gerry Boyce, pers. comm.). St. Thomas' burial ground was not the only established cemetery in the surrounding region, but it was used by Methodists, Presbyterians and others who were not members of the church (Boyce, pers. comm.).

By 1834, when Belleville was incorporated as a village, it contained 613 acres of land and was home to nearly 1,700 people (Mika and Mika, 1986). The 1840 Religious Census reveals that Thurlow township, one of the oldest townships and one that eventually gave rise to the city of Belleville, housed at least 10 different religious denominations (Boyce, 1967). The population continued to rise, new churches were built and immigrants from the British Isles, Holland, France, Germany and other countries flocked to Belleville (Boyce, 1991; 1967; Mika and Mika, 1986). By the 1860's the population had reached 6,277 and by the late 1870's it had surpassed 11,000 (Mika and Mika, 1986). On May 1, 1874 a by-law preventing further interments at St. Thomas' was introduced. From 1874 onward, St. Thomas' parishioners were buried at the Belleville municipal cemetery (Bellestedt, 1969; Diocese of Ontario Archives [DOA] series I, part 35A; Gerry Boyce, pers. comm.). Belleville continued to expand and by 1860 its total acreage was 5,937 and the population in 1966 had grown to 32,785.

The passing of the years and two fires, one in 1876 and the second in 1974, took their toll on both the tombstones and the burial records stored at the Church. Thus, when the current St. Thomas' congregation decided to add a new parish hall to the church, it was thought that the building site contained only a minimal number of burials (McKillop et al. 1989)

2.2 FIELD METHODS.

The excavation of the cemetery was a carefully controlled process. Once the existing tombstones had been mapped and moved, a bulldozer was employed to remove some of the noncultural overburden soil. The locations of many of the individual burials were established through shovel-shining; shovel testing disclosed the rest. The large crew of archaeology students and trained volunteers, supervised by Dr. McKillop, were divided into teams. One group was responsible for locating burials through shovel-shining and test pitting, while the other recorded, measured, photographed, excavated and conducted shovel tests to the subsoil to ensure that all burials had been removed. Due to the well-drained sandy soil and excellent preservation, burials were located and removed in one day, thereby reducing any chance of accidental mixing of elements from different burials (Ibid; Saunders et al., 1991).

Once the excavation was underway, it became apparent that the cemetery was far larger than originally anticipated. Shortly after the work had begun, Gerry Boyce, a Belleville historian, began to examine the parish burial records for St. Thomas' at the Anglican Diocesan Archives in Kingston. The documents indicated that approximately 1500 individuals had been interred during the 53 year period the cemetery was in use.

Plans for the new parish hall were modified so that only one third of the total area believed to contain graves had to be excavated. Consequently, the parish burial records include information on the entire cemetery, but the excavated skeletal material is only a portion of the total number of individuals buried at St. Thomas'. According to the parish register, the 595 individuals removed from 579 graves during the excavation represent about 40% of the total burials (Saunders et al., 1991).

2.3 DOCUMENTATION

Since this thesis depends upon the sex and age information derived from the coffin plates and burial records, it was necessary to examine these sources to ensure that they were accurate.

The first step in this process was to cross check the data on the coffin plates to that listed in the parish burial records. This work was undertaken by historian Gerry Boyce. Mr. Boyce examined the nameplates, recording all the legible characters, then compared the nameplate data to the information in the parish registers. In some cases he could discern actual names and ages from the coffin plates, in other instances he had only initials and year of burial. The final list of "identifieds" includes individuals of known age and unknown sex. Cases of this type resulted when age was the only information clearly visible on the nameplate. Without additional information, identity could not be firmly established. When only the name of the individual was present on the coffin plate, the parish registers were searched until the name in question was located. The associated age could then be recorded. Thus the "known" individuals are those for whom information was clearly visible on the coffin plate and, in most cases, could be confirmed by the parish records (Boyce, pers. comm.).

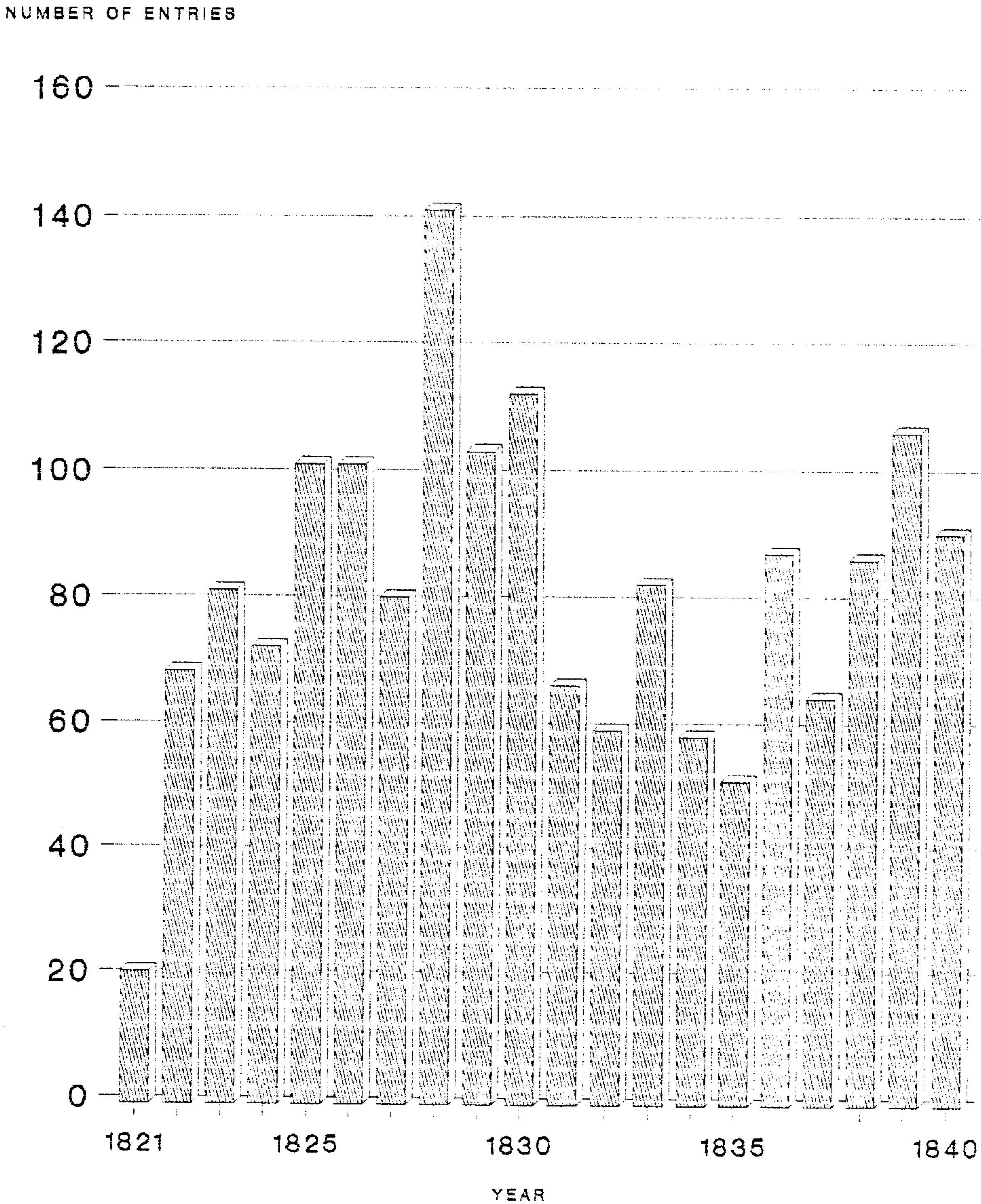
The majority of names and ages taken from the coffins were in agreement with the parish register. There were, however, four names that could not be matched (Boyce, pers. comm.). According to the coffin plates or domes, three of these individuals were male and one female. Two of the individuals died in 1865, one in 1871 and, in one instance, the year of death is unknown. Unfortunately, these unlisted names suggest that the register is incomplete, and perhaps a sex and age profile based on this information would be a misrepresentation of the actual cemetery sample. In order to address this possibility, the parish register was assessed according to Drake's (1974) methodology.

This process involves the detailed examination of parish records in order to establish whether they can be considered complete for the purpose of analysis. The procedure is as follows: establish that there are one hundred entries per year; look for obvious gaps in the register; look for other evidence of under registration; look for evidence that people buried in other churchyards were registered in the parish records; show that the sexes and ages in the records were not themselves estimates.

1. The initial step in the evaluation of parish registers was to establish that there were at least one hundred entries, comprised of marriages, baptisms and burials, in the book per year. This figure is, in a sense, arbitrary. Its purpose is to give the researcher a foundation for statistical analysis. Years containing below average entries are extremely difficult to evaluate. Normal and expected fluctuations cannot be differentiated from under-registration (Drake, 1974).

Unfortunately, at St. Thomas' the first twenty years were marked by annual totals falling just short of the minimal requirement (Figure 1). This does not mean that the parish records cannot be used, it simply suggests that careful consideration be given to this particular period. It is quite

FIGURE 1. TOTAL NUMBER OF ENTRIES* IN THE ST. THOMAS' REGISTER, FIRST TWENTY YEARS



* INCLUDES BAPTISMS, MARRIAGES AND BURIALS

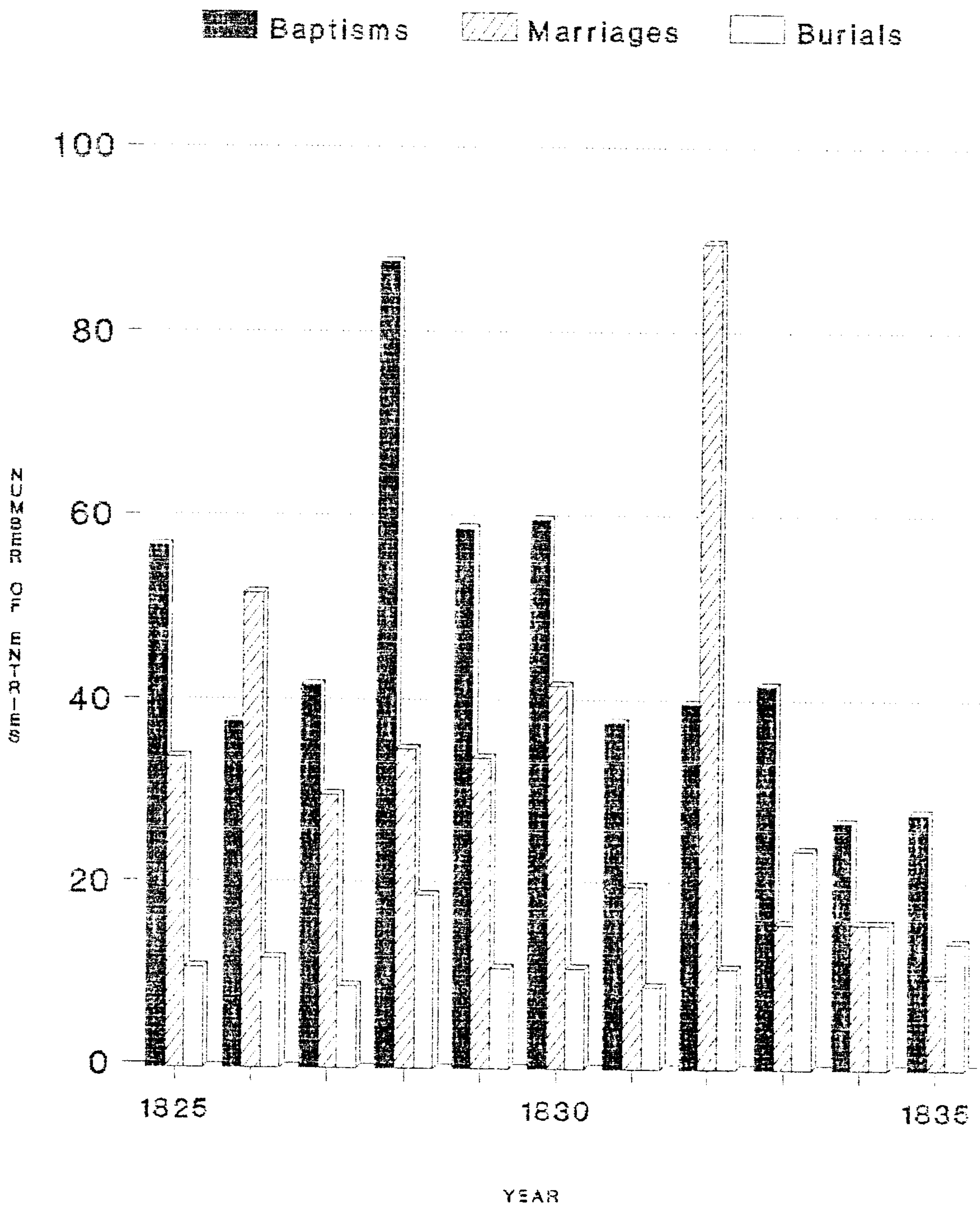
possible that the congregation at St. Thomas' was relatively small during the first twenty years and therefore incapable of generating one hundred entries per annum. Vestry minutes are available for this period but they have not yet been analyzed (Boyce, pers. comm.). By 1840, the situation had altered, one hundred entries or more was commonplace. Bellstedt (1969) supports the hypothesis that the congregation was much smaller during the first twenty years stating that, "eventually, [however] the church became too small to accommodate the congregation" (1969: 8) and in 1850, a committee to obtain plans for a new church was formed (ibid). For a detailed look at the number of entries per year over the entire fifty-three years of the cemetery's use, see Appendix 1.

Entries for 1821 were exceptionally low because the records did not begin until July of that year. Thereafter, a gradual increase in the number of events took place; fluctuations were minor. The exception to the general pattern was 1830-31 when the total number of entries dropped from 112 to 67. A decrease of this magnitude could conceivably indicate a period of under-registration, but the drop in number does not coincide with any change to the parish itself. Thomas Campbell remained rector during this period and there were no notations in the register indicating a disturbance in regular activities (DOA microfilm series I, part 35A). The decrease in total activities seems to have been caused by a reduction in baptisms and marriages. Since burials do not contribute substantially to the overall decrease (11 in 1830, 9 in 1831), there is no reason to suspect that burials have been under-reported (Figure 2).

Entries continued to remain low for the following four years. The reduction may have been partly related to the response of the Belleville town council to the cholera epidemics of 1832 and 1834. Belleville escaped the 1832 epidemic by posting armed guards on all roads leading into the town (Mika, 1986). Since St. Thomas' parish included individuals from the surrounding area who may have been prevented from entering the town (ibid.), the below average number of proceedings recorded during this period could have been a reflection of reduced activity at St. Thomas' as a result of cholera. No one in the town itself actually contracted cholera (Mika, 1986) and burials for 1832 reflect this with neither a noticeable rise nor a drop in number. The evidence is inconclusive. Although the possibility of under-registration cannot be excluded, it appears that burials were largely unaffected.

Another possible explanation for the decrease in entries is the poor health of Rector Campbell. Although the course of his illness was not fully documented, it is known that he became confined to his house in January 1835 and that he died the following September (DOA microfilm series I, part 35A). Perhaps 1831 marked the beginning of the rector's decline. The lowest number of entries during the twenty years under

FIGURE 2. ST. THOMAS' PARISH REGISTER ENTRIES
1825 - 1835



consideration occurred in 1835, but once again burials do not appear to have been affected (14 burials took place).

2. The first two of Drake's (1974) tests involved examining the register for obvious gaps between years or months. The first entry for St. Thomas' cemetery was in July, 1821. Fortunately, the burial sequence was uninterrupted and there were no gaps of one year duration from July 1821 to May 1874, at which point the cemetery closed. Two gaps between months took place: in 1823, no entries were made for April; in 1835, no entries were made for June. Possible explanations for this absence will be discussed below. Drake (1974) notes that gaps of less than two to three months duration are not considered significant, therefore St. Thomas' passes this test of completeness.

3. The next consideration was continuity of record keepers. Changes in handwriting and/or in the name of the individual recording the events had to be examined to ensure that they did not correspond to gaps in the records. At St. Thomas' Church, the rector served as the record keeper (DOA microfilm series I, part 35A), thus it is necessary to prove that the death of a rector did not result in a period of under-registration.

From 1821 until his death in 1835, Thomas Campbell was the rector of St. Thomas' Anglican Church. In January of 1835, Thomas Campbell was confined to his house. For the following seven months, various officiating ministers attended to the duties of rector of St. Thomas' Church, including Thomas Campbell whenever his health permitted. During this period one month contains no entries. In May, Job Deacon conducted a funeral and in July, John Cockran took on the duties of rector, although he was not officially ordained until after Campbell's death in September (DOA microfilm series I, part 35A). The fact that there was a change in ministers over the same month, June 1835, during which no entries were recorded is probably no coincidence. On no other occasion when a change in ministers took place does a gap in the records occur.

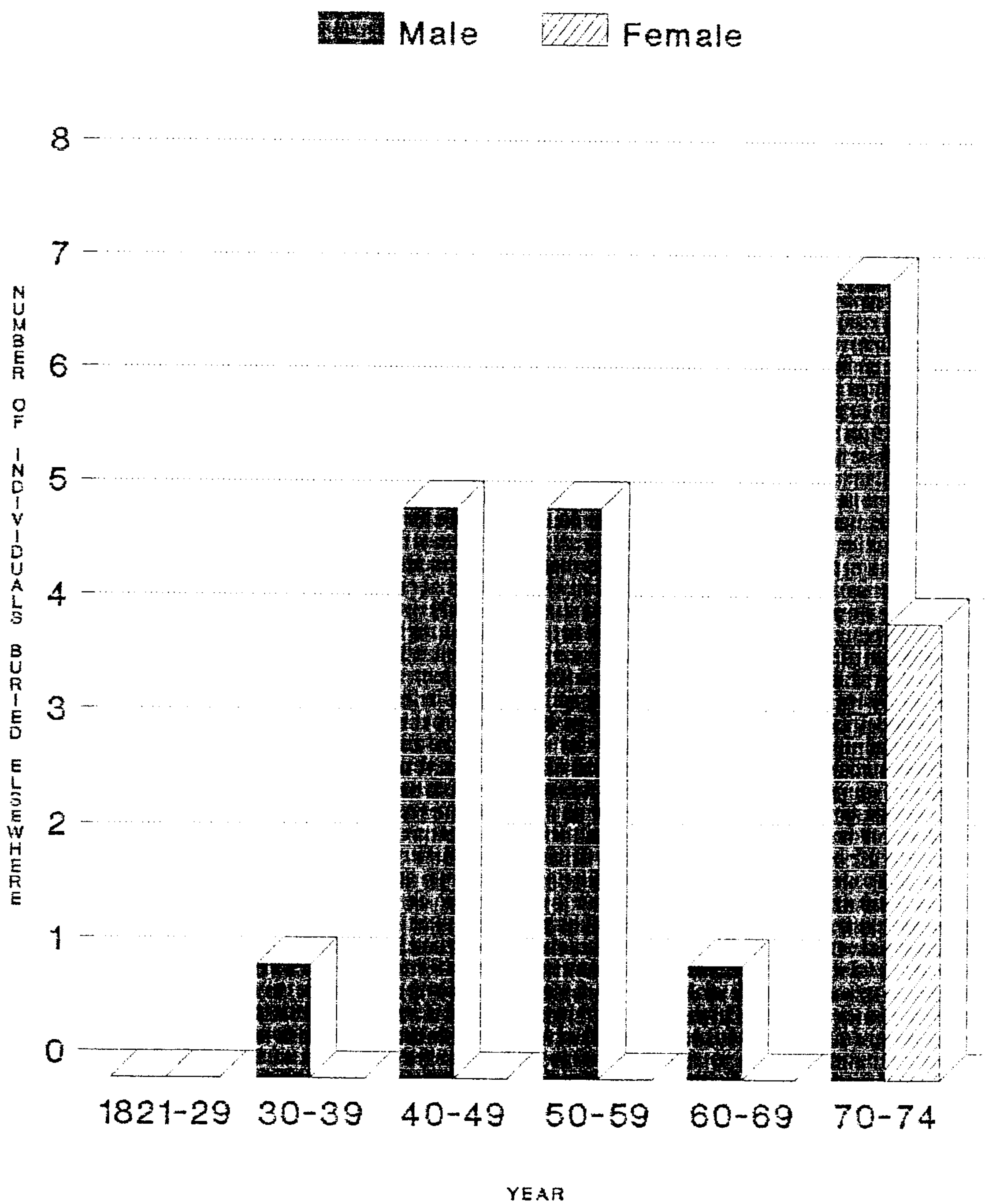
Although there is no way of knowing whether the absence of data is due to the absence of a minister during this period, or due to the fact that records were not being kept, three different ministers did manage to officiate and record events between January and May, 1835, without producing obvious gaps. This suggests that during the changeover from Deacon to Cockran events could have been relocated to another parish or forestalled until the arrival of the new minister. Only death could not have been delayed. The question is, if a death had occurred during June, 1835, would the individual have been buried at St. Thomas'? If this was the case, would there have been enough deaths during that particular month to skew a sex and age profile produced from the parish records? In order to answer the latter question an examination of deaths occurring during the month of June for the decade 1831-

1840 was undertaken. On average there were only 1.5 deaths during the month of June. It is unlikely that one or two unrecorded burials, even assuming that both were adults, could cause a bias in the profile of an adult sample of approximately 857, although it would reduce the number of individuals who could be cross-referenced with the coffin plates.

4. The next step involved reviewing the records for evidence that people buried in other cemeteries were registered in the St. Thomas' parish records. For various reasons (for example, a parishioner died in another city but due to the affiliation with St. Thomas' the death was recorded in this register), there were twenty-three cases in which the name of an individual buried elsewhere was recorded in the St. Thomas' register during the period 1830-1874 (Figure 3). These cases are clearly marked. The minister in question made an effort to report that the burial did not take place at St. Thomas'. Therefore, it appears that all funerals entered in the parish register represent individuals interred at St. Thomas', unless otherwise noted.

5. The last step in this evaluation process was to prove that the sexes and ages in the records were not estimates. Satisfaction of the accuracy of documented sex was a straightforward procedure. Most individuals were recorded by name and/or were given titles such as "Mrs." or "Mr." and/or were defined by some relationship to another

FIGURE 3. PEOPLE REGISTERED AT ST. THOMAS' AND BURIED ELSEWHERE



individual, all of which contributed to remove any ambiguity about an individual's sex.

There were only 13 individuals (age unknown) for whom it was impossible to determine sex from the records. These were cases in which surnames were accompanied by the notation "burial" but no entry followed. Occasionally, there was a blank in place of a given name. Assuming that all 13 individuals of unknown sex were adult, this represents only 1.52% of the 857 adults listed in the burial register and cannot be considered a significant proportion. All calculations based on adult sex ratios take these thirteen cases into account by presenting ranges rather than specific counts.

Establishing the accuracy of recorded age-at-death was much more problematical. People are often unaware of the actual age of a friend or relative and when required to state an age, simply provide an estimate. Such practices could only be ruled out in cases where age was recorded in years and months or in years, months and days. In all other instances, one could never be certain whether some element of guess work was involved. Ages that are even rather than odd and those ending with (0) are particularly questionable since, as noted by Drake (1974), people are more likely to guess even numbers than odd, and ages ending in (0) tend to be favourites of those doing the guessing.

In order to test for bias of this nature, all of the recorded ages were divided into categories based upon the digit in the units column. The totals are presented in Figure 4. Clearly, ages ending in 0 and in 5 are more frequently represented than other ages. To test whether there is a significant difference between ages ending in odd numbers and those ending in even, a binomial test was run on SPSS PC+ v.2.0. Binomial tests indicate "whether the observed distribution of a dichotomous variable is the same as that expected from a specified binomial distribution" (Norusis/SPSS INC., 1988: C100). See Appendix 2. for a description of the binomial test.

In this case the expected frequencies for both odd and even numbers were 0.5. The first binomial test excluded ages ending in 0 and 5 in order to determine whether the less commonly used numbers exhibited a bias in favour of even ages. The number of cases ending in an even digit were totalled as were those ending in an odd digit. The binomial test was conducted using these totals (see Figure 4). The observed proportion was 0.5071 for odd ages, indicating that there is no significant difference between odd and even ages ($p=0.7678$). A second binomial test, including ages ending in 0 and 5, was run. The observed proportion for odd ages was 0.4987, indicating once again that there was no significant difference between odd and even ages ($p=0.9711$). Although some of the ages may be estimates, there is no apparent bias which would preclude comparisons to skeletally-derived age

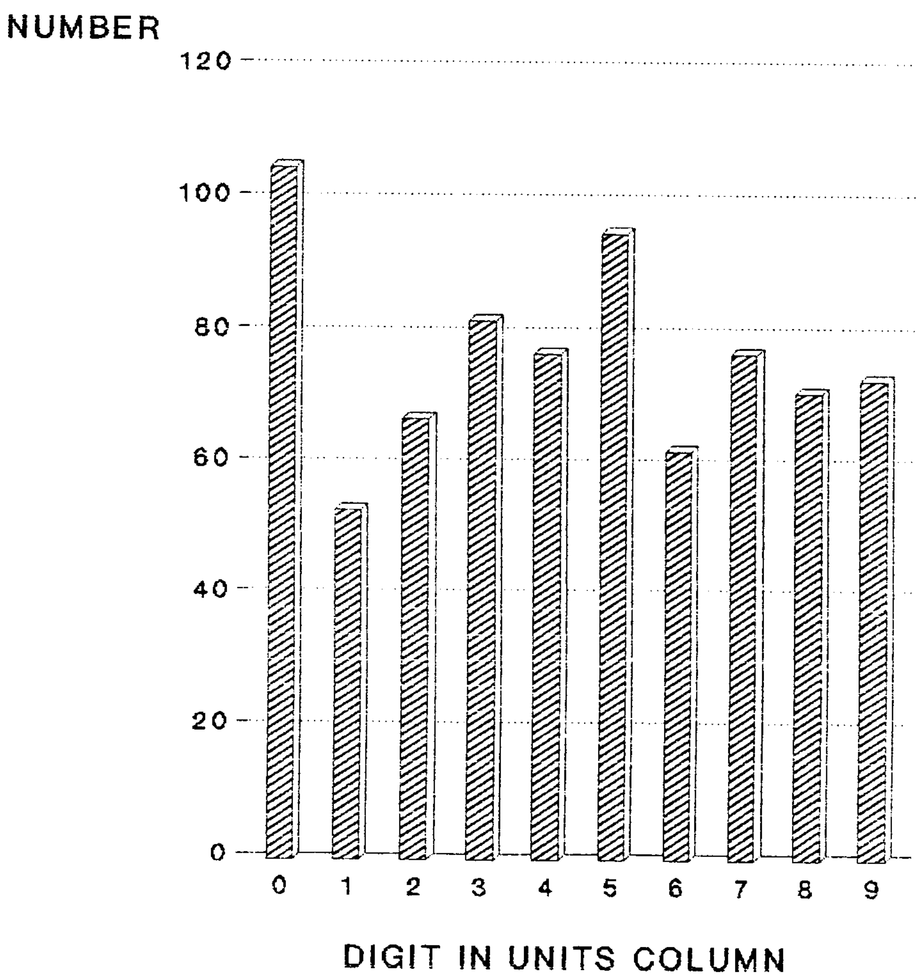
FIGURE 4. NUMBER OF RECORDED AGES BY DIGIT IN THE UNITS COLUMN

UNITS	0	1	2	3	4	5	6	7	8	9
NUMBER OF CASES	105	53	67	82	77	95	62	77	71	73

DISTRIBUTION OF DIGITS IN UNITS COLUMN

OF AGES RECORDED IN THE

ST. THOMAS' PARISH REGISTER



estimates.

In addition to problems associated with individual records of known age, problems exist as a result of the absence of exact ages. There are 95 individuals for whom age is unknown. This represents a maximum of 11.1% of the adult burial sample if all of the 95 individual records under consideration are adult. Of those 95 cases, only 61 are confirmed adults. Thus at least 7.1% of the adult group is of unknown age.

Every decade has at least five cases of unreported age at death (Figure 5). The responsibility for these absences belongs to a number of rectors, suggesting that the absences are random and that no particular age group has been under-recorded. In order to test this hypothesis, two specific time periods were considered. Twenty-five of the individuals of unknown age (26%) were buried during 1830-1839. When the documented ages of all the people buried during this period are examined, no visual evidence of under-reporting of any age group is evident. The mortality pattern for 1830-39, is similar in appearance to the mortality pattern of the remaining years (compare Figures 6 and 7). The same can be said for the period 1860-1865 when an additional twelve (13%) individuals of unknown age were interred.

A log-linear test with chi square residuals was performed on Systat using a BIT 386 microcomputer to determine whether the two subsample age distributions are significantly

FIGURE 5. DISTRIBUTION OF INDIVIDUALS OF UNKNOWN AGE IN THE ST. THOMAS' PARISH REGISTER (N=95)

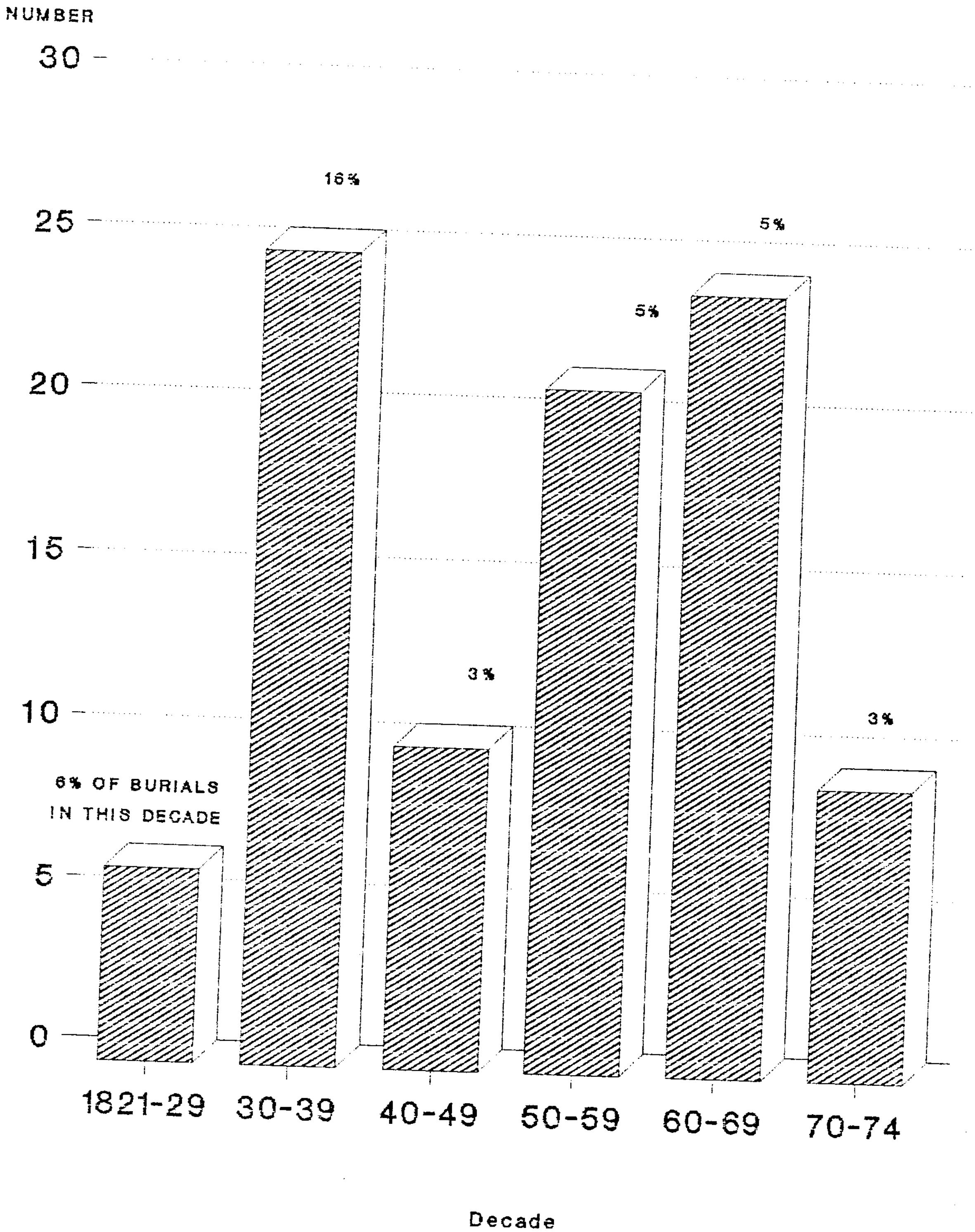


FIGURE 6. DISTRIBUTION OF RECORDED AGES
1830-39 COMPARED TO 1821-74

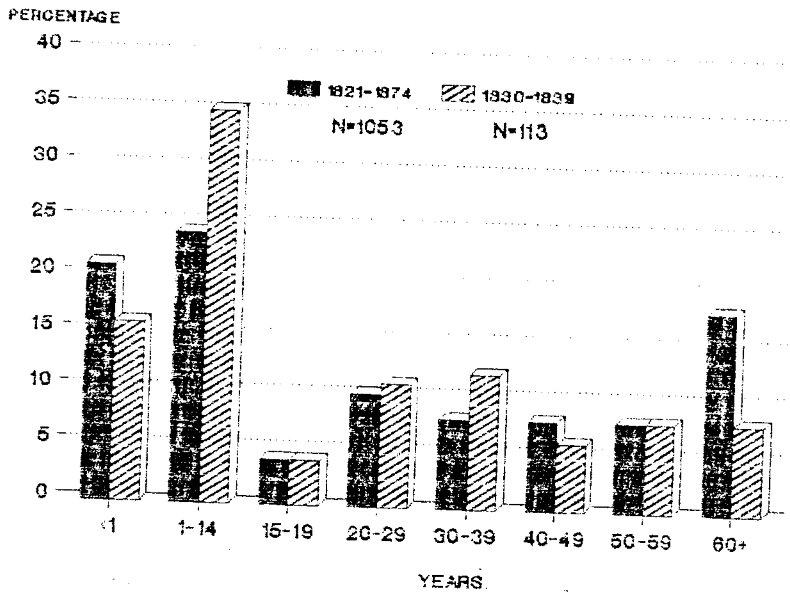
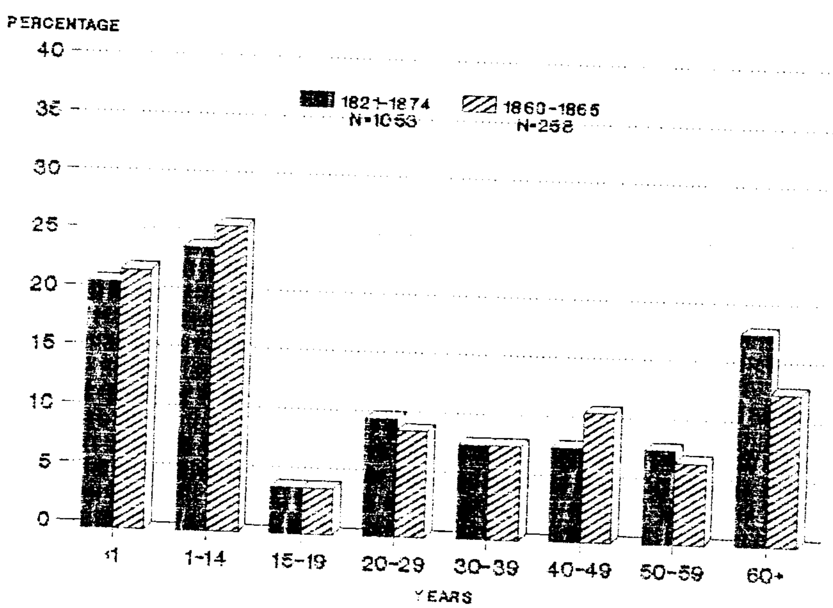


FIGURE 7. DISTRIBUTION OF RECORDED AGES
1860-65 COMPARED TO 1821-74



different from the rest of the sample. An explanation of the log-linear test is presented in Appendix 2. The chi square residuals are given in Tables 1 and 2. A residual greater than 1.64 can be considered significant at the 0.05 level (Wilkinson, 1988).

The data indicate that there is a significant difference between the 1830-39 distribution and the distribution spanning 1821-1874 (Likelihood Ratio Chi square= 14.43, p=0.044, df=7), but that no significant difference exists between the 1860-65 distribution and the distribution spanning 1821-1874 (Likelihood Ratio Chi square= 6.06, p=0.533, df=7). Table 1 shows an abundance of individuals in the "one to fourteen years" category and an under-representation of individuals in the "sixty years plus" category for 1830-39. While the abundance of children does not present any difficulty for this adult-based study, the under-representation of older adults could be a problem. There are two possible explanations for this finding. First, proportionately fewer people over the age of sixty died and were buried at St. Thomas' during 1830-39 than died and were buried during the remaining time the cemetery was in use. Second, for whatever reason, the ages of a significant number of people sixty years and older were not recorded in the parish burial records.

There is good reason to believe that the former explanation is the correct one. Belleville in the 1830's was still a relatively new community. The first Loyalists had

TABLE 1. RAW DATA AND CHI SQUARE RESIDUALS FOR 1830-39

AGE RANGE	1830-39 NUMBER OF BURIALS BY AGE	1821-29 1840-59 1866-74 BURIALS BY AGE	1830-39 CHI SQUARE RESIDUALS	1821-29 1840-59 1866-74 RESIDUALS
< 1 YEAR	18	224	-1.13	.37
1 -14 YEARS	40	255	2.13 *	-.70
15-19	5	39	.36	-.12
20-29	12	103	.26	-.08
30-39	13	83	1.21	-.40
40-49	7	85	-.64	.21
50-59	9	79	.16	-.05
60 +	9	185	-2.26 *	.74

Chi square= 14.43 d.f.=7 p=.044

TABLE 2. RAW DATA AND CHI SQUARE RESIDUALS FOR 1860-65

AGE RANGE	1860-65 NUMBER OF BURIALS BY AGE	1821-29 1840-59 1866-74 BURIALS BY AGE	1860-65 CHI SQUARE RESIDUALS	1821-29 1840-59 1866-74 RESIDUALS
< 1 YEAR	57	224	.23	-.11
1 -14 YEARS	68	255	.56	-.28
15-19 YEARS	10	39	.11	-.06
20-29	22	103	-.52	.26
30-39	20	83	-.06	.03
40-49	29	85	1.39	-.69
50-59	19	79	-.07	.03
60 +	33	185	-1.51	.75

Chi square= 6.06 d.f.= 7 p=.533

* Significant at .05 level.

only established themselves in the front townships of Hastings County in 1784 and the first large group of loyalists had only arrived in the Belleville area in 1789 (Boyce, 1967). Population growth due to immigration was most rapid in the dozen years before 1837 (ibid). Since most migrants are adolescents and young adults between the ages of 15 and 35 (Browning, 1971; Grigg, 1977), it is quite probable that during and prior to the 1830's, there were proportionately fewer individuals over the age of sixty in Belleville. The under-representation of individuals sixty years and older in the parish records during 1830-39, is more likely the result of the demographic features of the period than of under-reporting of the ages of older adults.

Thus, it appears that the individuals of unknown age who died between 1830-39 and 1860-65, were randomly distributed throughout the age range. If this holds true for the remaining unknowns, then the absence of exact ages will not introduce significant bias into the age profile of the population.

In conclusion, it is evident the St. Thomas' parish register is not perfect in every respect. The coffin data indicate that some individuals were buried without record, but an assessment of the documents, following Drake (1974), suggests that these are isolated examples. Reductions in parish events occurred during 1831-1835 but the number of burials was unaffected. No significant gaps in the registers were detected. Parishioners buried elsewhere were so noted in the register. Individuals of unknown sex represent a nonsignificant proportion of the total adult sample (1.52%). Those of unknown age appear to be randomly distributed throughout the age range. Consequently, a sex and age profile created from the register data should be unbiased and reliable. It can confidently be compared to the skeletally-derived sex and age profile from the identified portion of the cemetery sample.

CHAPTER THREE
3.0 SEX DETERMINATION

3.1 INTRODUCTION

Methods of determining the sex of an individual based upon skeletal remains can be broadly divided into two groups: morphological observation, also referred to as "subjective"; and metric, or "objective" techniques. Various researchers have expressed a preference for metric analyses which employ a variety of measurements in order to develop regression formulae. It has been argued that such metric approaches result in fewer indeterminate cases, that they broaden the range of bones that may be employed to determine sex, are easier to teach and are more repeatable than morphological assessments (Giles and Elliot, 1963; Henke, 1977; Kajanoja, 1966; Keen, 1950; Kelley, 1979; Richman et al., 1979; Washburn, 1948). Others have observed that techniques involving measurements usually require better preservation of the remains, they may employ ill-defined landmarks, vague definitions of procedures, and measurements which inherently have a limited descriptive ability (Meindl et al., 1985; Stewart, 1954; Weiss, 1972). Furthermore, it has been noted that the discriminatory power of existing functions such as that developed by Giles and Elliot (1963) are severely limited in series where the pattern of sexual dimorphism does not follow the original sample (Buikstra and Mielke, 1985).

Many investigators have recognized that both approaches have comparable levels of accuracy and that each contains inherent advantages and disadvantages (Acsádi and Nemeskeri, 1970; Buikstra and Mielke, 1985; El-Najjar and McWilliams, 1978; Krogman and Īşcan, 1986; Stewart, 1979; Workshop of European Anthropologists, 1980). The two methods are complementary in nature and both should be considered whenever possible. It was therefore decided that the two types of data would be collected for the Belleville project. The decision to focus this thesis on morphological indicators of sex rather than metric was not made arbitrarily. The St. Thomas' sample is particularly well-suited to addressing the following issues related to aspects of morphological sex assessment.

1. Most osteologists identify the pelvis as the single most important region for the determination of sex since the female pelvis is modified during adolescence to accommodate the birth of an infant. Second in importance is the skull, which reflects sex characteristics such as size and rugosity (Acsádi and Nemeskeri, 1970; Bass, 1987; Brothwell, 1981; El-Najjar and McWilliams, 1979; Krogman and Īşcan, 1986; St. Hoyme and Īşcan, 1989; Steele and Bramblett, 1988; Stewart, 1979; Ubelaker, 1984; Workshop of European Anthropologists, 1980). This assertion will be tested by comparing the accuracy of the pelvic and skull morphological methods in

addition to the individual criteria upon which these methods depend.

2. As El-Najjar and McWilliams (1979) note, researchers rarely agree upon the recommended weights of each particular feature. Regarding the pelvis, criteria reflecting anterior pelvic growth at adolescence tend to be preferred over those indicative of posterior growth, since evidence of the former is frequently clearer (Coleman, 1969; Phenice, 1969; St. Hoyme and Īşcan, 1989). Despite this general concordance, however, there is still some disagreement over the relative value of each criterion. This ongoing dispute will be discussed below.

With respect to the skull, the weighting of individual criteria appears to be a matter of personal preference, based largely upon experience, rather than empirical evidence (Krogman and Īşcan, 1986; Stewart, 1979; El-Najjar and McWilliams, 1979; Hrdlicka, 1952). The problem is that it has not yet been determined whether the characteristics deserve the weightings attributed to them. The rankings may simply be a reflection of the observers' biases.

It is proposed here that an examination of both pelvic and craniofacial growth will indicate which characteristics within each of these categories will provide the most accurate assessments of sex. This prediction can then be tested on the personally identified subsample of the St. Thomas' cemetery material.

3. The Phenice method (1969) is frequently cited as the

best pelvic discriminator because of the ease with which observations can be made and the high level of accuracy (96%) that may be obtained (Bass, 1987; Buikstra and Mielke, 1985; Phenice, 1969; Stewart, 1979; Weiss, 1972). Recently, some controversy concerning this technique has arisen.

Phenice identifies the ventral arc as the most useful of the three criteria. Kelley (1978) reaffirms this finding using 362 hipbones of unknown sex from Native Californian archaeological sites. After comparing a metric assessment of sex from these unknown archaeological cases to the Phenice criteria, he observes that there were fewer indeterminate cases for the ventral arc than for either subpubic concavity or ischiopubic ramus.

Sutherland and Suchey (1991; 1987) examined 1284 well-documented pubic bones obtained during autopsies. They agree with Phenice, citing an accuracy of 96% using the ventral arc alone and only 70% using the ischiopubic ramus (subpubic concavity was not assessed since the procurement technique employed during dissection damaged the region and made observations impossible). In addition, they identified a precursor condition, a faint line following the course of the ventral arc, which appeared at about age fourteen in their sample and was the most frequent condition by age twenty (Sutherland and Suchey, 1991: 504).

In contrast, Lovell (1989) initiated intraobserver tests of the Phenice criteria using 50 pubic bones from 36 medical school cadavers and found that most researchers participating in her test felt that the ventral arc was ambiguous. No differential success in the application of the three criteria was noted.

MacLaughlin and Bruce (1990) selected 273 skeletons of documented age and sex from three European skeletal collections: a 17th to 18th century English cemetery (85 individuals); a modern dissecting-room sample from the Netherlands (136 cases); and another modern dissecting-room sample from Scotland (52 individuals). They observe that the subpubic concavity proved to be the single most reliable indicator and, in fact, tends to be better than the three features in combination. Yet, even when using the ventral arc, they could achieve only 87% accuracy on the English sample, 72% on the Dutch, and 75% on the Scottish collection.

Due in part to the controversy surrounding the ventral arc, detailed dissections of the pubis were completed by Anderson (1990) and Budinoff and Tague (1990) in order to determine the anatomical and developmental bases for this characteristic. Using 23 cadavers, aged 63-94 years, and 8 cadavers, aged 14-21 years, Anderson (1990) observed that the ventral arc marks the origin of the female gracilis muscle with contributions from adductor magnus and adductor brevis muscles. He notes that in the male the origins of the same muscles, "parallel the contour of the pubic symphysis and inferior ramus" (Anderson, 1990: 452). Although he does not state it in this manner, Anderson's observations on the male indicate that the muscular origins in question conform to the male version of the ventral arc as described by Phenice (1969). Anderson's findings are in agreement with those of Budinoff and Tague (1990). Having dissected only 7 cadavers, they state that, "although males and females generally differ in the course taken by the ridge of bone on the ventral aspect of the pubic corpus, the sexes are identical in the muscular and ligamentous attachments to that ridge" (Budinoff and Tague, 1990: 76).

While both studies agree with respect to the relationship between the ventral arc and the muscular origins of some of the adductors, they focus on slightly different aspects of their findings and, consequently, present disparate conclusions. Anderson (1990) concentrates upon the lateral position of the origins with respect to the pubic symphysis. Although he recognizes that the female muscular origins are associated with the ventral arc, he fails to realize that the male sites of muscle origin conform to the bony ridge paralleling the pubic symphysis; a ridge, identified by Phenice (1969) as the male counterpart to the female ventral arc. Thus Anderson was able to conclude that:

> Simply stated, a more medial origin of the three muscles is seen in males because the symphyseal surface does not impinge upon the ventral surface of the os pubis. With females, however, the invasion of the symphysis onto the ventral surface of the pubic bone prohibits the same medial origin. Because this symphyseal wrap-around is the most lateral expanse of the symphyseal surface, which is within the joint capsule and covered by the

> ligamentous fibers of the capsule, these muscles cannot attach at the same site as in males. When the pubescent pubic elongation begins in females, the musculature that was once attached along the symphyseal border is "pushed" laterally by the increased growth that takes place at the inferior margin of the symphysis... Therefore, males are unlikely to ever mimic the female condition
> (Anderson, 1990: 456-457).

On the other hand, Budinoff and Tague (1990) focus on the relationship between the muscular attachments and the ridges, rather than on the degree of lateral placement of the adductor origins. Having taken a slightly different perspective, they conclude that, "the lateral placement of the ventral arc is directly related to pubic length" and imply that males with long pubes might be interpreted as having a ventral arc (1990: 78).

Thus Anderson's work suggests that the ventral arc should have an extremely high level of accuracy, comparable to that achieved by Sutherland and Suchey (1987). Budinoff and Tague (1990) are less optimistic concerning the use of the ventral arc as a single indicator and state that it cannot be considered independent of pubis length.

This analysis will address these issues by assessing each of Phenice's (1969) criteria individually. MacLaughlin and Bruce observe that theirs was the only study to, "evaluate the relative contribution of each of the three features to successful sex identification" (1990: 1390). In the present investigation, the accuracy of each characteristic was determined. The individual results were compared to the levels of accuracy achieved using all possible combinations of these features.

4. Finally, the St. Thomas' material is an excellent data base for testing the assertion of Meindl et al. (1985) that the proportion of males in the sample will be underestimated (following a morphological analysis of either the pelvis or the skull), while few errors will be made in assessing "true females". This view contrasts with Weiss' (1972) contention that a morphological analysis of the skull generally causes indeterminate cases to be classified as male, thus creating a significant bias in favour of males.

Because the St. Thomas' Church cemetery is a large documented sample, it is possible to test specific hypotheses associated with morphological approaches to skeletal sex determination, in addition to a simple analysis of the accuracy of these methods.

3.2 SEX AND THE SINGLE SKELETON: EMPLOYING THE BONES OF THE PELVIS TO ASSESS THE SEX OF HUMAN SKELETAL REMAINS

3.2.1 INTRODUCTION

As indicated above, the pelvis is considered to be the single best region for assessing the sex of human skeletal remains. Because of the distinct functional requirements of the female pelvis, adolescent differential growth of the bony pelvis occurs in both degree and direction. Consequently, most pelvic sex indicators differ in form, not simply in size (Coleman, 1969; St. Hoyme, 1984; St. Hoyme and Īşcan, 1989).

As a result, an investigator examining a single pelvis or even a single hipbone may confidently assign sex, whereas an observer attempting to determine sex based on a single skull is handicapped because of population differences which strongly affect the larger/smaller type features characteristic of sex differences of the skull (Acsádi and Nemeskeri, 1970; Krogman and Īşcan, 19862; St. Hoyme and Īşcan, 1989; Steele and Bramblett, 1988; Ubelaker, 1984; Workshop of European Anthropologists, 1980).

Yet, as demonstrated by the above discussion, pelvic indicators of sex are not entirely free from controversy. In this chapter the following four questions concerning sex characteristics of the pelvis will be examined:

(1) What levels of accuracy and reliability can be achieved using this method?

(2) Which trait(s) is/are the most accurate and reliable?

(3) With regard to the Phenice method: which of the three traits is the most useful?

(4) Does the morphological method of assessment produce a bias in favour of one sex, and if so, in which direction is the error?

3.2.2 THE TRAIT LIST

The features chosen for examination in this study were compiled from trait lists produced by many physical anthropologists including: Bass, 1987; Krogman and Īşcan, 1986; Phenice, 1969; St. Hoyme, 1984; and Stewart, 1979. Seventeen traits, representing the two functional divisions of the pelvis and reflecting differential growth patterns, were chosen. A copy of the data sheet employed is presented in Table 3 (see also Figure 8).

3.2.3 LITERATURE REVIEW

After examining two primary papers, Coleman's (1969) practical analysis of pelvic growth and St. Hoyme's (1984) theoretical approach to growth of the posterior pelvis, it was possible to generate some hypotheses and expectations about the relative worth of the traits in Table 3.

Having analyzed and measured 300 anterio-posterior x-rays taken of the pelves of thirty individuals from the ages of 9-18 years, Coleman (1969) states that the most significant example of sexually dimorphic growth is the greater lateral migration of the female ischia necessary to expand the pelvic cavity. He also notes that the bones of the true pelvis are more dimorphic than those of the false and that features relying on both anatomical systems for their final form are the most variable and therefore the least useful.

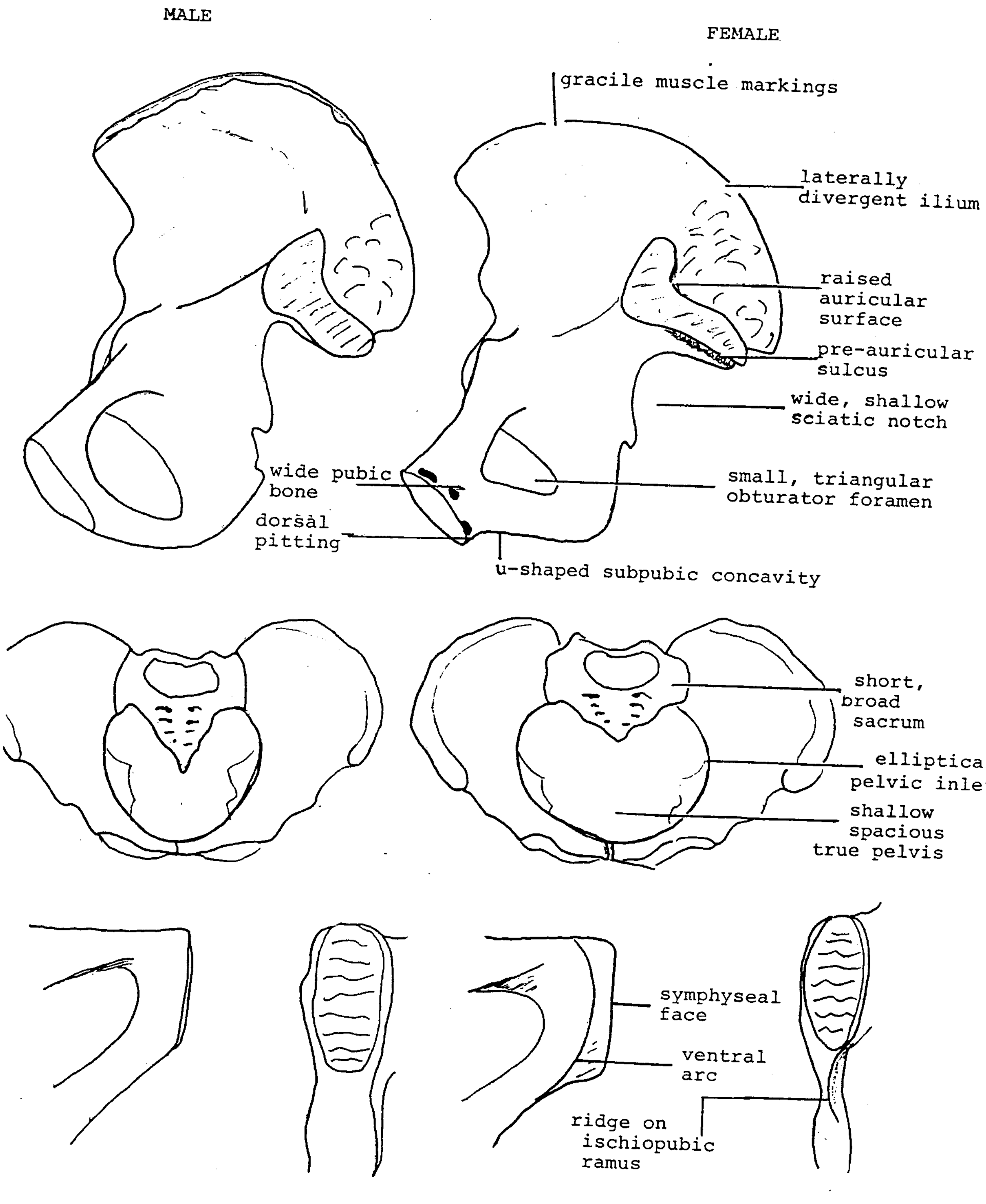
St. Hoyme (1984) postulates enlarging growth in the female posterior pelvis, analogous to that of the pubis. She identifies three possible growth patterns: (1) primarily on the iliac portion of the sacral-iliac joint, resulting in a raised auricular surface, a wide sciatic notch and the

TABLE 3. VISUAL SEX DETERMINATION OF THE ADULT PELVIS

GROUP/CASE: ____________ DATE: ____________
BURIAL : ____________ LOCATION: ____________
CONDITION : ____________
OBSERVER : ____________
SKULL ASSOCIATED: ____________

	MALE		FEMALE	
PUBIC BONE				
Subpubic Concavity (Subpub)	v-shaped angle	____	u-shaped angle	____
Ischiopubic ramus (Ischram)	ridge absent	____	ridge present	____
Ventral arc (Venarc)	absent	____	present	____
Shape of pubic bone (Pubshape)	narrow	____	wide, rectangular	____
Dorsal pitting (Dorpit)			present	____
Sciatic notch (Scinot)	small, close, deep	____	wide, shallow	____
ILIUM				
Auricular surface (aursurf)	not raised	____	raised post-sup margin	____
Pre-auricular sulcus (Sulcus)	absent thin grooves	____	large circular depressions	____
Ilium shape (Ilium)	high, vertical	____	laterally divergent	____
PELVIS				
Pelvic Inlet (Pelinlet)	heart-shape	____	elliptical	____
True Pelvis (Truepel)	relatively small	____	oblique, shallow spacious	____
Obturator foramen (Obfor)	large, ovoid	____	small triangular	____
Acetabulum (Acetab)	large, directed laterally	____	small, ant-lat	____
Muscle Markings (Musmark)	marked, rugged	____	gracile, smooth	____
SACRUM				
Shape (Sacrum)	longer, narrower, more evenly distributed curvature	____	shorter, broader tendency to curve at S1-S2,S3-S5	____
Number of segments (No.segs)	5+	____	5	____
Posterior View (Postview)	sacroiliac joints visible	____	not visible, in inferior part	____

FIGURE 8. Pelvic Traits



presence of both the preauricular sulcus and post-auricular groove, (2) growth concentrated on the sacral portion of the joint, resulting in a widened sacrum, a narrower "male"looking iliac crest, (3) a combination of both of the above.

If the theoretical views of St. Hoyme (1984) are combined with the observations of Coleman (1969) the following hypotheses result:

1. Features resulting from the modification of the female prepubescent pelvis to accommodate the birth of an infant will produce the most reliable indicators. Since lateral migration of the female ischia are responsible for increasing the pelvic outlet, traits which result from this migration will be particularly useful. According to Coleman (1969) the relevant characteristics are:

(a) size and shape of the true pelvis

(b) width of the subpubic angle

(c) shape/length of the pubic bone

St. Hoyme (1984) notes that the ventral arc marks the former border of the pubis and, therefore, is related to pubis length. Budinoff and Tague (1990) support this view. More recently, St. Hoyme and İşcan (1989) have added that the subpubic angle is associated with pubis shape and the presence of a ventral arc. They claim that all three characteristics are expressions of anterior pelvic growth and therefore, must be counted as a single point of evidence. Thus the ventral arc was added to Coleman's list with the caution that one must examine the co-occurrence of traits (b), (c) and (d) as expressions of the same phenomenon.

2. In her earlier paper, St. Hoyme (1984) identifies five features of the posterior pelvis which she believes would be as accurate as those of the anterior. More recently, however, St. Hoyme and Īşcan have indicated that evidence for remodelling in the posterior pelvis might not be as clear as that from the anterior and consequently, "this may lead to lower weightings for this group of factors" (St. Hoyme and Īşcan, 1989: 85). The posterior pelvic features expected to be only slightly less useful than those from the anterior are as follows:

(e) auricular surface

(f) presence or absence of the preauricular sulcus

(g) sciatic notch shape

(h) dorsal pitting resulting from the instability of the female sacro-iliac joint

(h) sacrum width/shape

Only the latter feature conflicts with Coleman's findings. He observed, "no sex differences in either the differential growth pattern or the amount of growth undergone by the local areas of the sacrum" (Coleman, 1969: 143). According to his analysis, the shape and posterior view of the sacrum should be of little value as a sex indicator.

Tague (1988) examined archaeological Amerindian samples and a sample of Blacks from the Hammann-Todd collection and found that while sexual differences in the degree of resorption of the pubis and preauricular area were present, in general, only a minority of females exhibited resorption more severe than that found in males. His investigation suggests that the preauricular sulcus and dorsal pitting should be given less weight than other features of the posterior pelvis.

3. Coleman identified other localized sites of differential growth that are not directly related to pelvic outlet increase. He feels that although these regions do exhibit sexual differences, they are not as obvious as those described above. He observed infero-lateral growth at the junction of the female ischial ramus and the pubic bone, an area corresponding to the attachment of the musculature of the penis in the male. Male ischial-pubic ramus growth was strictly inferior in direction. Also noted was the less steep slant of the female ilium relative to the male. Coleman (1969) suggests this pattern was due to growth between the thoracic vertebrae lengthening the trunk faster than the iliac crests are growing and/or a greater increase in trunk muscularity among most males and some females.

4. Features dependent upon both functional divisions of the pelvis for their final adult configuration will show little sexual differentiation. According to Coleman (1969) only the pelvic inlet and the sciatic notch are affected by both systems. St. Hoyme (1984) disagreed with Coleman, suggesting that the shape of the sciatic notch results from

growth along the iliac surface of the sacro-iliac joint and, consequently, this feature should be extremely useful.

5. Characteristics which depend upon differing growth patterns for their final form will be more useful for morphological determination of sex than will those dependent upon differential rates of growth. Thus the acetabulum, which shows no evidence of significant sex differences in the amount of growth (Coleman, 1969), muscle markings and the obturator foramen (St. Hoyme, 1984) will be the least useful characteristics.

Based on these five hypotheses, the examined traits may be separated into blocks and ranked according to their expected level of accuracy (Table 4). Sciatic notch and sacrum shape are listed twice since Coleman (1969) and St. Hoyme (1984) disagree with respect to the placement of these features. It is suggested here that Coleman's ranking is more probable since his study dealt with observations while St. Hoyme's approach was more theoretical.

The hypothetical ranking can also be used to predict the order of accuracy for the Phenice criteria. Phenice (1969) claims that the ventral arc is the most useful, followed by the subpubic concavity and the ischiopubic ramus. The work of Coleman (1969) and St. Hoyme (1984) suggests a similar pattern. The ventral arc and subpubic concavity are expected to be equally effective while the ischiopubic ramus should be less effective.

TABLE 4. HYPOTHETICAL RANKING OF PELVIC TRAITS

MOST EFFECTIVE (Coleman, St. Hoyme)	true pelvis shape/length of pubis	subpubic concavity ventral arc
SECOND MOST EFFECTIVE (St. Hoyme, St. Hoyme and Īşcan)	auricular surface sacrum	sciatic notch
THIRD MOST EFFECTIVE (St. Hoyme, Tague)	preauricular sulcus	dorsal pitting
FOURTH MOST EFFECTIVE (Coleman)	ischiopubic ramus	ilium shape
FIFTH MOST EFFECTIVE (Coleman)	pelvic inlet	sciatic notch
LEAST EFFECTIVE (Coleman and St. Hoyme)	acetabulum muscle markings posterior view of the sacrum	obturator foramen sacrum

With respect to bias and direction of error of sex estimations, Meindl et al. (1985) suggest that the female pelvis has limited variability due to the combined functional requirements of childbirth and locomotion. If, in fact, the form of the female pelvis is more constant, one would expect fewer errors to be made in the assessment of female remains. The male pelvis, being more variable, would be more prone to misclassification. Therefore, when using pelvic criteria one could expect a bias in favour of females. Tague (1989) refutes this argument on the basis of a metric analysis of the size and shape of the true pelvis of 200 skeletons of known sex and a series of Amerindian samples of unknown sex. He failed to find consistently more variation in male pelves and as a result suggests that if the pelves themselves are not more variable, perhaps the visual clues employed by anthropologists are more variable in male skeletons.

If the trait list is examined it becomes evident that most of the features under consideration depend upon differential growth of the female pelvis for their final form (St. Hoyme, 1984). Since this is the case, one would expect borderline individuals to be classified as male due to the lack of evidence for alteration. Furthermore, the female form of each characteristic can vary or be modified by age. For example, the ventral arc may not be palpable until after twenty years of age (Sutherland and Suchey, 1987). Young females may be misclassified because the ventral arc is clearly absent. On the other hand, the subpubic angle narrows with age (Tague, 1989) and therefore older females are also subject to misclassification.

Despite the fact that Meindl et al. (1985) found the opposite to be true, it is argued here that the direction of error will be to misclassify females because of the many factors which contribute to the variability of the characteristic female pelvic form. The bias will be in favour of males.

3.2.4 HYPOTHESES

It was possible to develop four main hypotheses based upon the pelvic growth literature. They are as follows:

1. The pelvic morphological approach to skeletal sex determination can accurately identify the sex of an individual and can produce an accurate sex profile of a sample. This hypothesis will be accepted if the morphological method has an accuracy level greater than or equal to 90% and if the sex ratio produced from the skeletal estimates is statistically indistinguishable from the ratio developed from the parish records.

2. The four features of the anterior pelvis: true pelvis shape; subpubic concavity shape; pubis shape and presence of a ventral arc will be more useful than the five features of the posterior pelvis: auricular surface height; sciatic notch shape; sacrum shape; presence of a preauricular

sulcus and dorsal pubic pitting. The criteria for acceptance of this hypothesis include: a mean accuracy for the anterior features which is at least 10% higher than the mean accuracy of the posterior features and a mean intraobserver error level which is at least 5% lower for the anterior features than for the posterior features.

3. Of the Phenice (1969) criteria, the ventral arc will produce the best results, followed by the subpubic concavity and the ischiopubic ramus. This hypothesis will be accepted if the ventral arc has both higher accuracy and lower intraobserver error than the subpubic concavity which in turn has both higher accuracy and lower intraobserver error than the ischiopubic ramus. An overall rank for the ventral arc which is at least two places higher than the subpubic concavity and an overall rank for the subpubic concavity which is at least two places higher than the ischiopubic ramus will confirm this hypothesis.

4. Although Meindl and colleagues (1985) suggest that the male pelvis is more frequently assessed incorrectly than is the female pelvis, it is suggested here that any bias in the morphological approach to sex determination of the pelvis will be in favour of males. This hypothesis will be accepted if there is a greater percentage of females than males whose sex was incorrectly assessed.

3.2.5 METHODS AND RESULTS

3.2.5.1 Data Collection

Most of the adult skeletal material excavated from the St. Thomas' cemetery was transported to McMaster University for storage during the period of analysis. The boxes containing the remains were numbered according to the sequence in which they were excavated and were stored in roughly the same order.

The assessment of sex by examination of the pelvis was undertaken at the beginning of the analysis period and assessment by examination of the skull was left until the end, in order to minimize intraobserver bias.

The pelves were studied individually; burials were not examined in consecutive order. Observations were taken in succession, as listed on the data sheet, for each pair of hipbones and any difference between the two sides were noted. Features were assessed as either "male", "female" or "indeterminate". Every effort was made to consider each trait separately so as not to bias the remaining observations. Due to time constraints, however, it was not practical to observe each pelvis individually for each trait in order to further reduce bias. Once all the observations were made, each criterion was given equal weighting and sex was assigned according to the category (male or female) into which the majority of features fell. In cases where the sum of results produced equal numbers of features identified as "male" and "female", sex was usually determined by giving Phenice's (1969) criteria more weight than the rest, due to their allegedly high collective level of accuracy.

Once all of the pelves had been analyzed, sixty-two individuals were re-examined for degree of intraobserver error. The retested cases consisted of all individuals which initial examination had indicated were borderline (indeterminate) and a random selection of the remaining pelves (including 23 personally identified individuals). Since difficult cases were deliberately chosen for inclusion into this study, the intraobserver error calculated herein may be higher than average.

3.2.5.2 Data Analysis

The first step in an analysis of this type is to determine the precision of the method being employed. It is necessary to know whether the observer can reliably reproduce her/his results. If s(he) cannot, the utility of the methodology is questionable and the technique should be reconsidered.

The problem of precision in this study was addressed via a test for intraobserver error. According to statisticians, researchers must decide for themselves the level of error they are willing to accept. In the past, Korey (cited in Molto, 1979) employed a 5% cutoff point for intraobserver error on discontinuous cranial traits. Features with error levels greater than 5% were considered too difficult to score and were removed from the analysis. More recently, Nichol and Turner (1986) chose a 10% critical level for dental characteristics. This analysis will follow the lead of Nichol and Turner, since a 10% critical level will identify criteria difficult to assess but still permits some room for error.

The precision test consisted of calculating the percentage of disagreements for final sex assignment between trial one and trial two in the sixty-two cases for which two sets of observations were taken. Seven differences occurred, resulting in 11.3% intraobserver error. This is slightly higher than the acceptable level and it suggests that there is a problem with the methodology. It is proposed here that the high level of intraobserver error is the result of the combined effects of only a few unreliable traits. Since each feature was given equal weight in the final estimation of sex, cases that contained roughly equal numbers of male and female criteria would require only a few inconsistently assessed traits for the outcome of the second assessment to disagree with that of the first.

The solution is to identify the features exhibiting high levels of intraobserver error and to eliminate them from the trait list. Thus, the reliability of each criterion was calculated. Table 5 lists the traits and gives a breakdown of the percentage of all possible types of disagreements. The proportion of difference is the percentage of discrepancies

TABLE 5. PERCENTAGE OF EACH TYPE OF DISAGREEMENT AND RANKING OF CRITERIA ACCORDING TO LEAST AMOUNT OF INTRAOBSERVER ERROR

TRAIT	*PROPORTION OF DIFFERENCE	+ERROR OF INDETERMINATE CRITERIA	@INTRAOBSERVER ERROR	RANK
Ventral Arc	11.2%	11.3%	0.0%	1
Muscle Markings	11.2%	29.0%	0.0%	1
Dorsal Pitting	12.9%	6.4%	0.0%	1
True Pelvis	22.5%	13.0%	0.0%	1
No. Sacral Segments	0.0%	4.8%	3.2%	5
Obturator foramen	9.6%	1.6%	3.2%	5
Posterior View of Sacrum	9.6%	0.0%	3.2%	5
Subpubic Concavity	14.5%	8.0%	3.2%	5
Pubis Shape	6.4%	14.4%	4.8%	9
Ilium Shape	11.2%	17.7%	6.4%	10
Sacrum Shape	11.3%	4.8%	6.4%	10
Sciatic Notch	0.0%	17.7%	6.5%	12
Pelvic Inlet	33.8%	1.6%	9.7%	13
Acetabulum	0.0%	6.4%	11.2%	14
Auricular Surface	1.6%	9.7%	11.3%	15
Preauricular Sulcus	1.6%	1.6%	11.3%	15
Ischiopubic Ramus	6.4%	9.6%	11.3%	15

N=62

* Refers to disagreements between trials that resulted from an attempt to assess the criterion in only one of the two instances (probably due to fragmented nature of the material).

+ Refers to disagreements between trials caused by assigning a sex during assessment one but labelling the second attempt "indeterminate".

@ Refers to the percentage of cases, for each criteria, that underwent a reversal in sex assignment between trial one and trial two.

caused by observing fragmentary material and assessing the trait only once in the two trials. Error of indeterminate criteria is the percentage of differences caused by assessing a criterion indeterminate for one trial and assigning that criterion a definite sex during the other trial. Intraobserver error is the percentage of cases, for a given criterion, that exhibit a reversal in sex assignment.

Although there are no extremely high levels of intraobserver error, there are four criteria which exceed the acceptable level. Acetabulum shape and size, auricular surface height (raised or flat), presence or absence of a pre-auricular sulcus and ischiopubic ramus shape were all found to be unreliable in this study. Since the level of intraobserver error for each of these features is only slightly higher than the critical level, the problematic criteria were included in the rest of the analysis. The implications of incorporating these criteria in a morphological sex estimation trait list are considered in the discussion.

Having established the reliability of the traits and the observer's ability to make use of them, the next step was to establish the accuracy of the method and the individual criteria involved. One means of determining the accuracy of the morphological approach was to compare the known and the skeletally determined sex of each of the forty-nine identified individuals. Only two of the estimated sex assignments were incorrect. The percentage of error is, therefore, 4.1%.

Thus, by employing the trait list described above, a 95.9% level of accuracy was achieved.

Accuracy was also examined by comparing the sex profile of the entire cemetery sample, as represented in the parish records, to the profile of the excavated skeletal sample. Before such an analysis could be undertaken, it was necessary to first determine which assignment of sex should be used in those cases where two evaluations were made. As indicated above, there were seven instances of differences occurring between trials one and two. The expectation was that experience, and consequently expertise, was gained by the time the second set of observations were made. Therefore, the second assignment should be correct. This view was tested by using twenty-three cases for which documented sex was available, and two sets of observations had been taken, to perform a crosstabulation of known sex by each trait and by estimated sex. SPSS PC+ version 2.0 was used to produce the results presented in Table 6 (see also Figure 9). There is no pattern which is common to all of the criteria examined. Some features, such as the pelvic inlet, show a decrease in error, indeterminate cases and unobservable traits while others, such as the subpubic concavity, show a decrease in unobservable and indeterminate observations at the expense of a slight increase in percentage of error. Other assorted combinations also occur. The following is a summary of the trial 2 data found in table 6:

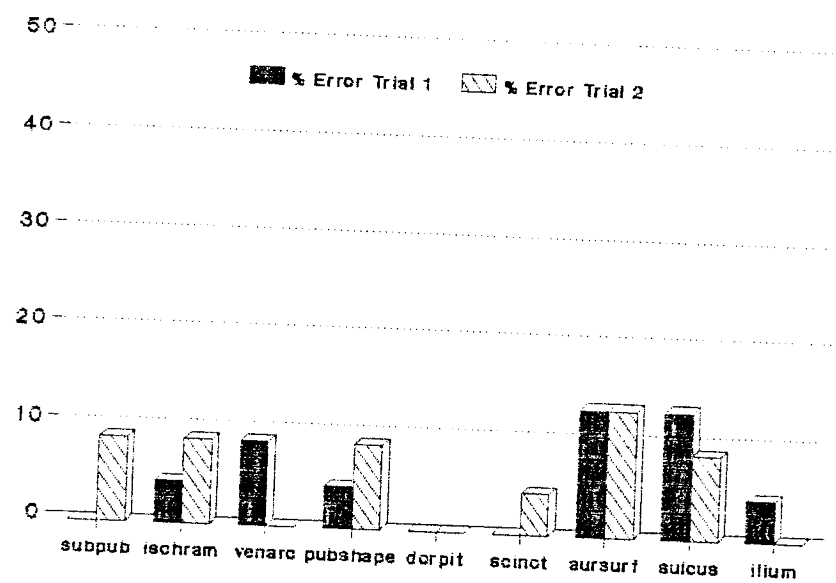
TABLE 6. COMPARISON OF %ERROR, % INDETERMINATE AND % DATA ABSENT BETWEEN OBSERVATIONS ONE AND TWO, CALCULATED FOR EACH PELVIC TRAIT

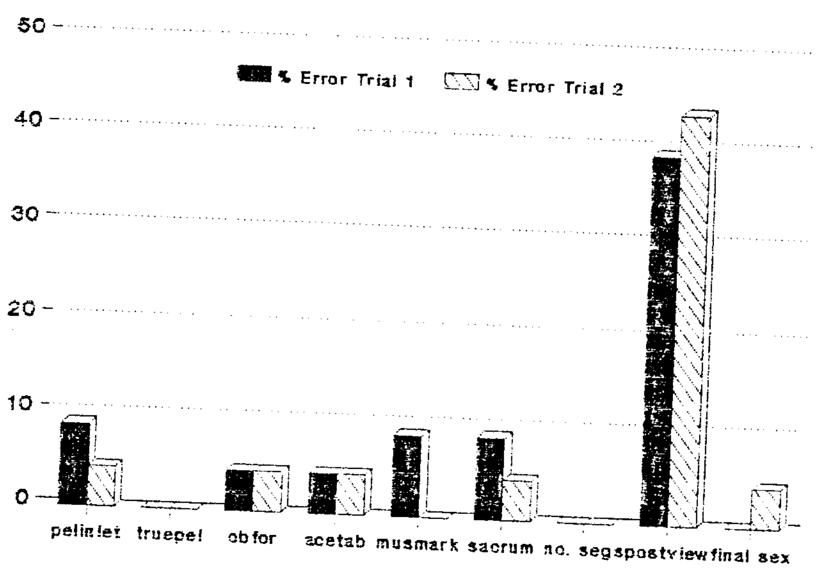
TRAIT	TRIAL NO.	% ERROR	% INDETERMINATE	% DATA ABSENT
subpubic concavity	1	0.0	4.3	47.8
	2	8.7	0.0	34.7
ischiopubic ramus	1	4.3	4.3	39.1
	2	8.7	8.7	34.7
ventral arc	1	8.7	4.3	30.4
	2	0.0	4.3	34.7
pubis shape	1	4.3	4.3	39.1
	2	8.7	4.3	43.4
dorsal pitting	1	0.0	34.7	47.8
	2	0.0	21.7	52.2
sciatic notch	1	0.0	21.7	0.0
	2	4.3	8.7	0.0
auricular surface	1	13.0	0.0	0.0
	2	13.0	13.0	0.0
preauricular sulcus	1	13.0	4.3	0.0
	2	8.7	0.0	0.0
ilium shape	1	4.3	17.4	13.0
	2	0.0	8.6	0.0
pelvic inlet	1	8.7	0.0	39.1
	2	4.3	0.0	4.3
true pelvis	1	0.0	0.0	43.5
	2	0.0	8.7	34.7
obturator foramen	1	4.3	0.0	47.8
	2	4.3	0.0	52.2
acetabulum	1	4.3	4.3	0.0
	2	4.3	0.0	0.0
muscle markings	1	8.6	26.1	13.0
	2	0.0	52.2	8.7
sacrum	1	8.7	0.0	34.7
	2	4.3	0.0	34.8
number of sacral segments	1	0.0	56.5	43.4
	2	0.0	56.5	43.4
posterior view of sacrum	1	39.1	0.0	17.3
	2	43.5	0.0	13.0

FIGURE 9.

Percentage of Error For Trials 1 and 2

First 9 Traits





Remaining Traits

and Final Sex Estimation

% error is reduced in 6 criteria
% error is increased in 5 criteria
no change in % error occurs in 6 criteria

% indeterminate is reduced in 6 criteria
% indeterminate is increased in 4 criteria
no change in % indeterminate occurs in 7 criteria

% data absent is reduced in 7 criteria
% data absent is increased in 5 criteria
no change in % data absent occurs in 5 criteria

It is possible to identify a general pattern in which trial 2 criteria exhibit reduced percentages of error, reduced indeterminate observations and reduced numbers of unobservable data. Since the final estimated sex is dependent upon the individual criteria, the evidence suggests that expertise and accuracy do increase with experience. Therefore, the second assignment of sex was employed in cases where a disagreement between the first and second set of observations occurred.

Next, SPSS PC+ version 2.0 was used to conduct a binomial test (for explanation see Appendix 2.) to compare the expected sex ratio, as determined by the parish records, to the sex ratio derived from a morphological examination of the excavated pelves. In order to perform this statistic, the expected probability of males to females for the entire cemetery sample had to be established. Unfortunately, the parish records did not always indicate the age of the individual and, consequently, the total number of adult males and females had to be expressed as ranges. The lowest number assumed that none of the age-unknown individuals were adult.

The highest assumed that all were adult. In order to encompass the full range of possibilities, two parish record frequencies were calculated. According to the register, there were 483-498 adult males and 343-349 adult females (D.O.A.M. series I, part 35A; Gerry Boyce, pers. comm.). This produced frequencies of 0.58-0.59 for males and 0.42-0.41 for females. The excavated sample consisted of 142 males and 107 females, resulting in frequencies of 0.57 and 0.43 respectively. Two binomial tests were run, the first using the frequency of 0.58 and the second using 0.59. The results indicate that there is no difference between the sex ratio of the cemetery population and the ratio of the excavated sample (p= 0.8053 and 0.5699, respectively).

The individual traits were also analyzed in order to determine the level of accuracy that each could achieve. Cross-tabulations of each feature by documented sex were performed using SPSS-PC+ version 2.0. Once again, for the cases that were assessed twice, the second trial was employed in the analysis because increased experience with the methods and material produced greater accuracy. The results of the crosstabulations, presented in Table 7, suggest that most of the pelvic criteria can be successfully employed to determine sex. In fact, 76% of the traits achieved levels of accuracy well above the level expected due to chance alone.

In an effort to assess the effects of age on accuracy, the known cases were divided into three age categories and the

TABLE 7. ACCURACY OF EACH PELVIC TRAIT

TRAIT	N	% CORRECT	% WRONG	% INDETERMINATE	RANK
Sacrum shape	34	94.1	2.9	3.0	1
Obturator foramen	32	93.8	6.2	0.0	2
Acetabulum	48	91.7	6.3	2.0	3
Preauricular sulcus	48	91.6	8.4	0.0	4
Ventral arc	38	86.9	0.0	13.1	5
Pubis shape	36	86.2	2.8	11.0	6
True pelvis	35	85.8	0.0	14.2	7
Sciatic notch	49	85.7	6.1	8.2	8
Subpubic concavity	37	83.8	5.4	10.8	9
Ilium shape	43	83.7	2.3	14.0	10
Ischiopubic ramus	40	80.0	5.0	15.0	11
Pelvic Inlet	45	80.0	6.6	13.4	11
Auricular Surface	49	73.5	14.2	12.3	13
Sacrum (post. view)	43	65.2	34.8	0.0	14
Muscle markings	44	56.8	2.3	40.9	15
Dorsal Pitting	28	35.7	0.0	64.3	16
# sacral segments	33	6.1	0.0	93.9	17
N=49					

TABLE 8. ACCURACY OF EACH TRAIT BY AGE CATEGORY

FEATURE	n	<25	n	25-44	n	45+
Subpubic concavity	7	85.8%	10	80.0%	16	81.3%
Ischiopubic ramus	7	57.2%	11	81.8%	18	83.3%
Ventral arc	7	71.5%	10	90.0%	17	94.1%
Pubis shape	6	83.3%	10	80.0%	16	87.5%
Dorsal pitting	6	50.0%	8	37.5%	11	36.4%
Sciatic notch	8	87.5%	11	81.9%	26	84.7%
Auricular surface	8	87.5%	11	63.7%	26	73.1%
Preauricular sulcus	7	85.8%	11	100.0%	26	88.5%
Ilium shape	8	62.5%	10	100.0%	21	85.7%
Pelvic inlet	8	75.0%	10	100.0%	23	73.9%
True pelvis	8	75.0%	7	100.0%	16	87.5%
Obturator foramen	7	100.0%	8	87.5%	14	100.0%
Acetabulum	8	87.5%	11	100.0%	25	88.0%
Muscle markings	7	42.9%	11	63.7%	22	58.9%
Sacrum	6	66.6%	9	100.0%	15	100.0%
No. sacral segments	6	16.7%	7	0.0%	16	6.3%
Posterior view of sacrum	8	25.0%	10	70.0%	21	71.5%*
ESTIMATED SEX	8	87.5%	11	100.0%	26	96.2%
N		N=8		N=11		N=26

* Statistically significant

cross-tabulations were rerun. It has been proposed that some features, such as the ventral arc, do not become distinctive until the third decade (Sutherland and Suchey, 1987), while others, such as the subpubic concavity, the preauricular sulcus and dorsal pitting lose their characteristic appearances with age (Tague, 1989; Kelley, 1979). The age categories, therefore, were created with these facts in mind to isolate the affects of age-related expressions of a trait on accuracy. The first grouping, which encompasses the growing period, consisted of adults younger than twenty-five years. By age twenty-five, the iliac crest has fused (Krogman and İşcan, 1986), the female ventral arc has formed (Sutherland and Suchey, 1987) and the superior border of the pubic symphysis is completing its growth (Tague, 1989).

The second division was comprised of individuals aged twenty-six to forty-four. This category encompasses those who had completed growth but had not yet experienced significant visual age-related bone degeneration.

The final group were forty-five years and older. By this point distinct degenerative changes can be seen on the pubis (Todd, 1920; Suchey et al., 1988) and auricular surface (Lovejoy et al., 1985) and obliteration of dorsal pitting and the preauricular sulcus has begun (Kelley, 1979).

Table 8 represents the results of an SPSS PC+ version 2.0 crosstabulation of known sex by each criterion, based upon the three age divisions outlined above. A chi-square test (p=0.273) indicates that there was no statistically significant difference between the levels of accuracy achieved in any of the three age categories, despite a perceived pattern which suggests that estimated sex is most accurate for the twenty-five to forty-four year category. It should be noted, however, that the sample sizes in this analysis were small. Perhaps a larger sample would produce statistically significant results.

Similarly, despite the apparent age-related increase in accuracy of four of the criteria, a Fisher's exact probability test indicates that this pattern is significant only in the case of the posterior view of the sacrum (p=0.031). The observed decrease in accuracy with age of the subpubic concavity and dorsal pitting features are likewise nonsignificant.

Accuracy and precision are equally as important, thus Table 9 represents the combined, overall ranking for the trait list. The significance of these results are discussed fully in Chapter 5.

Blocks of criteria were also analyzed for their collective effectiveness as sex indicators. Two by two tables, controlling for documented sex, of all possible combinations of criteria were performed using SPSS PC+. The probability of assigning the correct sex to a case, based upon the combined results of two criteria was determined in the following manner. If both traits indicated male, the sex

TABLE 9. THE EFFECTIVENESS RANKING OF THE PELVIC TRAITS

TRAIT	ACCURACY RANK (Table 7)	PRECISION RANK (Table 5)	TOTAL	OVERALL RANK
Ventral arc	5	1	6	1
Obturator foramen	2	5	7	2
True pelvis	7	1	8	3
Sacrum shape	1	10	11	4
Subpubic concavity	9	5	14	5
Pubis shape	6	9	15	6
Muscle markings	15	1	16	7
Dorsal pitting	16	1	17	8
Acetabulum	3	14	17	8
Preauricular sulcus	4	15	19	10
Sacrum (post. view)	14	5	19	10
Sciatic notch	8	12	20	12
Ilium shape	10	10	20	12
# sacral segments	17	5	22	14
Pelvic inlet	11	13	24	15
Ischiopubic ramus	11	15	26	16
Auricular surface	13	15	28	17

estimate was male. Similarly, if both traits indicated female, the sex estimate was female. If one indicated male and one was indeterminate, the sex estimate was male, and so on. The probability of achieving a correct sex assignment using this method was calculated by dividing the number of correct sex estimates (when compared to documented sex) by the total number of estimates.

For example, when the ventral arc and the true pelvis are used in combination, the probability of estimating "male" when the individual was in fact male was 23/23 or 1.0 (P[m/male]=23/23=1.0). The probability of estimating "female" when the actual sex was female was 9/10 or 0.90 (P[f/female]=9/10=0.90). The probability of making no decision (perhaps one criterion suggested "male" the other "female" or both were "indeterminate") when the individual was actually female was 1/10 or 0.10 (P[?/female]=1/10=0.10). Therefore, the probability of a correct answer was the probability of estimating male when the case was male plus the probability of estimating female when the case was female, divided by two ({P[m/male] + P[f/female]} / 2). In this example, probability of making a correct estimate was 0.95 ([1.0 + 0.90] /2 =0.95) and the probability of making no decision was 0.05 ([0.10+0] /2). If errors occur, the probability of error is calculated in the same manner as the probability of correct or indeterminate cases.

Since the original method of analysis (each criterion having equal weight with the majority decision defining the sex) produced an accuracy level of 95.9%, combinations which generate probabilities under 0.95 are less effective than the original technique and will not, therefore, be presented. Table 10 lists the probabilities equal to or greater than 0.95.

Table 10 demonstrates that some criteria were more useful than others and that they reappeared in various combinations. In an effort to increase the probability of obtaining a correct answer, combinations of three were considered. Only those features which produced good results in the combinations of two, were chosen for incorporation into the combinations of three.

None of these combinations managed to exceed the highest probability obtained using the combinations of two (0.98 for Obturator foramen/Ventral arc). Table 11 indicates the combinations of three which produced the highest probabilities of achieving a correct sex estimation.

One last attempt to increase the probability of estimating sex correctly was made. Table 11 indicates that only two combinations of three criteria produced probabilities of 0.96. They were: pubis shape/ subpubic concavity/ acetabulum and pubis shape/ ventral arc/ acetabulum. It was decided to combine these four characteristics and to run an SPSS-PC+ v2.0 crosstabulation of the combination controlling for documented sex. The result was a probability of 0.96 of

TABLE 10. PROBABILITY OF ESTIMATING SEX CORRECTLY

COMBINATION	PROBABILITY CORRECT	PROBABILITY NO DECISION	PROBABILITY ERROR
Obturator foramen/Ventral arc	p=.98	p=.02	p=0.0
Obturator foramen/True pelvis	p=.98	p=.02	p=0.0
Pubis shape/Acetabulum shape	p=.96	p=0.0	p=.04
Ventral arc/True pelvis	p=.95	p=.05	p=0.0
Obturator foramen/ilium shape	p=.95	p=.05	p=0.0
Sacrum shape/Subpubic concavity	p=.95	p=0.0	p=.05
True pelvis/Subpubic concavity	p=.95	p=0.0	p=.05
Sacrum shape/Acetabulum shape	p=.95	p=0.0	p=.05
Sacrum shape/Ventral arc	p=.95	p=0.0	p=.05
Sacrum shape/Pubis shape	p=.95	p=0.0	p=.05
Acetabulum shape/True pelvis	p=.95	p=0.0	p=.05

TABLE 11. PROBABILITY OF ESTIMATING SEX CORRECTLY COMBINATIONS OF THREE CRITERIA

COMBINATION	PROBABILITY CORRECT	PROBABILITY NO DECISION	PROBABILITY ERROR
Pubshape/Subpub/Acetabulum	p=.96	p=0.0	p=.04
Pubshape/Venarc/Acetabulum	p=.96	p=0.0	p=.04
Obfor/Subpub/True pelvis	p=.95	p=.05	p=0.0
Obfor/Venarc/Acetabulum	p=.95	p=.05	p=0.0
Obfor/Venarc/Pubis shape	p=.95	p=.05	p=0.0
Venarc/Subpub/Sacrum shape	p=.95	p=0.0	p=.05
Venarc/Subpub/True pelvis	p=.95	p=0.0	p=.05
Pubshape/Subpub/Sacrum shape	p=.95	p=0.0	p=.05
Acetab/Subpub/Sacrum shape	p=.95	p=0.0	p=.05
Truepel/Subpub/Sacrum shape	p=.95	p=0.0	p=.05
Pubshape/Subpub/True pelvis	p=.95	p=0.0	p=.05
Acetab/Subpub/True pelvis	p=.95	p=0.0	p=.05
Acetab/Venarc/Sacrum shape	p=.95	p=0.0	p=.05
Pubshape/Venarc/Sacrum shape	p=.95	p=0.0	p=.05
Pubshape/Acetab/Sacrum shape	p=.95	p=0.0	p=.05

estimating sex correctly with a 0.04 probability of making an error.

All analyses of the Phenice (1969) criteria were performed in conjunction with those described above. The issues surrounding the ventral arc, subpubic concavity and ischiopubic ramus raised in the introduction to this chapter are addressed in Chapter 5.

Similarly, the direction of error was assessed through an examination of the incorrect sex assignments in conjunction with the analyses performed above. Discussion of the results can be found in Chapter 5.

3.3 GROUP SEX: EMPLOYING THE BONES OF THE SKULL TO ASSESS THE SEX OF HUMAN SKELETAL REMAINS

3.3.1 INTRODUCTION

As Keen (1950: 65) observed, determining the sex of a single skull from a forensic case may prove to be difficult. Evidence of sex from the skull is indirect and as a result, sex-related features display a wide range of variation both within and between populations (Acsádi and Nemeskeri, 1970; Keen, 1950; Krogman and Īşcan, 1986; St. Hoyme and Īşcan, 1989; Steele and Bramblett, 1988; Ubelaker, 1984; Workshop of European Anthropologists, 1980). Since many of the cranial features employed to determine sex are associated with general size and muscularity, establishing the sex of an individual skull is exceedingly problematic unless the specimen exhibits extreme versions of the characteristics (Keen, 1950).

A researcher attempting to determine sex from a group of skulls is in a much better position than a researcher trying to determine the sex of an individual skull, since (s)he would be able to observe a large number of individuals and consequently, would have a better idea of the expected range of variation.

Since size is an important factor in sex determination from the skull, there will be a certain amount of overlap between male and female expressions of a feature in any population and, therefore, there will be a small proportion of cases for which sex cannot accurately be ascertained (Stewart, 1952). When one is employing a morphological trait list, the indeterminate cases will likely exhibit some features associated with males and some with females. Under such circumstances, it is necessary to know upon which features one should rely.

Unfortunately, as indicated previously, weightings developed in the past appear to have been a matter of personal preference. According to St. Hoyme and Īşcan, the distributions of many cranial criteria have never been reported for individuals of known sex or race and the differences which we have come to rely upon, "may simply reflect our stereotypes" (St. Hoyme and Īşcan, 1989: 71).

This study contributes to our understanding of cranial morphological sex characteristics by addressing the following issues:

(1) What levels of accuracy and reliability can be achieved using a comprehensive list of cranial features?

(2) Which criterion/criteria is/are the most accurate and reliable?

(3) Does this method of assessing sex produce a bias, and if so, in which direction is the error?

3.3.2 THE TRAIT LIST

The features chosen for examination in this study were compiled from trait lists produced by many physical anthropologists including: Bass, 1987; El-Najjar and McWilliams, 1979; Hrdlicka (Stewart, 1952); Keen, 1950; Krogman and Īşcan, 1986; Meindl et al., 1985; Stewart, 1979. Seventeen traits representing both the calvarium and the face were chosen. A copy of the data sheet employed is presented in Table 12 (see also Figure 10). Features were scored as either "male", "female" or "indeterminate".

3.3.3 LITERATURE REVIEW

As indicated above, most of the physical anthropologists whose recommendations were taken into consideration during the development of the cranial trait list, have variously ranked the features presented here. These rankings are displayed in Table 13. By examining the literature on craniofacial growth, it should be possible to predict the most accurate feature(s)

TABLE 12. VISUAL SEX DETERMINATION OF THE ADULT SKULL

GROUP/CASE: ______________ DATE: ______________
BURIAL : ______________ LOCATION: ______________
CONDITION : ______________
OBSERVER : ______________
SEX : ______________

	MALE		FEMALE	
GENERAL SIZE ARCHITECTURE	large & rugged	____	small and smooth	____
FOREHEAD	steeper, less round	____	rounded, full	____
FRONTAL EMINENCES	small	____	large	____
SUPRAORBITAL RIDGES	med - large	____	small - medium	____
ORBITS	squared, lower, smaller rounded margins	____	rounded, higher, larger, sharp margins	____
NASAL APERTURE	higher, narrower sharp margins	____	lower, wider, rounded margins	____
NASALS	larger, form sharp angle at midline	____	smaller, wider angle at juncture	____
MALARS	heavier, more laterally arched, muscle marked	____	light, more compressed smooth	____
ZYGOMATIC	posterior end extends as crest past ext. aud. meatus	____	posterior end does not extend as far	____
PARIETAL EMINENCES	small	____	large	____
MASTOID	medium to large	____	small to medium	____
OCCIPITAL AREA	marked muscle lines	____	protuberances not marked	____
OCCIPITAL CONDYLES	large	____	small	____
PALATE	larger, broader U-shaped	____	small, parabolic shape	____
TOOTH SIZE	larger, M1 5 cusp	____	smaller	____
MANDIBLE	larger, higher symphysis broader ascending ramus, gonial angle <125, gonial angle flares	____	small, lower corporal and ramal dimensions, gonial angle >125, gonial angle does not flare	____
CHIN FORM	square	____	rounded	____

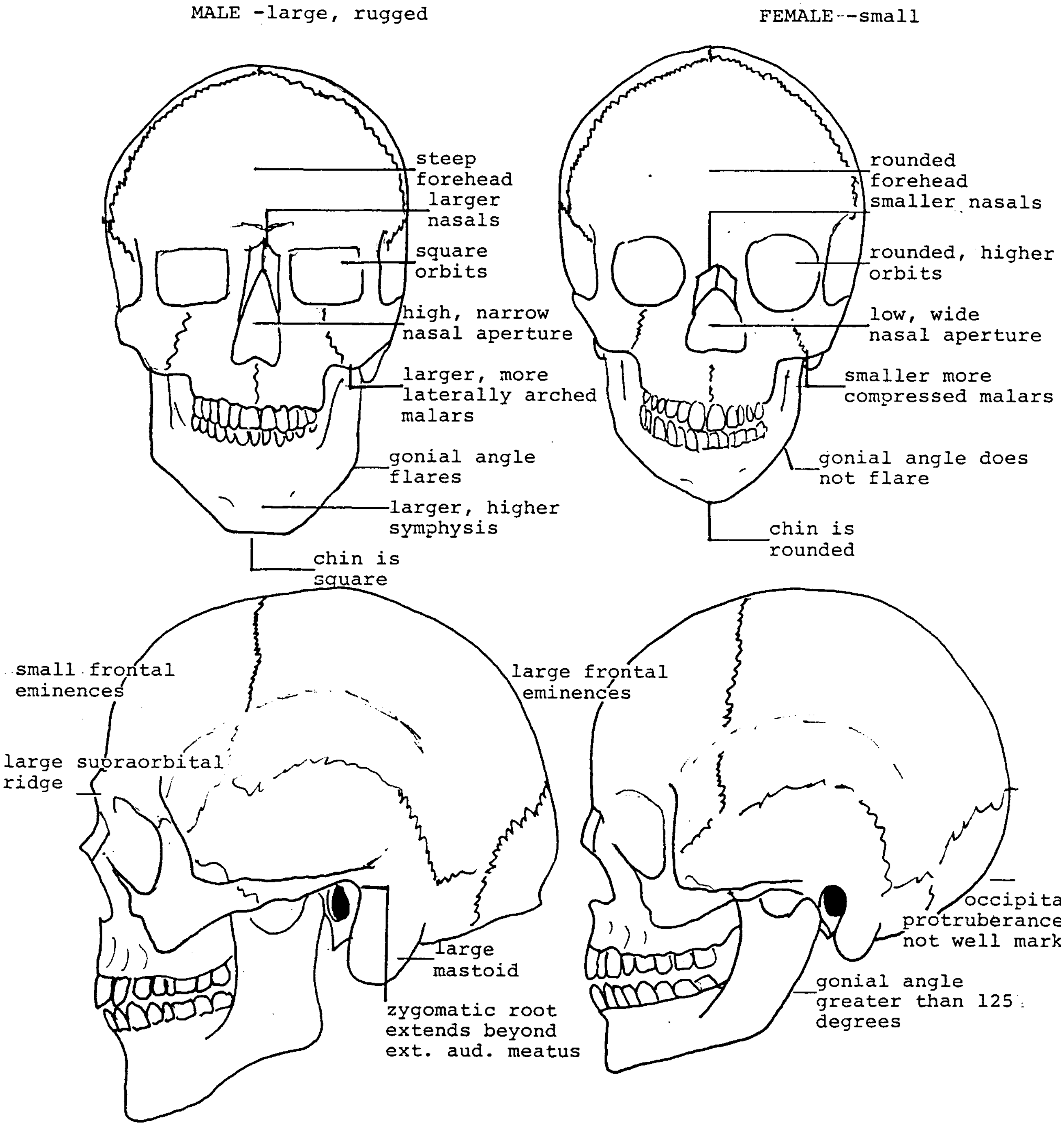

MALE -large, rugged
FEMALE--small
steep forehead
larger nasals
square orbits
high, narrow nasal aperture
larger, more laterally arched malars
gonial angle flares
larger, higher symphysis
chin is square
rounded forehead
smaller nasals
rounded, higher orbits
low, wide nasal aperture
smaller more compressed malars
gonial angle does not flare
chin is rounded
small frontal eminences
large supraorbital ridge
large mastoid
zygomatic root extends beyond ext. aud. meatus
large frontal eminences
occipita protruberance not well mark
gonial angle greater than 125 degrees

TABLE 13. RANKING OF SKULL SEX CRITERIA ACCORDING TO FIVE PHYSICAL ANTHROPOLOGISTS

RANK	KROGMAN AND ĪŞCAN (1986)	RANK	MEINDL ET AL. (1985)
1.	General size, architecture	1.	overall rugosity of occipital, mastoid, zygomatic
2.	supraorbital ridges	2.	relative depth of mandibular rami, morphology of chin, degree of gonial eversion
3.	mastoid size	3.	size of teeth and palate
4.	rugosity of occipital region	4.	size of pterygoid plates
5.	size and shape of malars	5.	supraorbital ridges
6.	orbit shape and size	6.	frontal profile
7.	mandible shape		
8.	palate shape and size		

RANK	HRDLICKA (1952)	EL-NAJJAR AND MCWILLIAMS (1979)	STEWART (1979)
1.	general size	1. supraorbital ridges	1. orbital borders
2.	supraorbital ridges	2. orbital borders	2. supraorbital ridges
3.	mastoid size	3. height of mandibular body	3. mastoid size
4.	rugosity of occipital and temporal region		4. rugosity of occipital region
5.	zygomatic arch size		5. malar shape and size
6.	malar shape and size		6. chin shape
7.	alveolar arch and teeth size		
8.	mandible shape		
9.	palate shape and size		
10.	base of skull		
11.	facial physiognomy		

for determining sex from the skull and to test both this prediction and the above rankings. According to many researchers, sexual dimorphism of the skull is visible largely as differences in the size of various features (Acsádi and Nemeskeri, 1970; Bass, 1987; Brothwell, 1981; El-Najjar and McWiliams, 1979; Krogman and İşcan, 1986; St. Hoyme and İşcan, 1989; Steele and Bramblett, 1988; Stewart, 1979; Ubelaker, 1984; Workshop of European Anthropologists, 1980). Most of these size differences result from the differential length in the growing period of males and females.

Infant facial features, regardless of sex, can be characterized as follows: nose is short, rounded, pug-like; nasal bridge is low; nasal profile is concave; forehead is bulbous and upright; cheekbones are prominent; face is flat and eyes are wide set (Enlow, 1983). As the face begins to grow, changes in dimensions and proportions occur. The longer the period of growth, the more pronounced these changes will be. Female facial features experience a reduction in growth rate beginning around the thirteenth year of life; maturation is completed soon afterward. At the same time, males enter a growth spurt which continues through adolescence with maturation completed in early adulthood (Enlow, 1983; Enlow, 1975; Goldstein, 1936; Nanda, 1955; St. Hoyme and İşcan, 1989). The difference in the duration of the growth period is a general phenomenon, which varies slightly between individuals. Thus, larger features are generally characteristic of males, but a certain amount of overlap is inevitable. One might, therefore, expect to see significant sexual differentiation only in features which must undergo a considerable increase in size and which do so over a long period of time, thereby allowing the effects of the male growth spurt to come into play.

In order to generate some hypotheses concerning the trait list employed in this study, it is necessary to examine the factors which are responsible for the final adult configuration of each feature. The duration of growth and the developmental patterns of the individual characteristics are outlined below.

The initial impression of the sex of a particular skull is usually based upon general size and architecture. A large, rugged cranium suggests the individual is male, while a small, gracile cranium suggests female. In an effort to determine whether sexual dimorphism in cranial size actually occurs, and to establish at what point it first becomes measurable, Baughan and Demirjian (1978) examined the cranial length and width and the stature of 73 girls aged 6 to 15 years and 47 boys aged 10 to 18 years. They discovered that even prior to puberty girls have a smaller cranium, both absolutely and in relation to stature, than boys of the same age. In addition, the authors report a spurt in cranial growth for boys, which was not evident in girls, and they indicate that this spurt contributes to adult sexual dimorphism (Baughan and Demirjian, 1978). Buschang et al. (1983) examined 26 males and 25 females between 4 and 16 years of age and they also found that males exhibit greater relative growth than females.

Whether the skull continues to increase in size throughout one's lifetime is still not clear. Goldstein (1936) reports an overall decrease with age in the dimensions he examined. On the other hand, Baer (1956) found a significant increase in male facial dimensions in the third decade but found no significant increase in either male neurocranial or female craniofacial dimensions. Tallgren (1974) examined 32 Finnish women aged 20-73 years over a period of 15 years and found no evidence for significant changes in measurements of the cranial vault and cranial base. Yet Israel (1981; 1973) discovered continued, although non-uniform, expansion of the craniofacial skeleton throughout adult life. His research on 26 males and 26 females between the ages of 24.9 and 78.8 years indicates that sella turcica, frontal sinus and the skull tables remodel at a rate twice that of anterior-posterior skull diameter and that this pattern is evident in both males and females (Israel, 1981).

Therefore, the data support the use of general size and architecture to establish a first impression, however, the literature also suggests that one must take age into consideration when employing these features. If the cranium continues to grow throughout adult life, then an older female might approximate a younger male in general size, and the unwary observer could misclassify the individual.

The size of the frontal eminences and supraorbital ridges and the shape of the forehead are all dependent upon the same process for their final adult form. According to Enlow (1983), growth of the inner table of the frontal ceases at age five or six, as growth of the frontal lobe is completed. Yet, the outer table is part of the nasomaxillary complex and as such, continues to remodel outward until nasal growth stops some years later. Since the nasal area of the female completes its growth several years before that of the male, there is less separation between the inner and outer tables (ibid). Greater separation of these tables in the male, results in a larger frontal sinus as well as greater protuberance of the supraorbital ridges. In addition, the anterolateral surface of the orbital roof is resorptive while the cutaneous surface of the supraorbital ridge is depository. These factors combine to create a protrusive superior orbital rim (Enlow, 1975). The forehead is also part of the nasomaxillary complex, which is displaced downward and forward from the calvarium during growth (ibid) Consequently, the male forehead changes from the bulbous, upright infantile shape to a steeper and less rounded form (Enlow, 1983).

Growth in the nasomaxillary region lasts longer in males than in females, but growth of the inner table is completed at approximately the same age. Therefore, the complex of features; frontal eminence and supraorbital ridge size and forehead shape, should provide good skeletal evidence for the sex of the individual.

Enlow (1975) notes that the nasal and jaw regions grow faster and to a greater extent than the orbits and soft tissues and as a result, adult eyes appear smaller than infant eyes, in proportion to the rest of the face. Since male features grow for a longer period than female features, the female orbits appear to be larger and to be located higher in the face than the male orbits. Additionally, the enlarged male supraorbital ridges decrease orbital height slightly (St. Hoyme and Īşcan, 1989), adding to the impression that male orbits are squared, situated lower in the face and smaller than those of the female.

The characteristic size and shape of the orbits are subject to the configuration of the rest of the skull. They appear to be larger or smaller depending upon the length of the face, the degree of forehead sloping and the size of the supraorbital ridges. Rather than being tangible features, they are merely impressions. As such, orbital size and shape may be useful in typical cases but will probably not be helpful in borderline individuals.

The orbit does not require a large increase in size in order to accommodate the growing eye and its surrounding tissues (Enlow, 1975). In contrast, the maxilla undergoes significant relative growth (Buschang et al., 1983a; Buschang et al., 1983b; Enlow, 1975; Goldstein, 1936). Thus, as the maxilla is displaced downward and the orbital floor (part of the maxilla) drifts with it, the orbit becomes unnecessarily large. In order to compensate for this drift and to maintain the proper orbital size, the orbital floor deposits bone (Enlow, 1975). Consequently, the floor of the nasal aperture, originally very close to the floor of the orbit, is almost twice the distance from the orbital floor by the time growth is completed (ibid). Since the duration of growth in this region is greater for males, the nasal aperture will be longer, yet the simultaneous forward projection of the nasal region, will cause the aperture to be situated higher on the male face than on the female face.

Nasal aperture form is, therefore, the result of downward and forward growth of the nasomaxillary region. Since a significant amount of growth occurs in this area and the duration of growth is longer in males, this feature should be fairly useful in skeletal sex determination.

Sexual dimorphism in nasal shape and form are also the result of size differences. The nasals are usually larger in males due to the extended duration of craniofacial growth. Nasal form results from the pattern of interaction of two structures adjacent to the nasal bones. The malar region becomes relocated posteriorly while the adjacent nasal region of the maxilla enlarges anteriorly, resulting in a protrusive nose (Enlow, 1975). Since the nasals increase in length but grow very little in width (ibid), they must gradually form a sharper angle in the midline in order to compensate for the divergent growth in the surrounding regions.

In order for this feature to be reliable, growth occurring in the malar and maxillary regions must consistently interact in such a way that pressure, created by the diverging patterns of growth, causes the nasals to bend in the midline. The shape of the nasal bones is, therefore, dependent upon two different systems in order to become manifest and, for this reason, nasal shape will probably not be the most useful criterion.

As indicated above, the malar becomes relocated posteriorly during growth. The zygomatic arch moves laterally by resorption on the medial side, within the temporal fossa, and by deposition on the lateral surface (Enlow, 1975). Thus the temporal fossa enlarges while the malar remains proportionately broad in relation to the face, jaw size, and masticatory musculature (ibid). Due to the extended growth of the male facial skeleton, the malars are larger and the zygomatic arches displaced more laterally than the corresponding structures in females. Dependent almost solely on differences in the duration of the male and female growth period, malar shape will probably be useful in all cases except those that are borderline.

Keen (1950) observed a difference between the form of the posterior root of the zygomatic of males and females. In males, the root is continuous with the supramastoid crest, which then becomes part of the temporal line. He indicates that this feature is dependent upon the masticatory apparatus, specifically, the development of the temporalis muscle (ibid). St. Hoyme and Īşcan (1989) suggest that this feature is a reflection of greater male robusticity, which can be seen elsewhere in the skeleton. This feature will only be reliable if the sample under examination exhibits sexually dimorphic robusticity. In typically gracile or robust populations, sexual differences of the posterior root of the zygomatic will not be significant.

The parietal eminence is the initial site of ossification for the parietal bone. Examination of the infant skull reveals that the parietal is bowed outward, the eminence being the most lateral point on the bone. Enlow (1975) noted that as the brain expands, the bones of the calvarium are displaced outwardly, and the arc of curvature of the whole bone decreases. This process accounts for the larger female parietal eminence. The male calvarium continues to grow after female growth is completed (Baughan and Demirjian, 1978), consequently, it takes on a flatter appearance and the parietal eminence becomes less marked than in females. Given that some overlap will occur between the length of the growing periods of individual males and females, one can also expect overlap between male and female parietal eminence size. Thus, this feature will likely be useful only in extreme cases.

Mastoid and occipital condyle size are reflections of extended male growth. They are continuous variables which are recorded simply as small, medium or large and are, therefore, highly subjective criteria. If, however, a large sample is being examined, one would have a fairly good idea of the expected range of variation. Consequently, these features would be much more useful. Since the St. Thomas' sample is quite extensive, both mastoid and occipital size should provide significant contributions to cranial sex determination.

Heavily marked occipital muscular attachments are, once again, a reflection of both personal muscular development and general robusticity. This criterion should be as effective as other indices of robusticity.

The "V" principle, described by Enlow (1975), best explains the enlargement of the human palate. Many facial bones, or parts of bones, have a roughly v-shaped form. During expansion, bone deposition occurs on the inner surface of the "V", while resorption occurs on the outer. Movement is toward the wide end of the "V". Thus, both displacement and enlargement occur. Such is the case with the palate. As it enlarges, it also widens (ibid). Growth in the nasomaxillary region continues several years longer in males than in females (Enlow, 1983). Thus, the male palate is both larger and broader. Due to the extended male growing period, palate shape and size should be quite useful for cranial sex determination.

Although many researchers mention tooth size as a possible source of sexual dimorphism (Acsádi and Nemeskeri, 1970; Hrdlicka, in Stewart, 1952; Krogman and Īşcan, 1986; Workshop of European Anthrcpologists, 1980), most seem to agree with St. Hoyme and Īşcan that, "the range of overlap in size, shape, and color [of teeth] is such as to make these features of dubious value for establishing race or sex" (1989: 72).

The direction of growth of the mandible is complex.

> The young child's mandible appears to be pointed. This is because it is wide, short, and more "V"-shaped. In the adult, the entire lower jaw becomes "squared". With the development of the chin, together with massive growth in the lateral areas of the trihedral eminence, eruption of the permanent dentition, lateral enlargement of each ramus, expansion of the masticatory musculature, and flaring of the gonial regions, the whole lower face takes on a "U"-shaped configuration, resulting in a considerably more full appearance...In the infant and young child, the gonial region lies well inside (medial to) the cheekbone. In the adult, the posteroinferior corner of the mandible extends laterally out to the cheekbone, or nearly so. This gives the posterior part of the jaw a square appearance.
>
> (Enlow, 1975: 6)

Enlow (1975) also notes that the ramus becomes progressively more upright during mandibular development. This accounts for the typical male gonial angle of less than 125 degrees. Clearly, the characteristic male mandible; larger, higher symphysis, broader ascending ramus and flaring gonial angle, is simply the result of continued male growth. Buschang et al. (1983b) indicate that the greatest relative growth of the face occurs in the mandible. This fact, combined with the knowledge that male craniofacial growth is not complete until some years after female craniofacial growth, suggests that the mandible, including chin form, will be exceptionally useful for sex determination.

In summary, a review of the literature concerning craniofacial growth has led to a number of predictions regarding the accuracy of the cranial traits employed in this study. Unlike the pelvic criteria, the cranial features did not fall into easily distinguishable categories. The suggested ranking is as follows:

General size and architecture should be used in order to gain an initial impression. Some indication of the expected degree of dimorphism visible in other features might be gained from this initial examination.

MOST EFFECTIVE Mandible chin form

SECOND MOST EFFECTIVE

Forehead, frontal eminences and supraorbital ridges should be considered a single point of evidence.
Palate

Measures of robusticity should be given consideration only if the range of variation for the group can be established and the evidence suggests that the entire sample does not fall into one of the two extreme categories. If a cursory examination warrants it, pay special attention to the zygomatic root, occipital and temporal muscle markings, mastoid and occipital condyle size.

THIRD MOST EFFECTIVE

Nasal aperture Malars

FOURTH MOST EFFECTIVE

Orbits Parietal eminences

LEAST EFFECTIVE

Nasals Teeth

The studies used to generate these predictions suggest that the inherent direction of error for cranial sex criteria favours females. The evidence supports the position that sexual dimorphism of the skull is the result of the longer period of male craniofacial growth (Enlow, 1983; Enlow, 1975; Goldstein, 1936; Nanda, 1955; St. Hoyme and Īşcan, 1989). Thus, one could expect borderline individuals to be classified "female", simply because they had not attained the typical male size and appearance. This prediction is supported by Meindl et al. (1985), who observe that females were rarely misclassified while males were occasionally mistaken for females, but it contradicts the findings of Weiss (1972). He claims that the larger/smaller type of sex characteristics of the skull produce, "an irresistible temptation in many cases to call doubtful specimens male" (Weiss, 1972: 240), thereby producing a bias in favour of males. Meindl and colleagues also note that female skulls appear more male-like with age and they conclude that while young males might be mistaken for females, older females might also be mistaken for males (Meindl et al., 1985).

3.3.4 HYPOTHESES

It was possible to develop three main hypotheses based upon the cranial growth literature. They are as follows:

1. The cranial morphological approach to skeletal sex determination can accurately identify the sex of an individual

and can produce accurate sex profiles of a sample. This hypothesis will be accepted if the method has an accuracy greater or equal to 90% and if the sex ratio produced from the skeletal estimates is statistically indistinguishable from the ratio developed from the parish records.

2. The features of the face are better indicators than the features of the calvarium . In order to accept this hypothesis the facial criteria must exhibit both a mean accuracy which is at least 10% higher and a mean intraobserver error which is at least 5% lower than the features of the calvarium

3. Any bias in the morphological method of cranial sex determination will favour females. This hypothesis will be accepted if there a greater percentage of females than males were misclassified.

3.3.5 METHODS AND RESULTS

3.3.5.1 Data Collection

Cranial sex determination was conducted toward the end of the four month period during which data from the St. Thomas' site were collected. In total, 245 skulls were sufficiently intact to permit some attempt at sex determination.

The crania and mandibles were examined in groups of approximately 80-90. As the skulls were brought into the lab by assistants, the researcher divided them into two groups based solely on general size and architecture. Once the initial eighty skulls were separated in this manner, the three largest and three smallest were examined in detail. Each of the six skulls exhibited extreme versions of almost every characteristic under consideration. Thus the expected range of variation for the sample was established. These six skulls remained available for comparison throughout the examination period and were used as the standard by which subjective assessments, such as "large", "medium" or "small", were made.

Each skull was then considered individually. The criteria were observed in the order in which they appear on the trait list. Borderline features were compared to the six extreme cases in an effort to achieve clarification. In some instances, this process was unsuccessful and the feature was left as "indeterminate".

After all the observations had been made, each criterion was given equal weighting and sex was assigned according to the category (male or female) into which the majority of features fell. As assessments were completed, the skulls were placed on the researcher's left if they had been designated female, and on the right, if male. Before each group was returned and the next was brought in, the skulls were once again compared to one another, on the basis of general size and physiognomy, to determine whether they appeared to belong in the division into which they had been placed. If doubt was raised, the skull was reassessed.

Once all cf the skulls had been analyzed, forty-nine individuals were re-examined for degree of intraobserver error. The retested cases consisted of some indeterminate individuals and a random selection of the remaining skulls.

3.3.5.2 Data Analysis

Precision was addressed via a test for intraobserver error. The forty-nine re-examined cases were assessed for differences in overall sex assignments (male and female) between trials. There were six disagreements, resulting in 12.2% error. This level of intraobserver error is unacceptably high. It was previously suggested that one or two unreliable criteria could adversely affect the level of error of the entire method. Therefore, the reliability of each feature was calculated. Table 14 lists the traits and provides a breakdown of the percentages of all possible forms of disagreements.

Although there are no extremely high levels of intra-observer error, there are two features that exceed the acceptable level of 10%. Both nasal aperture size, shape and placement and orbit size, shape and border sharpness were found to be unreliable in this study. Since they only surpass the critical value by a few percentage points, these problem-atic criteria were included in the rest of the analysis. The implication of incorporating these features into the trait list may be significant, therefore, it will be discussed in Chapter 5.

TABLE 14. PERCENTAGE OF EACH TYPE OF DISAGREEMENT AND RANKING OF CRITERIA ACCORDING TO THE LEAST AMOUNT OF INTRAOBSERVER ERROR

TRAIT	* PROPORTION OF DIFFERENCE	+ ERROR OF INDETERMINATE CRITERIA	@INTRA-OBSERVER ERROR	RANK
Occipital condyles	28.5%	24.5%	0.0%	1
Tooth size	36.7%	18.3%	0.0%	1
General size	42.8%	40.8%	0.0%	1
Mastoids	44.9%	44.9%	0.0%	1
Supraorbital ridges	26.4%	20.4%	2.0%	5
Parietal eminences	28.5%	26.5%	2.0%	5
Occipital markings	42.8%	34.7%	2.0%	5
Chin form	18.3%	14.2%	4.1%	8
Mandible	30.6%	24.5%	4.1%	8
Palate	30.6%	24.5%	4.1%	8
Malars	30.7%	18.4%	4.1%	8
Frontal eminences	34.6%	28.5%	4.1%	8
Forehead	42.8%	32.6%	4.1%	8
Zygomatics	30.6%	24.5%	6.1%	14
Nasals	32.6%	26.5%	6.1%	14
Nasal aperture	20.3%	4.0%	10.3%	16
Orbits	28.5%	14.3%	12.2%	17

N=49

* Refers to disagreements between trials that resulted from an attempt to assess the criterion in only one of the two instances (probably due to fragmented nature of the material).

+ Refers to disagreements between trials caused by assigning a sex during one assessment but labelling the second attempt "indeterminate".

@ Refers to the percentage of cases, for each criterion, that underwent a reversal in sex assignment between trial one and trial two.

Having established the levels of reliability for each trait, the next step was to establish the accuracy of using morphological features of the cranium to determine sex. One means of assessing accuracy involved a comparison between the documented and estimated sex of the forty-six personally known individuals for whom cranial material was available. Five of the forty-six sex estimates were incorrect. The percentage of error was thus 10.9%. By employing only the cranial traits described above, an 89.1% level of accuracy was achieved.

Accuracy was also examined by comparing the sex profile of the entire cemetery sample to the profile of the excavated sample. Before this analysis was undertaken, it was necessary to first determine which assignment of sex should be employed for the six undocumented cases in which two sets of observations were made, and in which a discrepancy occurred. It was believed that greater expertise would be gained by the time the second observations were made, and therefore, the second assignment should be employed. This prediction was tested by crosstabulating known sex by each trait and known sex by estimated sex for those cases in which documented sex was available. SPSS PC+ version 2.0 was used to produce the results presented in Table 15 and Figure 11.

As the data for final estimated sex indicate, trial 2 results were generally better than those for trial one. Ten characteristics exhibited higher accuracy for trial two; only four, for trial 1; and in three cases, the levels of accuracy

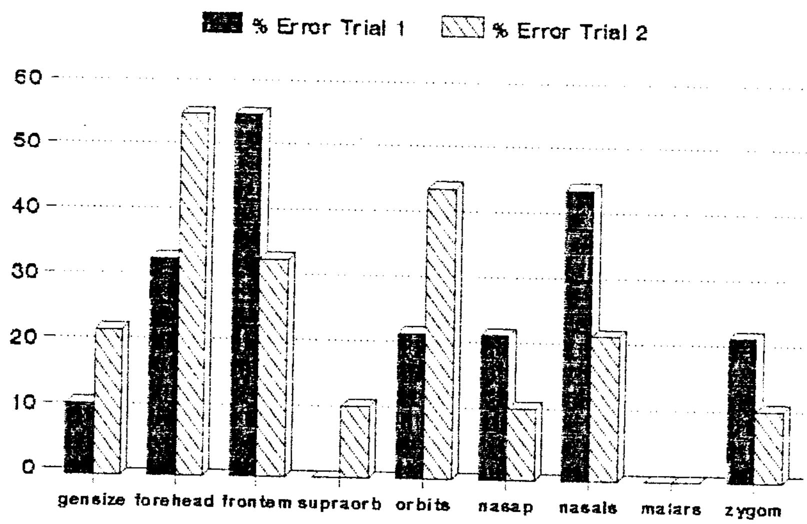
TABLE 15. COMPARISON OF % ERROR, % INDETERMINATE AND % DATA ABSENT BETWEEN THE FIRST AND SECOND SET OF OBSERVATIONS, CALCULATED FOR EACH CRANIAL TRAIT (N=9)

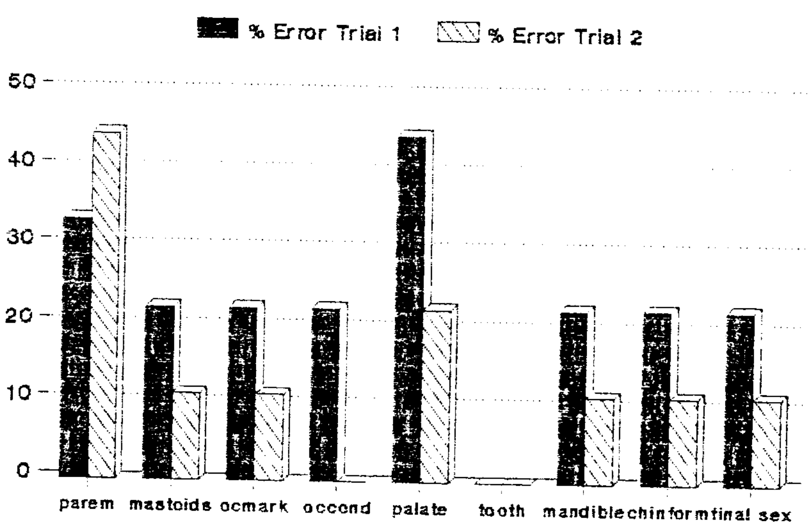
TRAIT	TRIAL NO.	% RIGHT	% ERROR	% INDETERMINATE	% DATA ABSENT
general size	1	22.2	11.1	66.7	0
	2	33.3	22.3	44.4	0
forehead	1	33.3	33.3	33.4	0
	2	22.2	55.6	22.2	0
frontal eminence	1	22.2	55.6	22.2	0
	2	33.3	33.3	33.4	0
supraorbital ridges	1	66.7	0.0	33.3	0
	2	77.8	11.1	11.1	0
orbits	1	55.5	22.3	22.2	0
	2	33.3	44.5	22.2	0
nasal aperture	1	66.7	22.2	11.1	0
	2	77.7	11.1	0.0	11.1
nasals	1	33.3	44.5	22.2	0
	2	55.6	22.2	22.2	0
malars	1	66.7	0.0	33.3	0
	2	88.9	0.0	11.1	0
zygomatics	1	66.7	22.2	11.1	0
	2	77.8	11.1	11.1	0
parietal eminences	1	0.0	33.3	66.7	0
	2	0.0	44.4	55.6	0
mastoids	1	11.1	22.2	66.7	0
	2	22.2	11.1	66.7	0
occipital markings	1	55.6	22.2	11.1	11.1
	2	44.4	11.1	33.3	11.1
occipital condyles	1	11.1	22.2	66.7	0
	2	11.1	0.0	88.9	0
palate	1	44.4	44.4	11.1	0
	2	55.6	22.2	22.2	0
tooth size	1	11.1	0.0	44.4	44.4
	2	0.0	0.0	77.8	22.2
mandible	1	55.6	22.2	22.2	0
	2	55.6	11.1	33.3	0
chin form	1	33.3	22.3	44.4	0
	2	55.6	11.1	33.3	0
final estimated sex	1	77.8	22.2	0.0	0
	2	88.9	11.1	0.0	0

FIGURE 11.

Percentage of Error For Trials 1 and 2

First 9 Traits





Remaining Traits

and Final Sex Estimation

remained the same. Thus it would appear that expertise was gained by the time the retests were made. Therefore, in cases where discrepancies occurred, the second assessment was employed.

Having established the validity of using the results of trial 2, SPSS PC+ version 2.0 was employed to conduct a binomial test (see Appendix 2. for description). As was the case for the assessment of the pelvic bones, two parish record frequencies were calculated in order to account for the possibility that some of the individuals whose ages were unrecorded were adults.

According to the register, there were 483-498 adult males and 343-349 adult females (D.O.A.M. series I, part 35A; Gerry Boyce, pers. comm.). This produced frequencies of 0.58-0.59 for males and 0.42-0.41 for females. Using cranial sex assignment only, the excavated sample consisted of 135 males and 108 females, resulting in frequencies of 0.56 and 0.44, respectively (compared to 0.57 and 0.43 for the pelvis). Two binomial tests were run, the first using the documented frequency of 0.58 and the second using 0.59. The results indicate ($p=0.4795$ and $p=0.3047$, respectively) that there is no significant difference between the documented sex ratio of the cemetery sample and the ratio established for the excavated sample.

The individual traits were also examined for their ability to reflect the sex of the individual. Cross-tabulations of each feature by documented sex were performed using SPSS PC + version 2.0. Once again, trial two results were employed whenever a discrepancy between test one and test two results occurred. Table 16 contains the outcome of the cross-tabulations. These results indicate that, individually, many of the cranial sex-related traits produce levels of accuracy only slightly better than what can be expected due to chance alone. In fact, 53% were less accurate than pure chance. The implication of these findings are discussed in Chapter 5.

Meindl et al. (1985) suggest that sex-dependent features of the skull are significantly affected by age. In an effort to assess the effects of age on accuracy, the known cases were divided into three age categories and the cross-tabulations were rerun. The first division consisted of individuals less than twenty-five years of age. Since male craniofacial growth is not completed until early adulthood (Enlow, 1983; Enlow, 1975), this category encompasses the growing period. The next division was comprised of individuals aged twenty-five to forty-four, and the last group were those aged forty-five and older. Table 17 illustrates the accuracy of each trait by age category.

Once again, despite the apparent age associated trends, a chi square test reveals no significant difference between the levels of accuracy achieved in the three age categories ($p=0.432$). In addition, a Fisher's exact probability test

TABLE 16. ACCURACY OF EACH CRANIAL TRAIT

TRAIT	N	% CORRECT	% WRONG	% INDETERMINATE
Nasal aperture	34	76.6	5.8	17.6
Zygomatics	47	70.3	4.2	25.6
Malars	38	68.4	5.2	26.4
Supraorbital ridges	46	60.9	4.3	34.8
Chin form	48	56.3	10.4	33.3
Occipital markings	45	53.3	13.4	33.3
Nasals	36	52.8	33.3	13.9
Mandible	47	51.1	8.4	40.5
Mastoids	47	44.7	6.4	48.9
Forehead	45	44.5	33.3	22.2
Orbits	39	43.6	30.7	25.7
General size	42	38.0	7.2	54.8
Palate	41	36.6	19.5	43.9
Frontal eminences	47	31.9	31.9	36.2
Parietal eminences	45	28.9	31.3	40.0
Occipital condyles	43	14.0	4.6	81.4
Tooth size	39	10.3	0.0	89.7

TABLE 17. ACCURACY OF EACH TRAIT BY AGE CATEGORY

FEATURE	n	<25	n	25-44	n	45+
General size	7	28.6%	9	44.4%	25	40.0%
Forehead	8	62.5%	9	33.3%	27	44.4%
Frontal eminences	8	49.0%	10	30.0%	28	25.0%
Supraorbital ridges	8	50.0%	9	77.8%	28	60.7%
Orbits	7	71.4%	9	44.4%	22	36.3%
Nasal aperture	7	85.7%	8	100.0%	18	66.6%
Nasals	7	57.2%	8	62.5%	20	50.0%
Malars	6	50.0%	9	77.8%	22	72.8%
Zygomatics	8	50.0%	10	80.0%	27	70.3%*
Parietal eminences	8	37.5%	9	11.1%	26	30.7%
Mastoids	7	42.9%	10	30.0%	28	50.0%
Occipital markings	6	16.7%	9	66.7%	28	60.7%*
Occipital condyles	7	14.3%	9	22.2%	25	8.0%
Palate	8	12.5%	9	44.4%	23	39.1%
Tooth size	7	14.3%	10	10.0%	20	10.0%
Mandible	8	50.0%	10	70.0%	27	44.4%
Chin form	8	87.5%	10	80.0%	28	35.7%

* Statistically Significant

indicates that only two individual criteria, the zygomatics and occipital markings, exhibit age-related patterns of accuracy. In both cases, accuracy increased with age. This pattern is contrary to that suggested by Meindl et al. (1985).

Since the accuracy and the precision of a trait are equally important, Table 18 illustrates the combined ranking for each criterion. The significance of these findings is fully discussed in Chapter 5.

Blocks of criteria were also assessed for their collective effectiveness as skeletal sex indicators. Two by two tables, controlling for documented sex, of all possible combinations of criteria were performed using SPSS PC+.

The probability, for each combination, of estimating sex correctly was calculated using the procedure previously described for the pelvis. The original method of analysis, in which each criterion was given equal weight, produced an accuracy of 89.1%. Therefore, the combination technique must produce probabilities greater than 0.89 in order to be considered an improvement. Unfortunately, the combination producing the best results, zygomatics and malars, achieved a probability of only 0.88.

In an effort to increase the probability of obtaining a correct answer, combinations of three were considered. Only those criteria which produced the best results in the combinations of two were chosen for incorporation into the combinations of three. Only one combination, zygomatics/ malars/

TABLE 18. THE EFFECTIVENESS RANKING OF THE CRANIAL TRAITS

TRAIT	ACCURACY RANK (Table 16)	PRECISION RANK (Table 14)	TOTAL	OVERALL RANK
Supraorbital ridges	4	5	9	1
Mastoids	9	1	10	2
Malars	3	8	11	3
Occipital markings	6	5	11	3
Chin form	5	8	13	5
General size	12	1	13	5
Zygomatics	2	14	16	7
Mandible	8	8	16	7
Nasal aperture	1	16	17	9
Occipital condyles	16	1	17	9
Forehead	10	8	18	11
Tooth size	17	1	18	11
Parietal eminences	15	5	20	13
Nasals	7	14	21	14
Palate	13	8	21	14
Frontal eminences	14	8	22	16
Orbits	11	17	28	17

nasal aperture, improved upon the original method. The probability of achieving an accurate estimation using these three criteria was 0.91. All other combinations of three resulted in probabilities below 0.88.

The direction of error was assessed through an examin ation of the incorrect sex assignments in conjunction with the analyses performed above. Discussion of the results can be found in Chapter 5.

3.4 CONCORDANCE OF PELVIC AND CRANIAL TRAITS

Both pelvic and cranial morphological sex criteria were available for two hundred and thirty-nine individuals. Disagreements between the sex estimate produced from pelvic observations and that produced from cranial observations occur in thirty-one cases. Concordance of pelvic and cranial sex estimations is 87%. Only four of the thirty-one disagreeing cases were personally identified individuals. In all four cases, the pelvic estimate was correct while the cranial estimate was not.

CHAPTER FOUR
4.0 AGE ESTIMATION

4.1 INTRODUCTION

Ever since Todd (1920) first demonstrated that the pubic symphyseal face undergoes regular age-related changes, analysis of the pubic symphysis has been the principal method of skeletal age estimation in both paleodemography and forensic anthropology in North America (Meindl and Lovejoy, 1989). In fact, despite numerous warnings (Brooks, 1955: 588; Brooks and Suchey, 1990: 237; Todd, 1920: 314), researchers have frequently relied upon only the pubic symphysis to estimate skeletal age-at-death (Meindl and Lovejoy, 1989). Meindl and Lovejoy (1989) attribute the trust that is placed in this method to the clarity and distinctiveness of some of its early age-related changes, and to the poor reviews that other methods have received when put to the test. For example, Brooks (1955), found that cranial suture closure was an extremely unreliable age indicator, whereas the pubic symphysis technique produced better results. Similarly, McKern (1957) recommended the use of epiphyseal and suture closure, for individuals over 17 years, only when the pubic symphysis was damaged or missing.

In contrast to these earlier studies, Meindl and colleagues (1983), assessed five methods of estimating age-at-death and discovered that the highest correlations between actual and estimated age were generated by summary age (weighted average of the results of all of the methods; 0.83), clinical age (a subjective seriation of the cases by summary age; 0.82), auricular surface (0.72), dental wear (0.70), cranial sutures (0.65), pubic symphysis (a modified version of Todd's [1920] study; 0.57) and proximal femur (0.53). These results suggest that more attention be directed toward other techniques, and that the pubic symphysis method is in need of further improvements.

Recently, Suchey and Brooks (Suchey et al., 1988) and Suchey, Brooks and Katz (Suchey and Katz, 1986) have developed an approach to pubic symphyseal age estimation (one for females and one for males) which they believe is superior to all previous methods.

The objective of the present analysis was to assess the accuracy of the Suchey-Brooks (Suchey et al., 1988) and Suchey-Katz (1986) methods of estimating age-at-death from the pubic symphysis and, similarly, the accuracy of the Lovejoy et al. (1985) auricular surface age estimation technique. The levels of accuracy, inaccuracy and bias were calculated by using the mean of each age category (as provided by the original authors). Both sets of data were then compared to their Gilbert-McKern (1973) and McKern-Stewart (1957) counterparts in order to establish whether the Suchey and colleagues' methods are, in fact, superior to other approaches.

Each phase in the pubic methods is represented by a range rather than a single age. Because these ranges rarely conform to the tidy blocks of ages, e.g. decades or five year blocks, that are used in sample age profiles, the phase ranges must be converted in order to fit the predefined categories of the age profile. There are many methods of making this conversion (see Jackes, 1985). In this study, the most straightforward method of conversion, which uses the mean age of each phase, was employed. Jackes (1985) system of pubic symphysis age distribution was also used. This approach involves casting all the cases within a phase over its 95% range (2 standard deviations) by assuming that the ages are normally distributed and calculating the probability of obtaining each age within the range (see Jackes [1985] for methodology). Both types of distributions were then compared to the documented age profile.

The levels of accuracy, inaccuracy and bias of the auricular surface technique were compared to those of the pubic symphyseal approaches.

4.2 LITERATURE REVIEW

4.2.1 PUBIC SYMPHYSEAL AGE ESTIMATION

The pubic symphyseal age estimation method has undergone considerable modification since Todd (1920) first introduced his system of analysis over seventy years ago. An in-depth treatment of the modifications and the tests of these modifications is beyond the scope of this analysis, thus the reader is referred to Jackes (in press) and Krogman and İşcan (1986) for greater detail.

Brooks (1955) conducted the earliest test of Todd's approach. She found that only 54% of the males and 31% of the females in her sample could be correctly assessed. Her primary concern was the tendency of the method to overage individuals (Brooks, 1955: 588). More recent tests of the Todd technique confirm this finding (Katz and Suchey, 1986; Meindl et al., 1983; Suchey et al., 1986), although Meindl and colleagues (1983) also noted that there was a tendency to underage in the fifth and sixth decades. Another criticism of Todd's method was the exclusion of a number of symphyses from the analysis because they did not conform to his ideal patterns, thus the total range of normal variation was greatly reduced (Jackes, 1985; Suchey et al., 1986). In addition, some of the individuals in Todd's sample were of unknown age and, consequently, estimates based upon anatomical examinations made by Todd or one of his colleagues were used in place of actual documented age, for comparison to skeletally estimated age-at-death (Suchey et al., 1986).

McKern and Stewart (1957) felt that Todd's technique was too static; that it was only useful on pubic symphyses which conformed to Todd's concept of "typical". Therefore, they developed a new approach to pubic symphyseal age estimation which they believed was more flexible and consequently, capable of accounting for a greater amount of individual variability. McKern (1957) reported that the method produced better results than age estimates made from epiphyseal and suture closure in individuals between 17 and 25 years of age. The correlation ratio of actual and estimated age was 0.86.

Bocquet-Appel and Masset (1982) argue that the McKern-Stewart method produces an age distribution very similar to that of the original reference population (Korean War dead). Jackes (1985) agreed with these findings noting that it was possible to identify the age estimation technique employed by an anthropologist simply by the age profile which is produced. The strong correlation between estimated and actual age achieved by McKern (1957) can be explained by the fact that his test sample was derived from his original reference population of Korean War dead. An additional problem with the McKern-Stewart method, noted by Meindl and Lovejoy (1989), is that the ages of the individuals in the original sample span only a few decades (17-50 years but only seven individuals were between 40 and 50 years of age). The highest age category is 36+ with a mean of only 41 years (McKern and Stewart, 1957). Thus, individuals older than fifty years will be significantly underaged. While this presents a problem for anyone attempting to estimate age, it is particularly vexing to researchers who wish to use the age estimates to produce life tables, since key features of the table, life expectancy for example, are dependent upon the oldest attained age.

Gilbert (1973) demonstrated that female age from pubes would be underestimated if the McKern-Stewart (1957) technique, developed for males, was employed to estimate female age-at-death, since pubic maturation occurs, on average, ten years later in females. Thus the Gilbert and McKern (1973) method for estimating female age-at-death was introduced. Unfortunately, Suchey (1979) found a considerable amount of variation in the stage assignments for this method made by twenty-three physical anthropologists on the same eleven female pubes. Only 51% of all of the age assessments made were correct, despite the broad age categories which are part of this technique. Most researchers had difficulties determining whether the ventral rampart was being built up or broken down.

Hanihara and Suzuki (1978) developed a method of estimating pubic age through multiple regression analysis. While their technique is better than other component systems in the 20-40 year range (Meindl and Lovejoy, 1989), it is not recommended for use on individuals over 40 years of age (Hanihara and Suzuki, 1978).

Nemeskeri and colleagues (Acsádi and Nemeskeri, 1970) also developed a method of estimating age-at-death from the pubic symphysis. Studies indicate that this approach assesses too many individuals at between 45 and 60 years of age

(Jackes, 1985). In addition, it focuses primarily on the very early and the very late changes making it possible to determine only whether a case is under 50, about 50 or over 50 years (Brooks and Suchey, 1990).

The most recent improvements are those of Suchey and Brooks (Suchey et al., 1988) for females and Suchey and Katz (1986) for males. Both methods examine the total pattern of change and reject the component analysis approach (Brooks and Suchey, 1990). The advantages of these techniques include the large sample sizes used to develop the method (females = 273 and males = 739); the fact that the documented ages-at-death of the individuals included in the original study cover most of the adult life span (14-99 years of age); and the variety of populations from which the cases were drawn (White, Black, Mexican, Oriental and Other). To date, no evaluations of these two methods, employing samples independent of the original reference group, have been made.

4.2.2 AURICULAR SURFACE METHOD OF AGE ESTIMATION

A relatively new approach, the auricular surface method of estimating age-at-death, was formally introduced by Lovejoy and his colleagues in 1985. Both this system and an earlier form of the technique, developed by Kobayashi (1967), were applied to Portuguese Mesolithic samples by Jackes (in press) and were found unsatisfactory due to the poor condition of the auricular surfaces and the high degree of inter-observer disagreement. Yet, in the original tests of the method, before it was fully developed, correlations with stated age ranged from 0.55 to 0.75 (Lovejoy et al., 1985). Meindl et al. (1983) were able to achieve a correlation of 0.72 between actual and estimated ages using the fully developed technique, while Bedford et al. (1989) produced slightly less impressive results with a correlation of 0.60 between actual and skeletally-derived age. Sokal and Rohlf (1981) have suggested that the coefficient of determination (r^2), which ranges from zero to one, is a better measure of the relationship between variables than is the correlation coefficient, particularly when one is comparing the relative importance of correlations of different magnitudes. This being the case, Lovejoy and colleague's age assessments exhibit a coefficient of determination ranging from 0.30 to 0.56; Meindl and colleagues achieved $r^2=0.52$ and Bedford and colleagues attained $r^2=0.36$.

The results of Murray and Murray (1991) were even more pessimistic. These researchers examined 189 individuals from the Terry collection and assigned only 44% correctly or within one category of actual age. They concluded that the auricular surface method overestimates the actual age of younger adults and underestimates the age of older adults, with a crossover between 35 and 40 years of age (Murray and Murray, 1991).

Because the method developed by Lovejoy et al. (1985) made use of the Hamann-Todd collection of skeletal material, there was some question concerning the "known" or actual ages of many of the individuals. It was suggested that some of the actual ages were simply estimates made by Todd, or his colleagues, following anatomical examinations of the questionable individuals (Brooks, pers. comm.). The same claim had been made by Katz and Suchey (1989) concerning a sample employed by Meindl et al. (1985) to develop and test a formal multifactorial method of age estimation, which included the auricular surface. Recently Meindl and colleagues stated that, "all of the ages used in our subsample [to develop and test the multifactorial method] were in fact the legal age at death recorded on (and copied from) the United States Revised Death Certificate filed at the Vital Records Division of Cleveland City Hall for each specimen used in the study" (Meindl et al., 1990: 350). Unfortunately, it was not made clear whether the five hundred specimens from the Todd Collection used to develop the auricular surface technique, specifically, were also represented by ages recorded on death certificates. The implications of using anatomical estimates in place of actual documented age will be discussed in Chapter 5.

4.3 METHODS AND RESULTS

4.3.1 DATA COLLECTION

Age estimations were conducted separately for the pubic symphysis methods and the auricular surface technique so that observations made using one approach would not bias those of the other. The Suchey-Katz and McKern-Stewart methods (for males) and the Suchey-Brooks and Gilbert-McKern methods (for females) were employed one after the other for each case. Due to time constraints, it was impractical to assess the pubic symphyses of 174 individuals on two separate occasions. Furthermore, in this study the focus is on the Suchey methods; the McKern-Stewart and Gilbert-McKern systems are used only as standards for comparison. In addition, the Suchey methods employ a total pattern approach rather than the component analysis used by the other techniques. Therefore, providing that the Suchey method was performed first, it was believed that no significant bias would result.

One hundred and fourteen males were examined following the procedure of Suchey and Katz (1986). Both casts and descriptions, representing the six phases, are employed in this system which requires the observer to focus upon overall patterns of change. Similarly, casts and descriptions were available for the McKern and Stewart (1957) method which focuses upon three components of the pubic symphysis: the ventral rampart, the dorsal demi-face and the symphyseal rim. Sixty females were analyzed following Suchey, Brooks and Katz (1988) which incorporates the use of both casts and descriptions in its total pattern approach. The Gilbert and McKern (1973) method, analogous to that of McKern and Stewart (1957), is a component system which employs both casts and descriptions.

Each set of pubes was examined and, after the data were recorded, grouped according to its Suchey-Katz or Suchey-Brooks phase (males and females were kept separate). Once all of the cases had been analyzed, attempts were made to seriate the individuals within each group, but the wide range of variation within each phase rendered the attempts useless. Comparisons were made, both within and between divisions, to ensure that all members within a phase appeared to be at relatively the same stage of development (or degeneration). If the right and left sides of a set of symphyses from one individual appeared to belong to different phases, both were noted but the younger one was chosen to represent the estimated age on the assumption that disease or pathology could cause the surface to appear prematurely old.

Once all of the symphyses had been examined, a random selection of 35 males and 25 females were re-examined for the purpose of conducting an intra-observer error test.

Two hundred and thirty-eight hip bones were analyzed using the Lovejoy et al. (1985) auricular surface method of age estimation. Both the descriptions (Lovejoy et al., 1985) and the photographs provided by Bedford et al. (1989) were employed. This approach focuses upon age-related changes in granulation, macroporosity, billowing, striations and transverse organization of the auricular surface of the ilium. If two auricular surfaces from one person seemed to be of different ages, the younger was chosen because disease and pathology can cause the surface to appear prematurely old (Rothschild, pers. comm.). Similarly, whenever a surface appeared borderline between two age categories, the younger one was chosen.

After each set of auricular surfaces was examined, it was grouped by age category. Once the individual analysis was completed, the surfaces were compared both within and between categories to ensure that those in the same age divisions presented similar appearances. Finally, following the recommendation of Lovejoy et al. (1985), attempts to seriate the surfaces were made. The observer eventually gave up in despair due to the complexities of such an endeavour. Far too much subtle variation was present to make seriation possible without more procedural information from the authors of the method (Lovejoy et al., 1985).

4.3.2 DATA ANALYSIS

Precision was addressed via a test for intra-observer error. Age estimates using the Suchey-Katz method were conducted twice for 35 males. Only 23 assessments agree; of the 12 that do not, 9 are within one age category and 3 differ by two or more categories. Thus, the intra-observer error is 34%. This can be compared to 39% intra-observer error (13 disagreements out of 33 retests) for the McKern-Stewart method.

Results for the Suchey-Brooks method are even poorer. Of the 25 females assessed twice, 14 disagreements occur between observation 1 and observation 2, resulting in 56% intra-observer error. Eight of the disagreements are within one age category while six differ by two or more categories. This can be compared to 52% intra-observer error (13 disagreements out of 25 retests) for the Gilbert-McKern system.

The auricular surface method produced an intra-observer error of 19.3%. All 238 surfaces were examined twice. Forty-six disagreements occur between the two sets of data. Most of the error (15.2%) is comprised of cases which disagree by only one age category (most of which span 5 years). Thus, only 4.2% of the disagreements differ by two or more age categories.

The accuracy of each method was assessed through a variety of approaches. The known subsample was employed to determine the percentage of accurate age estimates. The levels of inaccuracy and bias were computed using the mean of each age phase (as determined for the original reference population). An age profile of the entire sample was produced from the parish register data. This was compared to the profiles generated from each of the age estimation techniques. It should be noted that there were fifty-five individuals aged 17 and over in the known subsample. However, due to differential preservation, each method is represented by some fraction of this total.

Before calculating the accuracy, inaccuracy and bias values, it was necessary to first establish which set of observations to employ when cases examined twice (for the intraobserver error study) produced different results.

Nine documented age-at-death cases, were examined twice using the Suchey-Katz method. Six of these retested cases produced estimated age ranges which included the actual age for both observation 1 and observation 2. In two cases, trial 2 produced the correct age estimate while trial 1 did not. There was one case in which neither trial 1 nor trial 2 produced correct estimates. Although the data are scant, they suggest slightly better results for the second set of observations.

Nine documented age-at-death cases were examined twice using the Suchey-Brooks approach. Both sets of observations produced estimated age ranges which encompassed the documented age.

The auricular surface method of age estimation was employed twice on twenty cases of known age-at-death. In seventeen cases the age ranges were identical. In two cases trial 2 produced a correct age estimate while trial 1 did not. In one case trial 1 produced an age estimate one category younger than the actual age and trial 2 produced an estimate one category older than the actual age. Once again, although the sample size is small, the data suggest that the second set of observations are more accurate than the first.

Given the above results, in all tests of accuracy, inaccuracy and bias, whenever two sets of disparate observations were available for the same specimen, the second set of data was employed in the calculations. Although the McKern-Stewart and Gilbert-McKern methods were not tested in this manner (since they are used for comparison only), the second set of observations was also employed for these methods, in order to achieve a more accurate comparison to the Suchey-Katz and Suchey-Brooks techniques.

The accuracy of each system of skeletal age estimation was calculated by dividing the number cases in which the estimated age range encompassed the actual documented age by the total number of cases (of known age) upon which the technique was employed. The results are presented in Table 19. Figures 12-14 illustrate the success of the Suchey-Katz (1986), Suchey-Brooks (Suchey et al., 1988) and the auricular surface methods. The rectangles represent the age ranges associated with each phase while the individual squares represent the individual cases. Cases which were estimated correctly are indicated by squares within the rectangles. The diamonds in the graphs of the two pubic symphysis methods represent the mean age for each phase.

Lovejoy and colleagues (1985) argue that the correlation coefficient is not necessarily the best indicator of the power of an age estimator. They claim that average error (inaccuracy) and bias are more appropriate measures. "Inaccuracy is the average absolute error of age estimation

TABLE 19. ACCURACY OF EACH METHOD OF ESTIMATING AGE-AT-DEATH FROM THE PELVIS

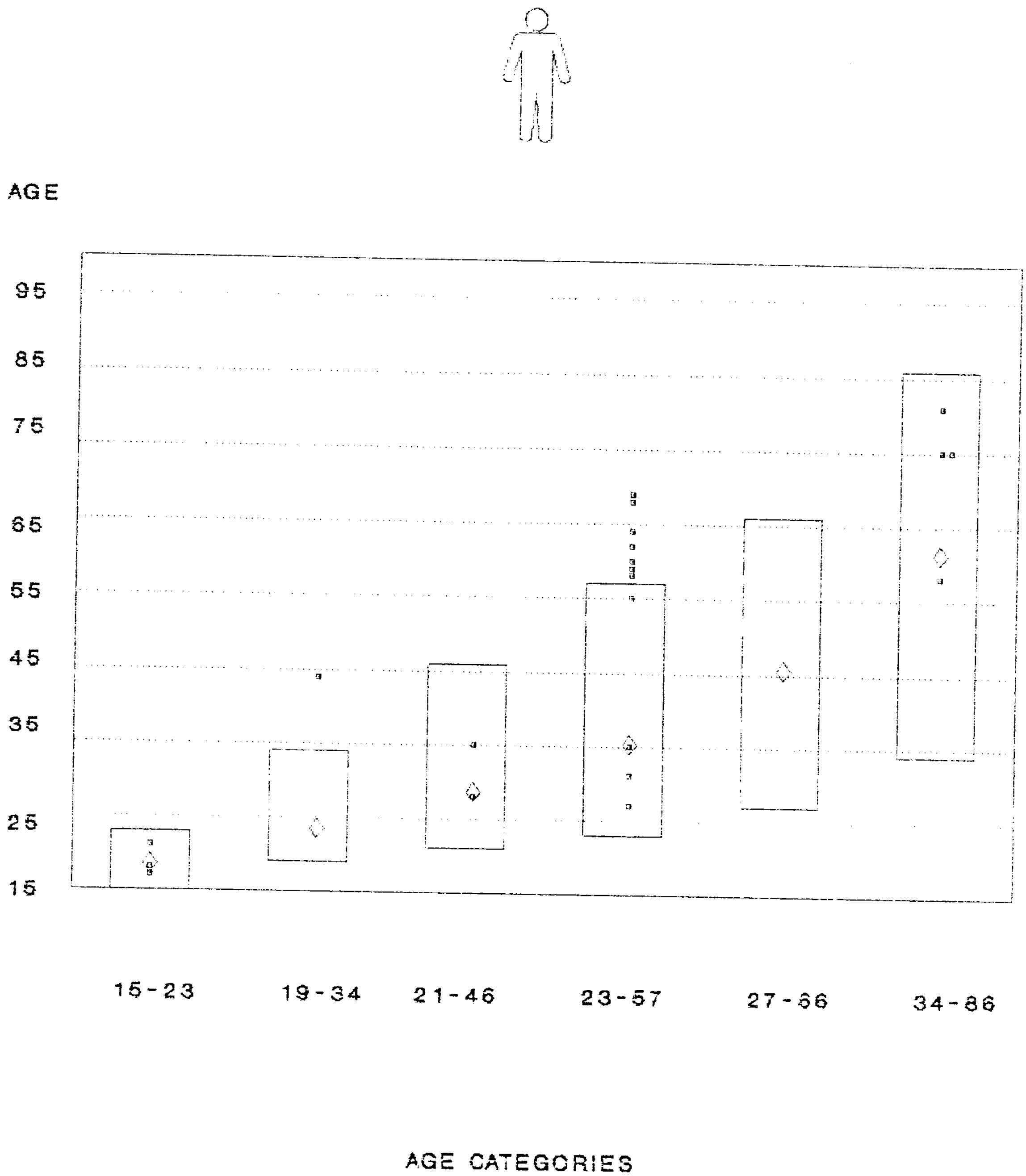
METHOD	NO. OF CASES CORRECT	NO. OF CASES EXAMINED	% CORRECT
Suchey-Katz (M)	13	21	70%
McKern-Stewart (M)	13	17	76%
Suchey-Brooks (F)	10	11	91%
Gilbert-McKern (F)	4	11	36% *

* The last category is not open-ended. If it were, the results would be:

METHOD	NO. OF CASES CORRECT	NO. OF CASES EXAMINED	% CORRECT
Gilbert-McKern (F)	5	11	45%
Auricular Surface (M&F) (Correct)	20	47	43% +
Auricular Surface (Correct *or* within one age category of being correct)	36	47	77%

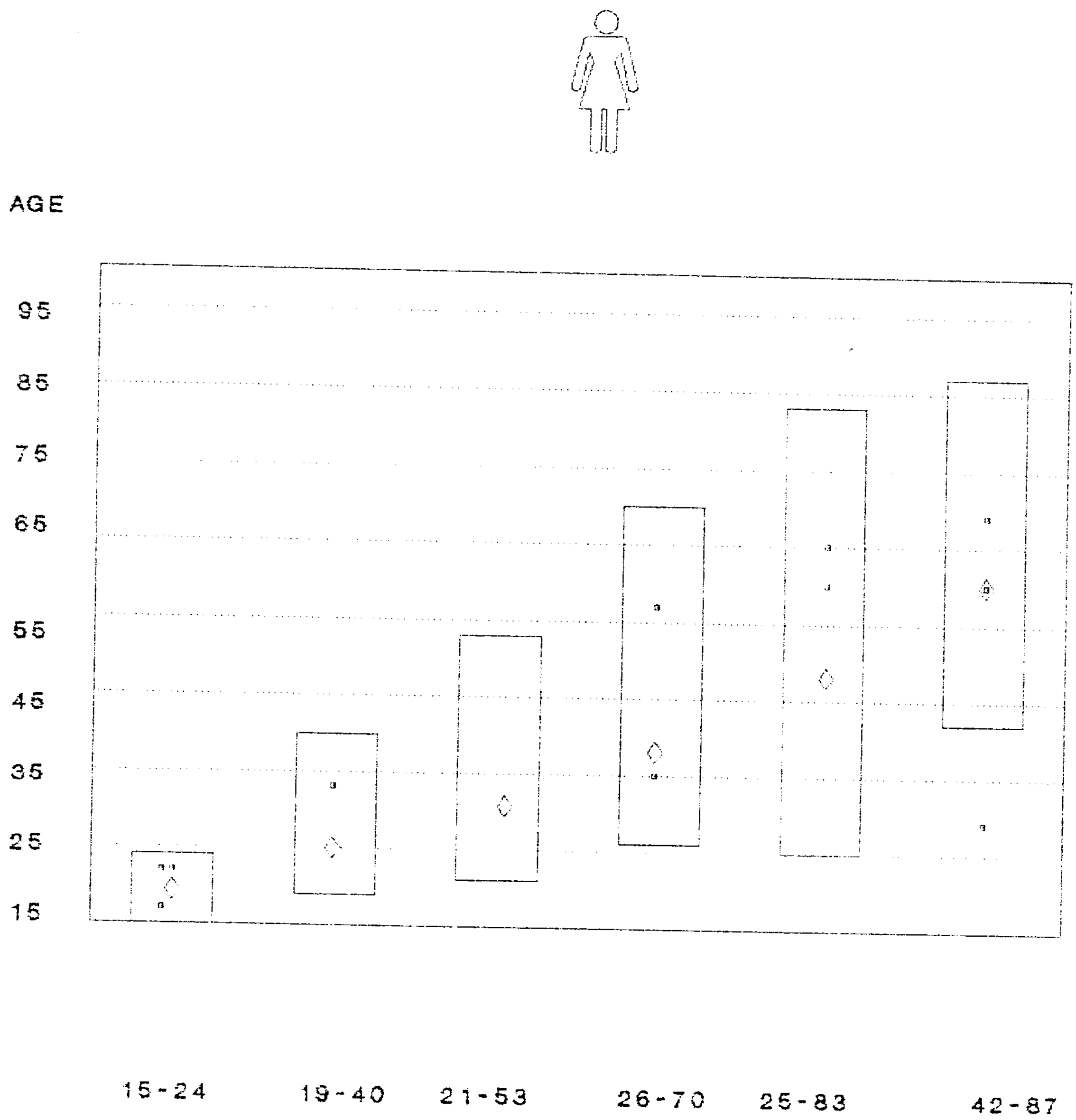
+ The youngest category is 20-24. Cases with actual ages younger than 20-24 were counted correct when assigned to this category.

FIGURE 12. SUCHEY-KATZ PUBIC SYMPHYSIS METHOD
COMPARISON OF ESTIMATES TO KNOWN AGES
ST. THOMAS CHURCH CEMETERY, BELLEVILLE



NOTE: DIAMONDS REPRESENT THE MEAN OF THE CATEGORY

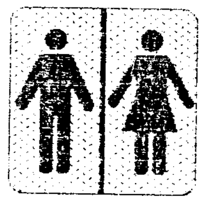
FIGURE 13. SUCHEY-BROOKS PUBIC SYMPHYSIS METHOD
COMPARISON OF ESTIMATE TO KNOWN AGES
ST. THOMAS CHURCH CEMETERY, BELLEVILLE

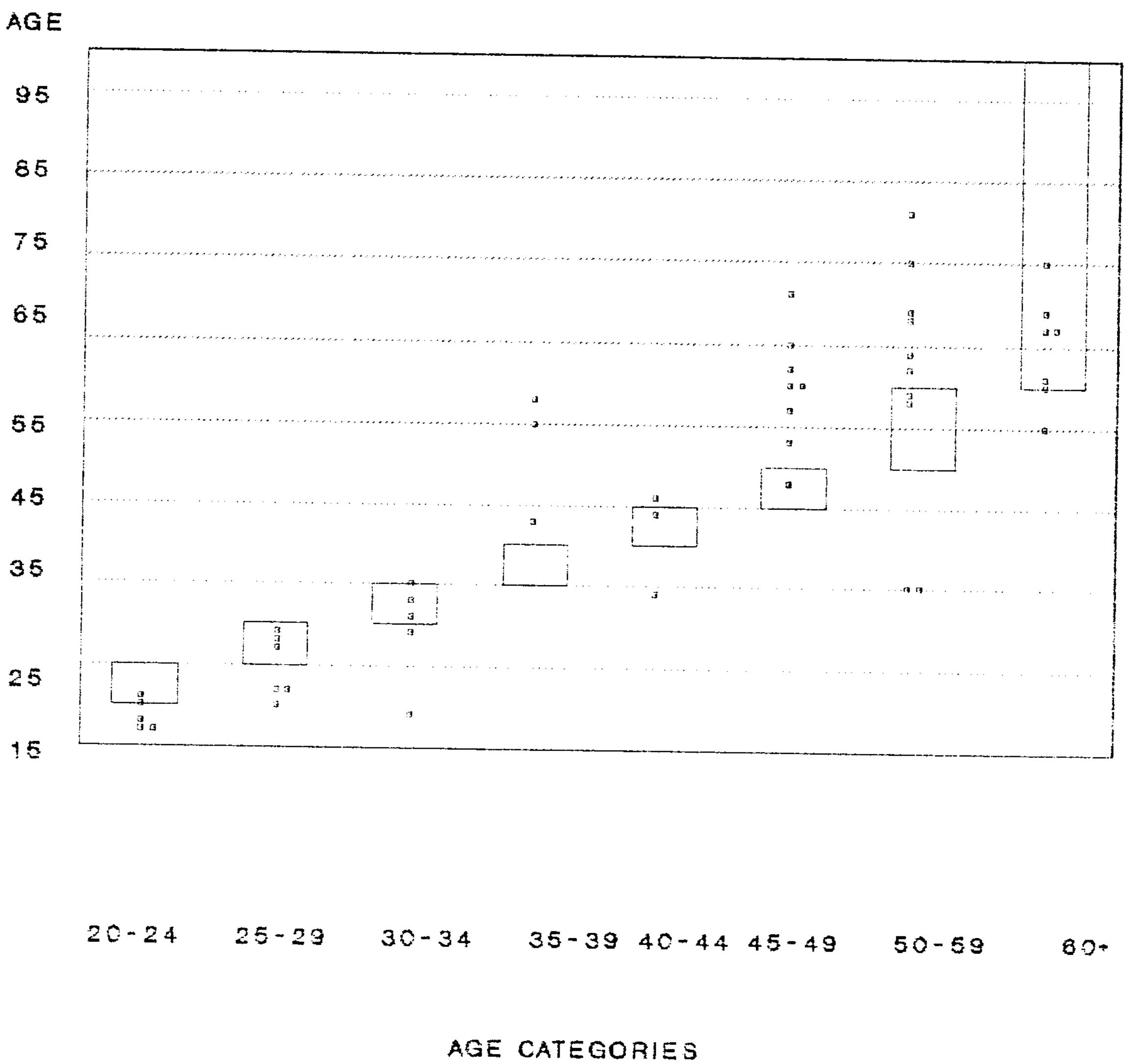


AGE CATEGORIES

NOTE: DIAMONDS REPRESENT THE MEAN OF THE CATEGORY

FIGURE 14. AURICULAR SURFACE METHOD
COMPARISON OF ESTIMATES TO KNOWN AGES
ST. THOMAS CHURCH CEMETERY, BELLEVILLE





for each individual, without reference to over- or underaging" (Lovejoy et al., 1985). It is calculated using the formula:

Inaccuracy (years) = $\Sigma|\text{estimated} - \text{actual age}| / N$

Bias is "the mean over or under prediction" (Lovejoy et al., 1985) and it is calculated using the formula:

Bias (years) = $\Sigma(\text{estimated} - \text{actual age}) / N$

The problem that arises when inaccuracy and bias are to be calculated is that both measurements require that each individual be represented by a single age estimation, not a range. Therefore, the mean age of the category (as established for the reference population by the authors of the methods) was used to represent the estimated age of an individual. The auricular surface method was treated slightly differently since the authors (Lovejoy et al., 1985) did not provide means (calculated from the original sample) for their age categories. For this method, ranges of inaccuracy and bias were produced by using the youngest age in the category, e.g. 20, and the oldest age of the category, e.g. 24. The results are presented in Table 20. Looking at this table, it is evident that the sample sizes of each age category are extremely small, casting doubt upon the validity of drawing conclusions from these data. For the most part, greater importance should be placed upon the normalized values (the overall bias and accuracy values for the method) due to the larger sample sizes. However, it must be noted that inaccuracy and bias vary in degree by age range; being higher

TABLE 20. INACCURACY AND BIAS OF METHODS OF ESTIMATING AGE-AT-DEATH FROM THE SKELETON

Age Range	Suchey-Katz	McKern-Stewart	Suchey-Brooks	Gilbert-McKern	Auricular Surface
17-29					
Inaccuracy	2.8	2.0	9.4	8.5	3.4 to 5.7
Bias	2.8	.7	7.3	7.1	1.7 to 5.7
N=	3	3	4	4	13
30-39					
Inaccuracy	3.6	5.5	11.2	4.0	7.5 to 10.5
Bias	-.6	-5.5	-4.8	4.0	4.5 to 10.2
N=	3	3	2	2	6
40-49					
Inaccuracy	N/A	N/A	N/A	N/A	5.3 to 1.8
Bias					-5.3 to -0.8
N=	0	0	0	0	4
50+					
Inaccuracy	21.5	30.4	11.3	14.6	14 to 15.3*
Bias	-21.2	-30.4	-11.3	-14.6	13.6 to -0.1
N=	11	11	5	5	24
Normalized over the age range					
Inaccuracy	15.3	21.0	9.6	10.4	9.4 to 10.8
Bias	-13.3	-20.5	-2.9	-3.3	-6.3 to 2.7
N=	17	17	11	11	47

* An artificial upper age limit of 90 years was created for this category to facilitate calculations.

in the older age categories. Table 20 indicates that most of the St. Thomas' sample is drawn from individuals in these older age categories. Therefore, one should expect the normalized values to be high. The purpose of presenting these data, in spite of the small sample sizes, was to make this association clear.

Despite the small sample sizes, the Suchey-Katz and the McKern-Stewart methods are directly comparable since the data for both were collected from exactly the same individuals. Likewise, the Suchey-Brooks and the Gilbert-McKern methods are comparable.

In order to provide some standard of comparison for the auricular surface method, Table 21 contains the levels of inaccuracy and bias obtained by Lovejoy and his colleagues (1985) as well as those achieved in this study. The implications of these findings are discussed in Chapter 5.

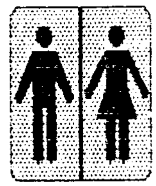
The ages recorded in the St. Thomas' parish register were used to generate three age profiles: one for males; one for females; and one with the sexes combined. Figures 15-23 compare the documented and the estimated profiles. The pubic symphysis methods are each represented by two age distributions. The first assigns the mean age of the estimated category to the individual cases which fall into that phase (for phases, means and standard deviations of each method see Appendix 3). For example, using the Suchey-Katz (1986) technique, all males estimated to be in phase III were

TABLE 21. INACCURACY AND BIAS OF THE AURICULAR SURFACE TECHNIQUE

ACTUAL AGE (years)	Lovejoy et al. Hammon-Todd	St. Thomas Belleville
20-29		
Inaccuracy	3.2	3.4 to 5.7
Bias	2.6	1.7 to 5.7
30-39		
Inaccuracy	7.2	7.5 to 10.5
Bias	1.6	4.5 to 10.2
40-49		
Inaccuracy	7.7	5.3 to 1.8
Bias	-2.9	-5.3 to -0.8
50-59		
Inaccuracy	11.1	12.1 to 11.9
Bias	1.9	-10.7 to -1.6
60+		
Inaccuracy	7.2	14.8 to 16.7
Bias	0.0	-14.8 to 0.5
All ages normalized		
Inaccuracy	7.3	9.4 to 10.8
Bias	-0.5	-6.3 to 2.7
N	108	47

Inaccuracy (years) = Σ |estimate - actual| / N
Bias (years) = Σ (estimate - actual) / N

FIGURE 15. COMPARISON OF DOCUMENTED,, AURICULAR SURFACE AND SUCHEY (SEXES COMBINED) AGE PROFILES



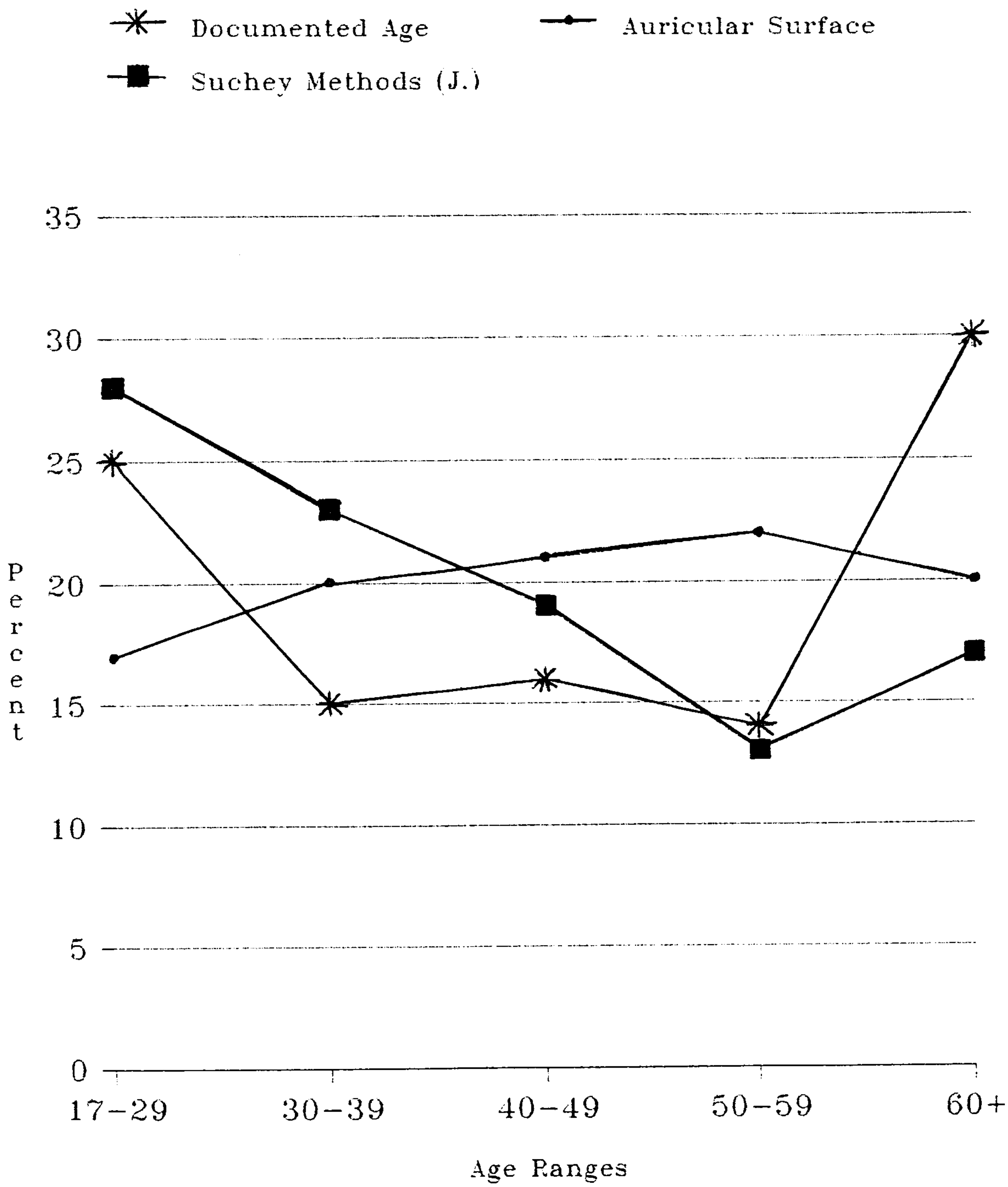


FIGURE 16. COMPARISON OF DOCUMENTED AND SUCHEY-KATZ ESTIMATED AGE PROFILES FOR MALES IN THE ST. THOMAS SAMPLE

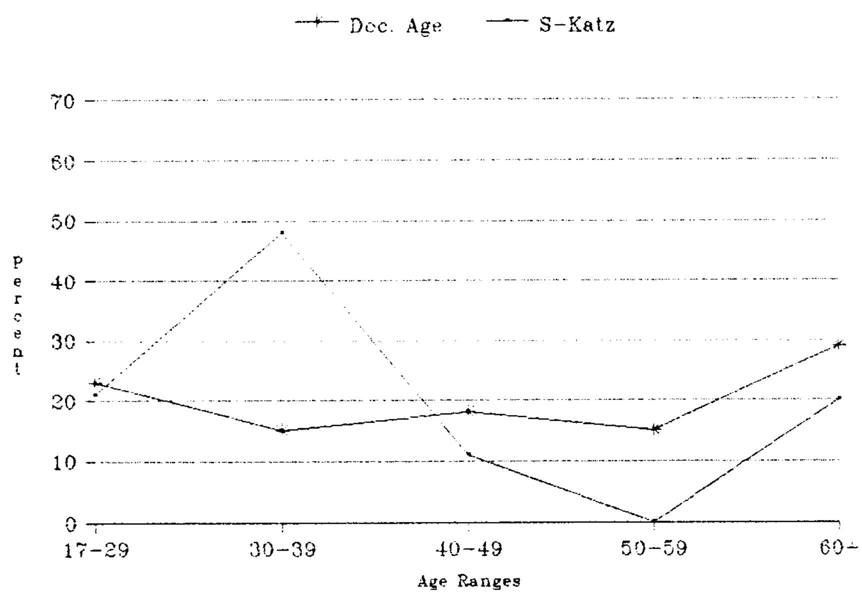


FIGURE 17. COMPARISON OF DOCUMENTED AND SUCHEY-KATZ (JACKES DISTRIBUTION) ESTIMATED AGE PROFILES FOR MALES IN THE ST. THOMAS' SAMPLE

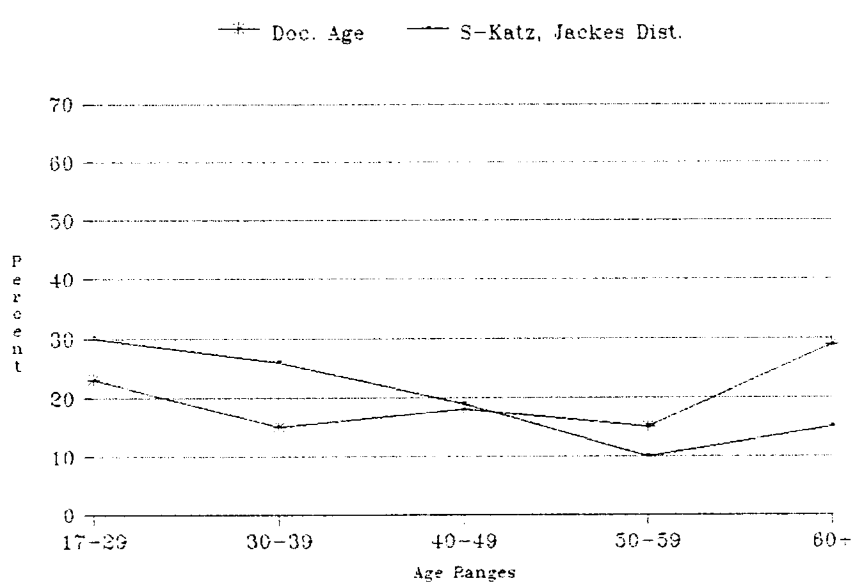


FIGURE 18. COMPARISON OF DOCUMENTED AND MCKERN-STEWART ESTIMATED AGE PROFILES FOR MALES

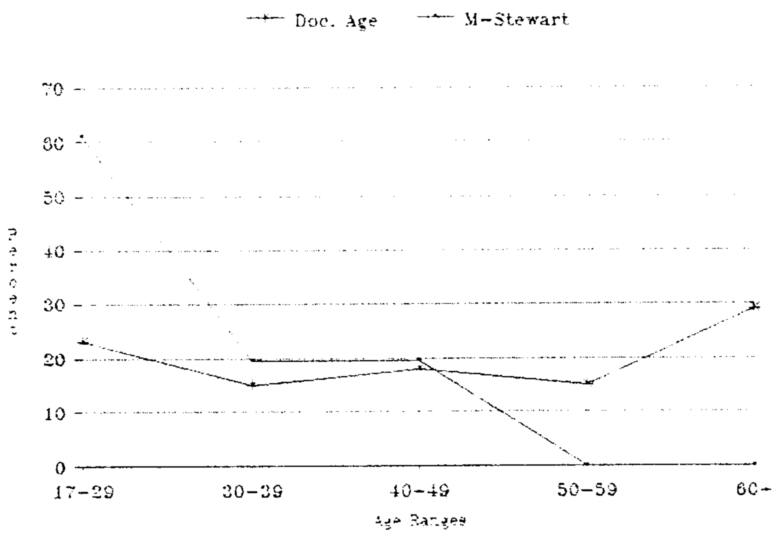


FIGURE 19. COMPARISON OF DOCUMENTED AND MCKERN-STEWART (JACKES DISTRIBUTION) ESTIMATED AGE PROFILES FOR MALES

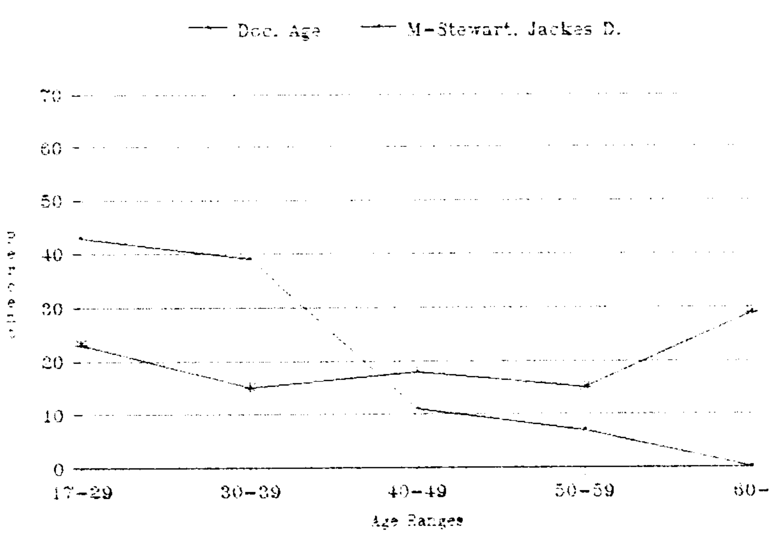


FIGURE 20. COMPARISON OF DOCUMENTED AND SUCHEY-BROOKS ESTIMATED AGE PROFILES FOR FEMALES IN THE ST. THOMAS' CEMETERY SAMPLE

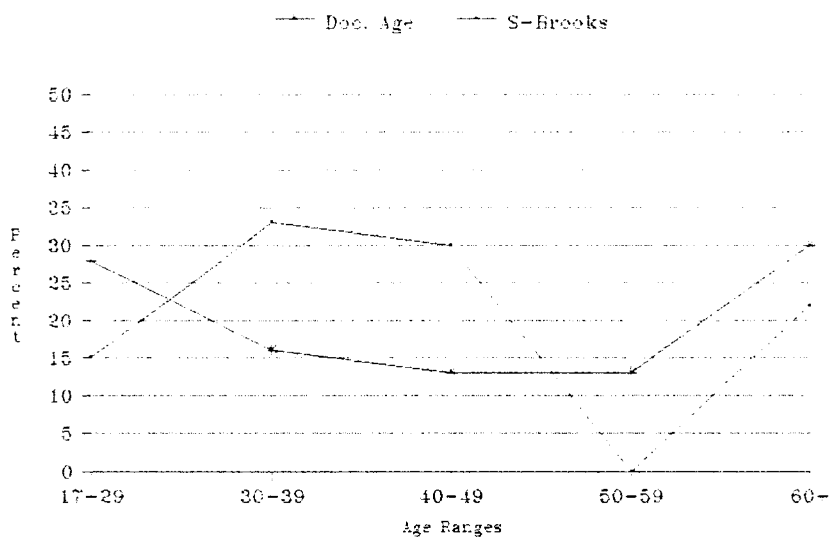


FIGURE 21. COMPARISON OF DOCUMENTED AND SUCHEY-BROOKS (JACKES DISTRIBUTION) ESTIMATED AGE PROFILES FOR FEMALES IN THE ST. THOMAS' CEMETERY SAMPLE

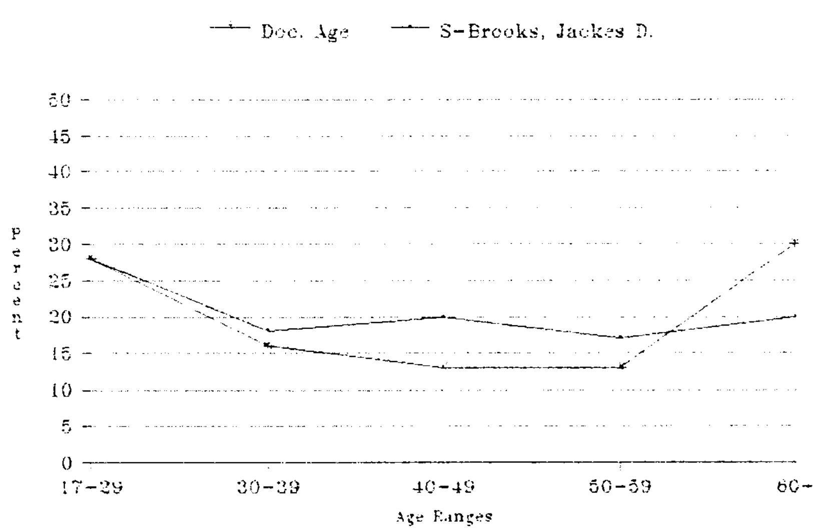


FIGURE 22. COMPARISON OF DOCUMENTED AND GILBERT-MCKERN ESTIMATED AGE PROFILES FOR FEMALES IN THE ST. THOMAS' CEMETERY SAMPLE

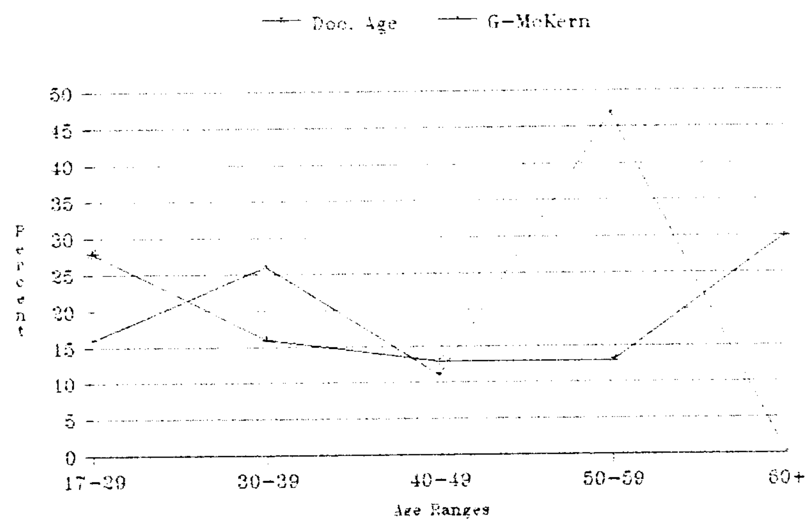
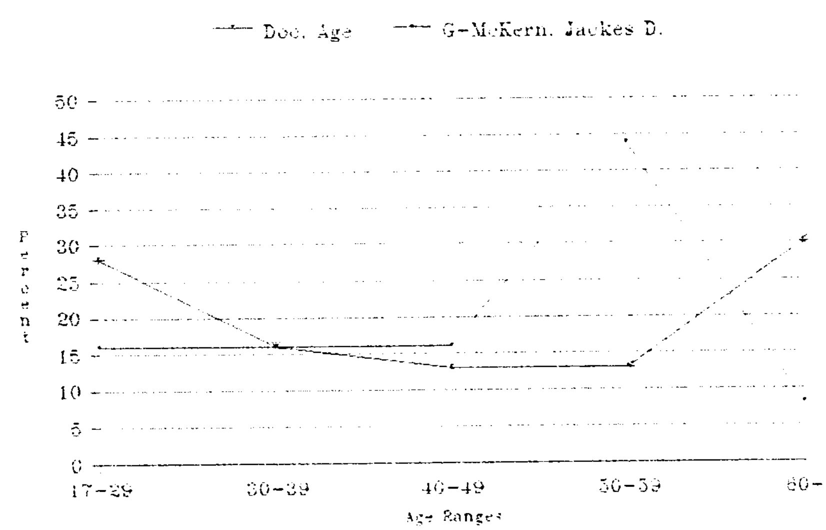


FIGURE 23. COMPARISON OF DOCUMENTED AND GILBERT-MCKERN (JACKES DISTRIBUTION) ESTIMATED AGE PROFILES FOR FEMALES IN THE ST. THOMAS' CEMETERY SAMPLE



assigned the age of 28.7 (the mean age for phase III). As a result, all phase III pubes fit into the 20-29 year slot of the age profile for the sample. Jackes (1985) argues that this approach can result in uneven distributions across all possible ages, since the means of the phases are not necessarily evenly dispersed. The McKern-Stewart (1957) method, for example, has eight phase means that fall into the 17-29 year slot.

As an alternative method of distribution, Jackes (1985) recommends casting all the cases within a phase over its 95% range (2 standard deviations) by assuming that the ages are normally distributed and calculating the probability of obtaining each age within the range (see Jackes [1985] for methodology). Figures 17, 19, 21 and 23 illustrate the results of distributing the age data according to the Jackes (1985) methodology (indicated by the name Jackes). In addition, the Suchey-Katz and Suchey-Brooks results (with Jackes distribution) were combined in order to produce a pubic symphyseal age profile comparable to that of the auricular surface and the documented sample (Figure 15). The distributions for the McKern-Stewart and the Gilbert-McKern systems were generated from the probabilities provided by Jackes (1985) while the probabilities for the Suchey methods were produced following Bennett (1979: 482), employing the table of areas under the normal curve (ibid: 496).

Since the purpose of creating an age profile for a sample is to identify patterns of mortality, it was not necessary to conduct statistical analyses on the estimated versus the actual distributions because, in this case, the visual representations are more expressive than the statistics.

CHAPTER FIVE
5.0 DISCUSSION

5.1 SEX DETERMINATION

Precision is a measure of an observer's ability to reproduce her/his results and is both a reflection of the researcher's capabilities as well as a commentary on the nature of the criteria being examined. Low precision (intraobserver error greater than 10%) indicates that a feature cannot be assessed reliably. Whether it be the researcher or the criteria which is/are at fault, the outcome is the same - an observer might correctly assess the sexes of a group of skeletons, but when required to repeat those results, has a greater than 1 in 10 chance of making an incorrect assignment. Unfortunately, in a practical situation, the researcher never knows whether that 1 in 10 error has been made. Consequently, features which exhibit high levels of intraobserver error (low precision) should simply be excluded from the analysis.

The overall degree of intraobserver error for both the pelvic (11.3%) and the cranial (12.2%) trait lists were slightly higher than the acceptable level of 10%, suggesting that at least some of the features employed in this analysis are difficult to assess. An examination of intraobserver error by trait identified four problematic pelvic criteria; acetabulum size and shape (11.2%), auricular surface height (11.3%), pre-auricular sulcus (11.3%) and ischiopubic ramus shape (11.3%), and two problematic cranial criteria; nasal aperture size and shape (10.3%) and orbital size and shape (12.2%). When considered individually, the amount by which these difficult features exceed the acceptable level is minimal. However, the combined impact of these criteria could significantly affect the outcome of a skeletal sex analysis, especially if the material were fragmentary and the problematic features were among the few traits still observable. It is therefore recommended that all criteria exceeding the 10% critical level of intraobserver error be excluded from the analysis.

An interobserver error test could not be conducted independent of the analysis so it remains unclear whether the difficulties associated with these six features arose due to some idiosyncrasy of this researcher or if they are problems inherent to the features themselves. While most of the criteria examined in this study were subjective, some were more subjective than others. Gauging acetabulum size by visual estimation, for example, is preferable to physical measurement in that it is quicker. However, one suspects that facility comes at the expense of precision. The same may be said for nasal aperture and orbital sizes. Similarly, the pre-auricular sulcus presents itself as varying degrees of resorption (Tague, 1988), the interpretation of which is not always a simple procedure. Some overlap between the male and female forms of the ischiopubic ramus does occur and the results of other investigators (Phenice, 1969; Sutherland and Suchey, 1987) suggests that this particular trait is unreliable. On the other hand, determining auricular surface height, as described by Bass (1987), seems a straightforward enough procedure so that perhaps the researcher was at fault.

It is recommended that an intraobserver error study be undertaken by all researchers in order to identify problematic criteria. The features may then be employed in the trait list at the discretion of the investigator. If others discover a precision problem with the six criteria identified by the author, the argument for abandoning these features is strengthened. Alternatively, if no criteria, or other criteria, are identified as problematic, a deficiency on the part of the observer is suggested and further practice is advised.

Accuracy is a measure of the number of correct assessments, usually presented in terms of percentage. The simple morphological method of estimating sex employed in this study produced an accuracy of 95.9% for the pelvic criteria. As a technique for group analysis, it produced a sex ratio statistically indistinguishable from the documented ratio in the parish registers. The cranial results were less spectacular.

Hrdlicka (Stewart, 1952) predicted that experienced investigators would be able to correctly identify sex, using only the skull and mandible, in 90% of the cases examined. Stewart (1979) assessed one hundred crania from the Terry skeletal collection and was able to correctly assign sex for 77% of the individuals. More recently, Krogman and İşcan (1986) stated optimistically that 92% accuracy could be achieved. The present investigation produced results similar to those of Hrdlicka; 89.1% accuracy was obtained. As a method of group analysis, the cranial trait list produced a sex ratio that was statistically indistinguishable from the documented ratio as determined from the parish burial records.

A comparison of the hypothetical pelvic trait list (Table 4) and the pelvic trait ranking which resulted from this analysis (Table 9), indicates that 67% of the criteria predicted to rank within the top 10, did so. Expressed slightly differently, the mean difference in ranking (calculated as $\Sigma|\text{estimated - actual rank}|/N$) was 5.6 places. These discrepancies can be translated into observations concerning the predictions, generated from a compilation of data from other researchers, and the hypotheses that were posited in Chapter 3.

It was discovered that, contrary to St. Hoyme (1984) but in keeping with St. Hoyme and İşcan (1989), the anterior features of the pelvis were more useful in correctly assigning sex than were the posterior features (3/4 anterior features are in the top 5). The criteria for acceptance of this hypothesis were met. The mean accuracy for the anterior features was 10% higher than the posterior traits (86% versus only 76%) and the intraobserver error was 5% less (2% versus 7%) for the anterior traits.

In contrast to Tague (1988), the pre-auricular sulcus scored very high on accuracy (91.6%), but due to low precision (11.3% intraobserver error) it placed 10th overall. This was one of the higher rankings for features of the posterior pelvis. Tague's (1988) assessment of the inaccuracy of dorsal pitting was upheld in this study (35.7% correct), although the precision of this feature was excellent (0% intraobserver error). This suggests that while it may be easy to assess, dorsal pitting is not a particularly useful sex criterion.

St. Hoyme (1984) appears to have been correct about the importance of the sacrum in skeletal sex determination. Coleman's (1969) assertion that there is no difference in the growth of the male and female sacrum is questionable in light of its capacity to reflect sex. However, Coleman's (1969) claim that features dependent upon two functional divisions (the false and true pelvis) are less sexually dimorphic than those dependent upon only one was borne out. Both the sciatic notch (85.7%) and the pelvic inlet (80%) exhibited relatively mediocre levels of accuracy and poor precision (6.5% and 9.7% intraobserver error, respectively).

Although Coleman (1969) noted no evidence of significant sexual differentiation in growth of the acetabulum, this feature can be extremely accurate (91.7%) but it is difficult to assess (11.2% intraobserver error). This suggests that if a researcher chooses to include acetabulum size in her/his analysis, the data should be evaluated metrically. The time saved through a morphological assessment of this feature cannot compensate for the low precision which results.

St. Hoyme (1984) indicated that the obturator foramen was of little value as a sex indicator, yet in this study it ranked 2nd overall (Table 9). In contrast, muscle markings proved to be ineffectual (56.8% accuracy), although they were observed consistently (0% intraobserver error). This suggests that they may contribute to a general impression of the true sex but, on their own, they produce results only slightly better than what might be expected from chance alone.

This investigation reproduced the original rankings of the Phenice criteria (Phenice, 1969). When precision and accuracy are combined, the ventral arc ranks first (accuracy 86.9%, intraobserver error 0%); subpubic concavity is second with an overall rank of 5 (accuracy 83.8%, intraobserver error 3.2%); and the ischiopubic ramus is third with an overall rank of 16 (accuracy 80%, intraobserver error 11.3%). This study differs from Phenice (1969) with regard to the accuracy that may be achieved by combining the three features. Phenice (1969) claimed 96% accuracy for the combined criteria, whereas this analysis managed to achieve only 88% accuracy. This is

TABLE 22. THE RESULTS OF THIS STUDY COMPARED TO THOSE OF PHENICE (1969), KELLEY (1978), SUTHERLAND AND SUCHEY (1987), LOVELL (1989) AND MCLAUGHLIN AND BRUCE (1990)

STUDY	ACCURACY OF PHENICE METHOD	ACCURACY OF INDIVIDUAL CRITERIA*		
	% Correct	VA	SPC	IR
Phenice	96%	-	-	-
Kelley	95%	-	-	-
Sutherland and Suchey	-	96%	-	70%
Lovell	83%	-	-	-
Bruce and McLaughlin				
English sample	83%	65%	87%	61%
Dutch sample	68%	55%	72%	59%
Scottish sample	59%	45%	75%	56%
Mean accuracy	70%	55%	78%	59%
This study	88%	87%	84%	80%

*Key to abbreviations

VA = ventral arc
SPC= subpubic concavity
IR = ischiopubic ramus

still higher than the average results reported by Lovell (1989), 83% accuracy (using 50 hip bones), and by McLaughlin and Bruce (1990), 70% accuracy (using 273 skeletons).

The results of the individual criteria vary in comparison to those of other researchers (Table 22). While Sutherland and Suchey (1987) had excellent results with the ventral arc (96% using 1284 specimens), McLaughlin and Bruce (1990) reported an average of only 55.1% success on the three samples they examined. In this analysis the level of accuracy fell between the values for the other studies (86.9%). McLaughlin and Bruce (1990) found the subpubic concavity to be the most useful single criterion, with an average accuracy of 78.2% (compare those results to 83.3% accuracy obtained in this study). Both Sutherland and Suchey (1987) and McLaughlin and Bruce (1990) had less success with the ischiopubic ramus, 70% and 61% accuracy, respectively, compared to this analysis, 80% accuracy. In contrast to Lovell (1989), in which the ventral arc was perceived by many participants to be most difficult to assess, this investigation found that the shape of the ischiopubic ramus is the most unreliable indicator (intraobserver error 11.3%) and, in fact, it is recommended that this feature be dropped from the pelvic trait list due to its low precision. The highest level of accuracy achieved in this study, for the Phenice criteria, was the result of combining the observations of the ventral arc and subpubic concavity; 92% accuracy was obtained.

McLauglin and Bruce (1990) suggested that the lower levels of correct sex identification found in their study, compared to those of Lovell (1989), Phenice (1969) and Sutherland and Suchey (1987), may have been due to population differences, theirs being a European rather than a North American sample. The St. Thomas' material is interesting in this sense since it is a 19th century North American cemetery sample that it is comprised largely of individuals who were recent British and Irish emigrants. One of the samples employed by McLaughlin and Bruce (1990) derived from a 17th-18th century English cemetery. In the case of the ventral arc and the ischiopubic ramus, the accuracy obtained in this study was superior to those of the English sample (86.9% compared to 64.9% and 80% compared to 61%, respectively). In contrast, the subpubic concavity fared slightly better in the English (87.1%) than in the St. Thomas' sample (83.3%). The St. Thomas' results are closer to those of Sutherland and Suchey's (1987) modern North American sample than to the historic English sample of McLaughlin and Bruce (1990), despite the fact that the St. Thomas' sample was comprised largely of English and Irish emigrants (Table 22).

A comparison of the hypothetical cranial trait list and the rankings resulting from this study (Table 18) indicates that the mean difference in ranking between the estimated and actual ranking was 6.4 places. Again, these discrepancies can be translated into observations about the hypotheses generated from a review of the cranial growth literature and presented in Chapter 3.

Neither the facial criteria nor the features of the calvarium were more useful than the other in determining sex from the skull. Of the top five criteria, three were from the face (60%) while two were from the calvarium (40%). It was suggested that due to the greater degree of relative growth of the face as compared to the calvarium, the sexually dimorphic features of the face would be more marked and therefore easier to assess.

This hypothesis proved to be incorrect since the criteria necessary for its acceptance were not met. Although the mean accuracy of the facial features was 10% greater than the features of the calvarium (49% versus 35%) the level of intraobserver error was not 5% lower for the facial traits (5% versus 2%). Perhaps the interaction of all of the growth patterns of the facial bones obscures the sexually dimorphic patterns of growth, thereby reducing the effectiveness of the facial criteria as sex indicators.

Although supraorbital ridges, frontal eminences and forehead shape are dependent upon the same growth process for their final form, the supraorbital torus was easier to assess (precision was lower) and it produced better results (60.9% accuracy compared to 44.5% and 31.9%).

The cranial ranking generated by this study (Table 18) may also be compared to the rankings produced by other physical anthropologists (Table 13). Table 18 was most similar to Hrdlicka's ranking (Stewart, 1952), exhibiting a mean difference of 2.4 places. He employed nine criteria that were directly comparable to those used in this study, 78% of which ranked within the top nine in this investigation. The differences involved features relating to the mouth. While Hrdlicka recommends alveolar arch and teeth size and palate shape and size, this study found these features performed rather poorly.

The Krogman and İşcan (1986) and the St. Thomas' rankings exhibit a mean difference of 3.25 places. Of the eight criteria employed by Krogman and İşcan, 75% tested in the top eight of this study. They recommend both orbit and palate shape and size while this analysis ranks the same features 17th and 14th, respectively.

Stewart (1979) employed only six criteria comparable to those used in this investigation. The mean difference between rankings was 3.7 places and 83% of his criteria tested in the top six of this study, the only disagreement being the value of the orbital borders: Stewart ranks it most useful while this author found it least useful.

Meindl et al. (1985) also employed six criteria comparable to those in this analysis. Their mean difference was 4.5 places and only 50% of their criteria tested in the top six of this study. Finally, El-Najjar and McWilliams (1979) employed only three criteria. The mean difference between their rankings and Table 18 was 6.3 places and only 33% of their criteria tested in the top three of this study.

By employing a trait list similar to that of Hrdlicka (Stewart, 1952), it was possible to virtually reproduce his degree of accuracy (90% versus 89.1%) and to approximate his ranking of criteria by usefulness. Although he does not explicitly state upon which population his ranking is based and his criteria most successfully employed, Hrdlicka does make reference to a variety of groups from "African and American Negroes" to "Whites" (Stewart, 1952: 129) to "Eskimos" and "northern Indian tribes" (ibid: 130), suggesting that the cranial criteria which he employed were not population specific. The fact that these traits were successfully and reliably applied to the St. Thomas' sample indicates that they are not sample specific and suggests that they may be effective on other populations.

Despite the fact that some pelvic and cranial features become more distinctive with age, e.g. the ventral arc (Sutherland and Suchey, 1987), while others lose their characteristic appearances with age, e.g. dorsal pitting (Kelley, 1979), the only statistically significant age-related effects on accuracy occurred for posterior view of the sacrum, zygomatic root extension and occipital muscle markings, all of which exhibited increased accuracy with age. Analysis of the complete pelvic and cranial trait lists similarly produced no significant fluctuation in accuracy due to age.

Meindl et al. (1985) claim that the success rate of assessing male skulls is lower for younger individuals, whereas with female skulls it decreases for older individuals. Unfortunately, due to already small sample sizes that resulted from dividing the known subsample into three age categories (n=6 to 26), it was felt that further division of the groups by sex would only produce invalid conclusions. Thus the males and females were evaluated together and no overall age-related differences in accuracy were discovered. This could mean that no differences exist, but it could also indicate that the male differences (reduced accuracy in younger individuals) were cancelled by the female differences (reduced accuracy in older individuals).

When the 55 assignments of sex (including retests) made from the skull were compared to the parish burial records, it was discovered that seven were incorrect. Three were cases of females thought to be males, while four were males thought to be female. The three cases of female mistaken identity were from two individuals; one was assessed twice, both times incorrectly. Burial 309 was a 60 year old woman. Burial 464, examined twice, was 46 years. Although the number of errors are small, the mistakes would seem to conform to Meindl et al.'s observation that, "greater age produces an increasingly male morphology" (Meindl et al., 1985). It should be noted, however, that there were seven other women in this sample over the age of 60 and all were assessed correctly.

The four cases of male mistaken identity were also concentrated in the older age categories. Burial 339 was 55; burial 397 was 75 and burial 429 was 43 years. Only burial 472, aged 20 years, conformed to Meindl et al.'s (1985) suggestion that younger males might be mistaken for females more frequently than were older males. The evidence is inconclusive.

Using the pelvis, 72 assignments of sex, including retests, were verified through the parish burial records. Of these 72 estimations, only two were incorrect. Burial 351A was a young female aged 22 years and three months. This individual was assessed twice. The initial results indicated that the specimen was female, the retest suggested male. The second case, burial 464, was extremely fragmentary. It was a woman, aged 46 years, who was incorrectly assigned as male. Although there were only two errors, the bias is toward males, as was predicted.

Tests of alternate methods of grouping the criteria revealed that no single pelvic feature produced better results than those generated by the complete trait list (95.9%). The most accurate single indicator was sacrum shape (94.1%). On the other hand, three combinations of pelvic criteria produced higher levels of accuracy than the trait list as a whole: obturator foramen shape and presence of the ventral arc (98%); obturator foramen shape and true pelvis (98%); pubis shape and acetabulum shape and size (96%). Since acetabulum shape and size exhibited low precision, the first two combinations are considered more reliable. While combinations of three traits could equal and even slightly surpass the accuracy of the complete trait list, none could produce results superior to the obturator foramen/ventral arc and obturator foramen/true pelvis combinations. This is not entirely surprising since the first three criteria, ranked by a combination of highest accuracy and greatest precision, are: ventral arc; obturator foramen; and true pelvis (Table 9.). Combining these three features does not increase accuracy above that obtained by the two combinations involving the obturator foramen. The ventral arc/true pelvis combination (95% accuracy), while an increase over the individual accuracies (86.9% and 85.8%, respectively) is not as useful as the combinations involving the obturator foramen. This suggests that true pelvis and the ventral arc are contributing similar information to the equation.

When this analysis is compared to Sutherland and Suchey (1987) and McLaughlin and Bruce (1990), it becomes evident that there is a certain degree of variability in the levels of accuracy that may be achieved by different researchers employing the same criteria on different populations. Therefore, despite the fact that this investigation discovered that the greatest accuracy for the pelvis was achieved by combining the results of only two criteria, the author does not advise focusing solely on these features when assessing sex. At the other extreme, the complete pelvic trait list produced an accuracy of 95.9% and had an unacceptably high level of intraobserver error (11.3%). Thus, it would appear that the best results, and likely the most widely applicable results, could be obtained by employing some subset of the total pelvic criteria. A review of the individual levels of accuracy (Table 7) indicates that some features produce results only slightly better than those expected by chance and, as indicated previously, other traits have low precision (Table 5) such that they should be excluded from the analysis. Table 9 ranks the pelvic criteria according to both high accuracy and precision. The first six traits (ventral arc, obturator foramen, true pelvis, sacrum shape, subpubic concavity and pubis shape) all have intraobserver error levels below 5% and are capable of successfully assigning sex in over 83% of the cases examined. Thus, these features are most highly recommended for use in determining sex from the bony pelvis. In cases of indecision, the results of this study indicate that emphasis should be placed upon the results of the obturator foramen/ventral arc and obturator foramen/true pelvis combinations.

No single cranial feature exhibited greater accuracy than the complete cranial trait list (89.1%). The best individual indicator was nasal aperture (76.6%), which had to be excluded from the analysis due to its poor precision. Similarly, none of the combinations of two cranial traits were capable of generating a higher level of accuracy than the complete trait list. Only one combination of three criteria, zygomatic root extension/malar size and shape/nasal aperture size and shape, improved upon the accuracy of the original method. Since this combination employs one criterion which exceeds the acceptable level of intraobserver error, and only increases accuracy by 1.9%, the author does not advise the use of this combination in preference to the complete trait list. Unfortunately, however, the complete list has an unacceptably high level of intraobserver error (12.2%). An examination of the individual levels of precision (Table 14) and accuracy (Table 16) make it possible to identify features with low precision, high inaccuracy (greater than 30% error) and high percentages of "indeterminate" ratings (70% or more); all of which contribute very little in the way of constructive information. Instead, the ambiguity of these features serves to increase the chance of error. It is therefore recommended that researchers focus their efforts on the first eight criteria in Table 18 (supraorbital ridges, mastoid size, malar size and shape, occipital markings, chin form, general size, zygomatic root extension and mandible shape). Not only did these features perform successfully in this study, but they are also among the traits most commonly recommended by other physical anthropologists (see Table 13).

The foregoing discussion indicates, not surprisingly, that the preferred area for skeletal sex estimation is the pelvis. It is capable of producing more accurate and precise results than can be achieved by examination of the skull. Even in cases where both sets of information are available, the cranial assessment serves only to re-enforce that made from the pelvis. In this analysis, whenever discrepancies occurred between pelvic and cranial sex determination, it was inevitably the cranial assessment that was at fault (n=4).

5.2 AGE ESTIMATION

Precision tests of the Suchey-Katz (1986) and Suchey-Brooks (Suchey et al., 1988) methods of estimating age-at-death from the pubic symphyseal face, resulted in dismally high levels of intraobserver error (34% and 56%, respectively). When techniques exhibit such poor precision, a serious flaw in either the method itself or the application of the method is indicated. The fact that the McKern-Stewart (1957) and Gilbert-McKern (1973) systems were as unreliable (39% and 52%, respectively) as the Suchey methods, suggests that it is not necessarily the specific technique that is at fault, but rather, that there are problems inherent to this entire approach to skeletal age estimation. Conversely, this inability to reproduce results may be the fault of this observer or may be the result of a combination of both factors.

During the analysis, some difficulties associated with the Suchey methods were identified by the author. In the male method, phases III and IV are typically differentiated by the completion of the oval outline surrounding the face of phase IV pubes. In addition, phase IV pubes may have a distinct rim and bony growths on the inferior half of the symphysis may occur (Suchey et al., 1988). Examination of the casts used to illustrate the six phases (produced by Diane France), reveals that phase III stage 2 pubes appear to be very similar to phase IV stage 1 pubes, the difference being degree of completion of the ventral rim. A review of the cases used in the intraobserver error test for the St. Thomas' sample indicates that phase III pubes were estimated consistently on two occasions, phase IV pubes were estimated consistently on fourteen occasions and in five cases the estimation changed from one phase to the other. These results illustrate the difficulty of reliably assessing phase III stage 2 and phase IV stage 1 pubes and partially explain the high level of intraobserver error.

An additional problem identified for both Suchey methods, was the difficulty in differentiating between phases III (2) or IV (1) and VI. Specifically, there was some question as to whether the ventral rim was in the process of forming or breaking down. This problem may be experience related. An examination of the data generated during the intraobserver error study reveals that the incorrect assessment was more frequently made during the first set of observations than in the second (4 to 2), although the sample size is far too small to statistically support this conclusion. On the other hand, Suchey (1979) discovered that researchers using the Gilbert-McKern (1973) technique also had problems determining whether the ventral rampart was in the process of being built up or broken down. Perhaps greater elaboration concerning the means by which one distinguishes between the two situations is required.

If the intraobserver error resulting from the two types of problems outlined above is eliminated, the Suchey-Katz (1986) method is reduced to 11.4% intraobserver error. On the other hand, even when these two main sources of intraobserver error are removed from the Suchey-Brooks (Suchey et al., 1988) method, intraobserver error is still unacceptably high at 24%. Although hardly conclusive, these results suggest that the Suchey-Katz method of estimating age-at-death for the males is more reliable than the Suchey-Brooks method for females.

The level of intraobserver error for the auricular surface method (19.3%) was much more reasonable than the levels obtained for the pubis methods, although it was still unacceptably high. Most of the error (15.2%) was comprised of cases that disagreed by one category, suggesting that minor refinements, rather than complete reversals of decisions, were taking place. Since most of the auricular surface age categories span only five years, a change in one category results in only a few years difference. In contrast, the Suchey-Katz and Suchey-Brooks age ranges are so extensive that a change of one category could mean an increase anywhere from 9-10 (Suchey-Katz) or 4-17 (Suchey-Brooks) years.

It is argued, therefore, that the auricular surface method can produce reliable results and, in this respect, is superior to the pubic symphysis approaches. This is somewhat surprising given the greater difficulty in applying the auricular surface technique as compared to the pubic symphysis methods. Many of the individual characteristics examined in the auricular surface system are difficult to assess, grain and density being the most problematic. In addition, the format of this technique is more complex and requires a slightly different approach than is usual in age estimation methods. The emphasis is on descriptions; there are no casts; and the pictures and photographs represent, not the range of variability, but rather the modal appearance for each age category. Clearly, however, the reliability of the auricular surface method makes the complexities involved in applying it worth mastering.

Superficially, the Suchey-Brooks method appears to provide the best age-at-death estimates. Accuracy (Table 19) was 91% and the normalized inaccuracy and bias values (Table 20) were also comparatively successful; one of the lowest levels of inaccuracy (9.6 years) in addition to the lowest degree of bias (-2.9 years). However, Figure 13 illustrates exactly why this technique is successful. The age ranges employed in the Suchey-Brooks system are so expansive that only the very young and the very old categories are mutually exclusive. In some cases, up to four age ranges can accommodate the same individual. It might work, but on its own, the system provides the researcher with very little practical information beyond the fact that the individual is younger or older.

In contrast, the Gilbert-McKern method produced an accuracy of 45% (Table 19) but fared only slightly worse than Suchey-Brooks in both inaccuracy (10.4 years) and bias (-3.3 years). The Gilbert-McKern approach is better in the younger age categories while the Suchey-Brooks method produced slightly better results in the oldest age category. In both systems there is a tendency to underage individuals over fifty years.

The Gilbert-McKern system has the practical advantage of smaller age categories, which may be of little consequence to paleodemographers, but which is invaluable for forensic purposes. On the other hand, the Suchey-Brooks method, distributed in the manner suggested by Jackes (hereafter S-B-J), produces the age profile most similar to the one developed from the parish register (Figure 21). The one drawback is its proclivity for placing some of the individuals over sixty years of age into the 40-49 year category. This tendency could result in major difficulties for researchers attempting to produce life tables, since many life table calculations depend upon the number of individuals in the oldest age category. This problem might be corrected if a more detailed description of the means by which one differentiates between ventral rim construction and destruction could be included in the methodology. None of the other female techniques (Suchey-Brooks, Gilbert-McKern and Gilbert-McKern modified following Jackes [G-M-J]) managed to even approximate the actual age profile (Figures 20-23). This suggests that the Suchey-Brooks and Gilbert-McKern methods may be adequate in individual cases, however, in the construction of an age profile, the inherent biases of these systems become magnified to such a degree that they distort the distribution and render it useless. Thus, the Gilbert-McKern method appears to be better in individual cases, while the S-B-J approach provides a more accurate picture of the sample's age distribution.

The auricular surface technique proved to be the second most exact method of estimating age-at-death, exhibiting an accuracy (within one age category of the actual age) of 77%. Figure 14 and Table 20 illustrate the tendency of this system to slightly overage younger individuals while underaging older individuals. This pattern of under- and overaging was also identified by Murray and Murray (1991). They noted that somewhere between ages 30 and 40, a crossover point occurred. Table 20 places this crossover at about 40 years.

Experience with the method does not appear to have played a role in the incorrect age assessments since the misclassified individuals were analyzed at various times throughout the entire period of this aspect of the study (June 28 to July 12, 1990).

The normalized inaccuracy and bias values (9.4-10.8 years and -6.3 to 2.7 years, respectively) indicate that the auricular surface technique is comparable to the Suchey-Brooks and Gilbert-McKern methods but is much more successful than either the Suchey-Katz or McKern-Stewart methods. Table 21 compares the levels of inaccuracy and bias in this investigation to those obtained by Lovejoy and colleagues (1985). While the inaccuracy values are comparable (in all ranges except 60+), the degrees of bias are not. Not only is bias generally higher in this analysis, but the patterns of bias also differ. Both studies indicate slight overaging in the younger age categories and both indicate slight underaging in the 40-49 year range. Whereas this investigation found underaging a problem in all older categories, Lovejoy et al. (1985) found no bias or only slight overaging in the same categories.

The fact that the patterns of bias identified herein are identical to those of Murray and Murray (1991) and both differ from the original study, suggests that other, nonage-related factors may be affecting the results. Murray and Murray (1991) verified that the auricular surface technique was equally applicable to all individuals without adjustments for population or sex. Until it is confirmed that the 500 individuals from the Hammon-Todd collection used by Lovejoy et al. (1985) to develop their methodology, had documented ages rather than anatomically estimated ages associated with them, the possibility remains that the pattern of bias discovered by Murray and Murray (1991), confirmed here, was built into the methodology as a result of these estimates. If this hypothesis is correct, other researchers will continually and inevitably discover a similar problem with their attempts to employ the auricular surface technique.

While this method may be useful in some individual cases, Figure 15 indicates that age profiles created from auricular surface age estimates will produce inaccurate patterns of mortality. When the results of the Suchey methods (using the Jackes distributions) are combined, the age profile that is created is closer to the actual profile than is the auricular surface method distribution. Of course, all of the problems associated with the individual Suchey sex-specific age profiles are present in the combined age distribution.

The least accurate system of age estimation was the Suchey-Katz technique (70% accuracy). The McKern-Stewart method fared slightly better exhibiting an accuracy of 76% (Table 19). Examination of the inaccuracy and bias levels (Table 20) indicate that the two techniques are equally effective in the younger age categories but that the Suchey-Katz method produces somewhat better results for individuals over 50 years. Both systems tend to underage. The slightly higher accuracy of the McKern-Stewart method, despite its poorer inaccuracy value, can probably be explained by the difference in the samples used in calculating accuracy. Four cases aged by the Suchey-Katz technique could not be assessed using the McKern-Stewart method due to the incompleteness of the specimen. It is likely that the potential for error is increased when analyzing fragmentary remains, thus the higher accuracy achieved for the McKern-Stewart method is explained.

The McKern-Stewart system has the advantage of smaller age categories which would make it more useful in individual cases, except for the fact that its oldest range is only 36+, making it useless for older individuals. Neither the Suchey-Katz nor the McKern-Stewart methods produced a particularly accurate age profile (Figures 16 and 18). Both profiles generated from the McKern-Stewart technique (with and without Jackes distribution) were absolutely and completely inappropriate, however, the main flaw with the Suchey-Katz (with Jackes distribution) profile was elevated percentages of death in the 17-29 and 30-39 year categories at the expense of those in the 60+ division. This inadequacy can be linked to the problem of differentiating between phases III (2) or IV (1) and VI, discussed previously. If this problem could be eradicated, the S-K-J method would be capable of generating an accurate age profile for the sample.

Thus, it would appear that the McKern-Stewart method provides a more useful estimate of age for younger individuals but the Suchey-Katz system is better for older individuals and if the age profile for the males in the sample is to be constructed, the S-K-J method should be employed.

CHAPTER SIX
6.0 SUMMARY AND CONCLUSIONS

6.1 SUMMARY

The precision and accuracy of the pelvic and cranial morphological methods of skeletal sex determination and three systems of pelvic age estimation were tested on a sample of approximately 278 individuals from the 19th century St. Thomas' Anglican Church Cemetery in Belleville, Ontario. Documentation in the form of parish burial, marriage and baptismal records exists for the entire cemetery but only 55 individuals over 17 years of age are personally identified.

Seventeen traits of the bony pelvis were examined individually and jointly in order to establish which combination of pelvic features produces the most precise and accurate determinations of skeletal sex. The Phenice (1969) criteria were given special attention, both individually and as an entire system, due to the recent controversy surrounding the rate of their success. The method which employs all seventeen traits to determine sex was examined in order to establish whether it exhibits a propensity to misclassify one sex more frequently than the other. A similar approach was taken for seventeen cranial traits.

The accuracy, inaccuracy and bias of the Suchey-Katz

(1986), Suchey-Brooks (Suchey et al., 1988) and auricular surface (Lovejoy et al., 1985) systems of estimating age-at-death from the human skeleton were calculated and compared to the older methods of McKern and Stewart (1956) and Gilbert and McKern (1973). Age profiles produced from the parish records were used as a standard of comparison for the profiles generated by each of the age estimation systems examined here. The Jackes (1985) method of pubic age distribution was applied to each of the pubic symphysis age estimation techniques and the resulting age profiles were also compared to the parish register profile.

The results of this analysis indicate that both the pelvic and cranial methods of morphological sex determination produce sex ratios indistinguishable from the documented ratio. On an individual basis, however, the precision and accuracy of the pelvic approach is superior to that of the cranial approach. Features of the anterior pelvis are more useful than those of the posterior pelvis. No difference exists between criteria of the face and calvarium. Three features, posterior view of the sacrum, zygomatic root extension and occipital muscle markings, exhibit increased accuracy with age, although use of the complete pelvic or cranial trait list conceals these relationships and produces no age-related affect on accuracy.

Of the Phenice (1969) criteria, the ventral arc is the single most useful trait exhibiting both high precision (0%

intraobserver error) and high accuracy (86.9%). The subpubic concavity ranked second and the ischiopubic ramus last. In fact, the high intraobserver error of the ischiopubic ramus (11.3%) suggested that this feature is unreliable and that it should be excluded from the analysis. The greatest success with the Phenice criteria was obtained using the ventral arc and subpubic concavity in combination (92% accuracy). Despite being more closely related to McLaughlin and Bruce's (1990) 17th and 18th century English sample, the St. Thomas' material produced results similar to those of Sutherland and Suchey's (1991, 1987) North American sample.

Most of the cranial criteria employed in this study were previously recommended for use by Hrdlicka (Stewart, 1952). It is noteworthy that the accuracy obtained here (89.1%) is almost identical to what he achieved (90%) and the individual criteria rankings also compare favourably. These results suggest that, when employed as a group, some sexually dimorphic cranial features can be used successfully on a variety of populations. However, some variation in their individual usefulness exists.

NOT IN BIB — TOO BAD - it's much (1930's) clearer than Stewart.

Because so few errors are made when conducting pelvic and cranial morphological assessments of sex, determining the direction of error for each of these approaches is difficult. Only two incorrect assignments of sex were made using the pelvis. Both times females were mistaken for males. Seven misclassifications were made using the skull; four were male and three female. Although hardly conclusive, the evidence suggests that the initial hypotheses are correct. Pelvic assessments are biased in favour of males because pre-adolescent pelves are more male-like. In order to achieve a classic female pelvic form the adolescent female must undergo changes related to accommodating the birth of an infant. If a female pelvis does not obtain the classic form, it is considered male. On the other hand, cranial assessments favour females because pre-adolescent skulls are more female-like. If a skull never achieves a classic male appearance, it necessarily appears female.

When both precision and accuracy are taken into account, the auricular surface technique of estimating skeletal age-at-death produces the best results for individual cases. It has the lowest level of intraobserver error and its accuracy is comparable to both the Suchey-Brooks and Gilbert-McKern systems and is superior to both the Suchey-Katz and the McKern-Stewart methods. In addition, because there is a tendency for the pubic symphyses to become damaged in cemetery samples such as St. Thomas', as coffin lids decay and cave inward, the auricular surfaces are usually better preserved. In this case 249 adults were represented by at least one auricular surface, whereas only 174 individuals were represented by at least one pubic symphysis. One major problem with the auricular surface method is its tendency to overage individuals less than 40 years and underage individuals greater than 40 years. This pattern was identified by Murray and Murray (1991) and was confirmed in this study.

Problems with phases III (2) or IV (1) and VI were identified for both of the Suchey methods and contributed significantly to the levels of intraobserver error. The accuracy of these methods were acceptable mainly because of the wide age ranges covered by each phase. Practically speaking, the estimates provided by the Suchey methods are too broad to be useful in individual cases. When used to produce an age profile, however, the Suchey systems (with the Jackes [1985] distribution method) are fairly successful although the pattern is distorted by the difficulty associated with phases III, IV and IV.

6.2 CONCLUSIONS

Intraobserver, and whenever possible, interobserver error tests should be conducted as a routine part of every skeletal analysis. They can be used in studies involving morphological sex determination to identify criteria that cannot be reliably assessed. The degree of error can assist the researcher in deciding whether to exclude the problematic features from the analysis or to include them but to give them less weight than other traits. When employed in conjunction with estimations of age-at-death, intra- and interobserver error tests identify which methods a given researcher exhibits greater competence with and which methods are generally more reliable.

Precise and accurate sex profiles may be obtained through assessments of either the pelvis or the skull but on an individual basis the pelvic results are superior. Even when both types of information are available, the cranial data should serve only to confirm the pelvic assignment. When used in combination, the most reliable and accurate results may be obtained from an examination of the following six pelvic criteria: ventral arc, obturator foramen, true pelvis shape, sacrum shape, subpubic concavity and pubis shape. Readers are advised to avoid using most of the other features since they tend to be less precise and less accurate and therefore introduce greater ambiguity into the equation. If forced to rely upon a cranial assessment, the following eight criteria produced the least amount of intraobserver error and the greatest degrees of accuracy: supraorbital ridges, mastoid size, malar size and shape, occipital markings, chin form, general size, zygomatic root extension and mandible shape.

The results of the age estimation tests confirm many of the conclusions reached by other researchers (Jackes, in press; Meindl and Lovejoy, 1989; Meindl et al., 1983). None of the techniques examined in this investigation are precise and accurate enough to be employed exclusive of other systems. Although the auricular surface method produced the best results on an individual basis, the distinctive pattern of over- and underaging that is exhibited by this technique indicates a notable problem with the approach and culminates in an obviously inaccurate age profile for the entire sample (see Figure 15).

Unfortunately, in this analysis, the Suchey- Brooks and Suchey-Katz methods proved to be unreliable. Elimination of the difficulties associated with phases III (2), IV (1) and V1 of the Suchey-Katz system, identified in this study, would make this method much more useful than is currently indicated. Even so, the large age ranges produced by this technique make it more valuable for analysis of samples (when the Jackes method of distributing pubic ages is applied) than for analysis of individuals.

These findings suggest that Bouquet-Appel and Masset (1977) may be justified in bidding "farewell to paleodemography". Average age-at-death (e_o on life tables) depends upon accurate adult age estimations. Jackes (in press) observed that, all too frequently, archaeological samples produce results which deviate from model tables and it is assumed that cultural or postdepositional processes are responsible. Few people consider the alternative explanation; that adult underaging is the cause of the unexpected results.

Figures 12-14 and Table 20 illustrate the degrees of underaging for the methods examined in this study. Clearly average age-at-death calculations based on these results would be incorrect and Figures 15-23 indicate how inaccurate age profiles based on skeletal age estimations can be. All is not lost. Jackes (in press) has suggested some means of circumventing these problems so that paleodemographic analyses may continue. She notes, however, that the value of such approaches may be limited.

Due to the small sample sizes employed in many of the tests conducted in this investigation, the conclusions reached are by no means definitive. Rather, they are a starting point for further analyses. Future studies should attempt to identify both pelvic and cranial criteria that can be applied successfully to any population. It is also necessary to isolate and exclude criteria that are inherently difficult to assess. The effects of age on the accuracy of each trait should also be examined in greater detail and on larger samples. Although very few features in this study exhibited statistically significant differences in accuracy due to age, a number of age-related trends were identified. It is clear that too many suppositions concerning skeletal sex criteria have been made over the decades and it is time that we stopped perpetuating these assumptions and began methodologically testing them.

In addition, it appears that revisions to the auricular surface technique are required in order to eliminate the distinctive pattern of under- and overaging noted by Murray and Murray (1991) and confirmed here. It would also be useful to known whether other researchers find it difficult to differentiate between the Suchey methods' phases III (2) or VI

(1) and VI. Since errors of this type contributed significantly to the intraobserver error levels of this investigation, their elimination would result in a considerable improvement in the reliability of the method.

The importance of archaeologically excavated historic cemetery material is made evident by this study. Historic cemeteries provide the advantage of large samples containing individuals whose ages span the entire range of human life expectancy. Many of the other biases associated with medical school cadavers and dissection room samples can also be avoided through the use of historic archaeological samples.

The people buried at St. Thomas' were from various socioeconomic classes (DOA series I, part 35A; Boyce, 1991; 1967), consequently, many were unaffected by poverty and nutritional deficiencies. In the past, dissection room samples and medical school cadavers, composed largely of poor, elderly individuals (St. Hoyme and Īşcan, 1989), formed the basis of studies concerning sex determination and age estimation. Current investigators have speculated that the effects of poverty could easily have been mistaken for sex- and age-related characteristics (ibid). Greater emphasis on historic cemeteries will help eliminate this concern.

As our cities continue to change and expand, excavation and relocation of historic cemeteries will become increasingly frequent. Researchers must recognize the valuable contributions that analysis of such material can provide and must communicate the importance of skeletal assessments to those in charge of cemetery relocation projects.

REFERENCES CITED

Acsádi, G. and J. Nemeskeri

1970 — Determination of Sex and Age From Skeletal Finds. IN History of Human Life Span and Mortality, pp. 73-95. Budapest: Akademiai Kaido.

Anderson, B. E.

1990 — Ventral Arc of the Os Pubis: Anatomical and Developmental Considerations. American Journal of Physical Anthropology, 83: 449-458.

Baer, M. J.

1956 — Dimensional Changes in the Human Head and Face in the Third Decade of Life. American Journal of Physical Anthropology, 14(4): 557-575.

Bass, W. M.

1987 — Human Osteology. Columbia: Missouri Archaeological Society, Inc.

Baughan, B. and A. Demirjian

1978 — Sexual dimorphism in the Growth of the Cranium. American Journal of Physical Anthropology, 49: 383-390.

Bedford, M. E., K. F. Russell, and C. O. Lovejoy

1989 — The Utility of the Auricular Surface Aging Technique (abstract). American Journal of Physical Anthropology,

1989 — The Auricular Surface Aging Technique. Sixteen colour photographs with descriptions. Kent: Kent State University.

Bellstedt, I.

1969 The Parish Story. St. Thomas' Anglican Church 1818-1968. Belleville: St. Thomas' Anglican Church.

Bennett, K. A.

1979 Fundamentals of Biological Anthropology, pp. 482, 496. Dubuque: Wm. C. Brown Company.

Bocquet-Appel, J. and C. Masset

1982 Farewell to Paleodemography. Journal of Human Evolution, 11: 321-333.

Boyce, G. E.

1991 Belleville's Ethnic Mosaic. Paper Presented to the Ethnic Festival Committee: President's Dinner, January 23, 1991. Belleville, Ontario.

1967 Historic Hastings. Belleville: Ontario Intelligencer Limited.

Breathnach, A. S. (ed.)

1965 Frazer's Anatomy of the Human Skeleton. London: J. & A. Churchill Ltd.

Brooks, S. T.

1955 Skeletal Age at Death: The Reliability of Cranial and Pubic Age Indicators. American Journal of Physical Anthropology, 13: 567-597

Brooks, S. and J. M. Suchey

1990 Skeletal Age Determination Based on the Os Pubis: A Comparison of the Acsádi-Nemeskeri and Suchey-Brooks Methods. Human Evolution, 5(3): 227-238.

Browning, H. L.

1971 Migrant Selectivity and Growth of Large Cities in Developing Societies In Rapid Population Growth. pp. 273-313. Prepared by a Study Committee of the Office of the Foreign Secretary National Academy of Sciences. Baltimore: The Johns Hopkins Press.

Brothwell, D. R.

1981 Digging Up Bones, pp. 59-75. New York: Cornell University Press.

Budinoff, L. C. and R. G. Tague

1990 Anatomical and Developmental Bases for the Ventral Arc of the Human Pubis. American Journal of Physical Anthropolology, 82: 73-79.

Buikstra, J. E. and J. H. Mielke

1985 Demography, Diet, and Health. IN The Analysis of Prehistoric Diets. R. I. Gilbert and J. H. Mielke, eds. pp. 359-422.

Buschang, P. H., R. M. Baume and G. G. Nass

1983 Craniofacial Growth and Size Patterns During Post Natal Development. Growth, 47: 217-223.

1983 A Craniofacial Growth Maturity Gradient for Males and Females Between Four and Sixteen Years of Age. American Journal of Physical Anthropology, 61(3): 373-382.

Coleman, W. H.

1969 Sex Differences in the Growth of the Human Bony Pelvis. American Journal of Physical Anthropology, 31: 125-152.

Diocese of Ontario Archives, General Synod Archives, Anglican Archives, Toronto, Ontario [DOA].

Microfilm #22, Series I Parish Registers, Part 35A: St. Thomas' Church, Belleville - County of Hastings 1821-1872.

DRAKE (1974) ? where is he?

El-Najjar, M. Y. and K. R. Mc Williams

1978 Forensic Anthropology. Springfield: Charles C. Thomas.

Enlow, D. H.

1982 Handbook of Facial Growth 2nd Edition. Toronto: W. B. Saunders Company.

1975 Handbook of Facial Growth. Toronto: W.B. Saunders Company.

Gilbert, B. M.

1972 Misapplication to Females of the Standard for Aging the Male Os Pubis. American Journal of Physical Anthroplogy, 38: 39-40.

Gilbert, B. M. and T. W. McKern

1972 A Method for Aging the Female Os Pubis. American Journal of Physical Anthropology, 38: 31-38.

Giles, E. and O. Elliot

1963 Sex Determination by Discriminant Function Analysis of Crania. American Journal of Physical Anthropology, 21: 53-68.

Goldstein, M. S.

1936 Changes in Dimensions and Form of the Face and Head With Age. American Journal of Physical Anthropology (old series), 22: 37-89.

Grigg, D. B.

1977 E. G. Ravenstein and the "Laws of Migration". Journal of Historical Geography 3(1): 41-54.

Hanihara, K. and T. Suzuki

1978 Estimation of Age from the Pubic Symphysis by Means of Multiple Regression Analysis. American Journal of Physical Anthropology, 48: 233-240.

Henke, W.

1977 On the Method of Discriminant Function Analysis for Sex Determination of the Skull. Journal of Forensic Sciences, 6: 95-100.

Howells, W. W.

1969 The Use of Multivariate Techniques in the Study of Skeletal Populations. American Journal of Physical Anthropology, 31: 311-314.

HRDLICKA? GOOD PRIMARY SOURCE!

Israel, H.

1979 The Dichotomous Pattern of Craniofacial Expansion During Aging. American Journal of Physical Anthropology, 47: 47-52.

1973 Age Factor and the Pattern of Change in Craniofacial Structures. American Journal of Physical Anthropology, 39: 111-128.

Jackes, M. K.

In Press Paleodemography: Problems and Techniques IN The Skeletal Biology of Past Peoples: Advances in Research Methods. S. R. Saunders and M. A. Katzenberg (eds). New York: Alan R. Liss.

1985 Pubic Symphysis Age Distributions. American Journal of Physical Anthropology, 68: 281-299.

Kajanoja, P.

1966 Sex Determination of Finnish Crania by Discriminant Function Analysis. American Journal of Physical Anthropology, 24:29-34.

Katz, D. and J. M. Suchey

1989 Race Differences in Pubic Symphyseal Aging Patterns in the Male. American Journal of Physical Anthropology, 80: 167-172.

1986 Age Determination of the Male Os Pubis. American Journal of Physical Anthropology, 69: 427-435.

Keen, J. A.

1950 A Study of the Differences Between Male and Female Skulls. American Journal of Physical Anthropology, 8: 65-79.

Kelley, M. A.

1979 Sex Determination with fragmented Skeletal Remains. Journal of Forensic Sciences, 24: 154-158.

1979 Parturition and Pelvic Changes. American Journal of Physical Anthropology, 51: 541-546.

1978 Phenice's Visual Sexing Technique for the Os Pubis: A Critique. American Journal of Physical Anthropology, 48: 121-122.

Kobayashi, K.

1967 Trend in the Length of Life Based on Human Skeletons from Prehistoric to Modern Times in Japan. Reprinted from Journal of the Faculty of Science, University of Tokyo Sec. V. 3(2): 107-162.

Krogman, W. M. and M. Y. İşcan

1986 The Human Skeleton in Forensic Medicine. Springfield: Charles C. Thomas.

Lanphear, K. M.

1989 Testing the Value of skeletal samples in Demographic Research: A Comparison with Vital Registration Samples. International Journal of Anthropology, 4 (3): 185-193.

Lovejoy, C. O., R. S. Meindl, R. P. Mensforth and T. J. Barton

1985 Multifactorial Determination of Skeletal Age at Death: A Method and Blind Tests of Its Accuracy. American Journal of Physical Anthropology, 68: 1-14.

Lovejoy, C. O., R. S. Meindl, T. R. Pryzbeck, R. P. Mensforth

1985 Chronological Metamorphosis of the Auricular Surface of the Ilium: A New Method for the Determination of Adult Skeletal Age at Death. American Journal of Physical Anthropology, 68: 16-28.

Lovell, N. C.

1989 Test of Phenice's Technique for Determining Sex From the Os Pubis. American Journal of Physical Anthropology, 79: 117-120.

MacLaughlin, S. M. and M. F. Bruce

1990 The Accuracy of Sex Identification in European Skeletal Remains Using the Phenice Characters. Journal of Forensic Sciences, 35: 1384-1392.

McKern, T. W.

1957 Estimation of Skeletal Age From Combined Maturational Activity. American Journal of Physical Anthropology, 15: 399-408.

McKern, T. W. and T. D. Stewart

1957 Skeletal Age Changes in Young American Males, Technical Report EP-45. Natick, MA: US Army Research and Development Center, Environmental Protection Research Division.

McKillop, H., S. Marshall, G. Boyce and S. Saunders

1989 Excavations at St. Thomas' Church, Belleville, Ontario: A Nineteenth Century Cemetery. Paper Presented at the Ontario Archaeological Society Symposium 1989. London, Ontario.

Meindl, R. S. and C. O. Lovejoy

1989 Age Changes in the Pelvis: Implications for Paleodemography. IN Age Markers in the Human Skeleton. M. Y. Iscan (ed), pp. 137-168. Springfield: Charles C. Thomas.

Meindl, R. S., C. O. Lovejoy, R. P. Mensforth

1983 Skeletal Age at Death: Accuracy of Determination and Implications for Human Demography. Human Biology, 55(1): 73-87.

Meindl, R. S., C. O. Lovejoy, R. P. Mensforth and L. Don Carlos

1985 Accuracy and Direction of Error in the Sexing of the Skeleton: Implications for Paleodemography. American Journal of Physical Anthropology, 68: 79-85.

Meindl, R. S., K. F. Russell, C. O. Lovejoy

1991 Reliability of Age at Death in the Hamann-Todd Collection: Validity of Subselection Procedures Used in Blind Tests of the Summary Age Technique. American Journal of Physical Anthropology, 83: 349-357.

Mika, N. and H. Mika

1986 — Belleville, the Seat of Hastings County. Belleville: Mika Publishing Co.

Molto, J. E.

1979 — The Assessment and Meaning of Intraobserver Error in Population Studies Based on Discontinuous Cranial Traits. American Journal of Physical Anthropology, 51: 333-344.

Murray, K. A. and T. Murray

1991 — Test of the Auricualr Surface Ageing Technique. Journal of Forensic Sciences, 36(4): 1162-1169.

Nanda, R. S.

1955 — The Rates of Growth of Several Facial Components measured from Serial Cephalometric Roentgenograms. American Journal of Orthodontics, 41: 658-673.

Nichol, C. R. and C. G. Turner

1986 — Intra- and Interobserver Concordance in Classifying Dental Morphology. American Journal of Physical Anthroplogy, 69: 299-315.

Phenice, T. W.

1969 — A Newly Developed Visual Method of Sexing the Os Pubis. American Journal of Physical Anthropology, 30: 297-302.

Richman, E. A., M. E. Michel, F. P. Schulter- Ellis, R. S. Corruccini

1979 — Determination of Sex by Discriminant Function Analysis of Postcranial Skeletal Measurements. Journal of Forensic Sciences, 24: 159-163.

St. Hoyme, L. E.

1984 Sex Differentiation in the Posterior Pelvis. Collegium Antropologica, 8: 139-153.

St. Hoyme, L. E. and M. Y. İşcan

1989 Determination of Sex and Race: Accuracy and Assumptions. IN Reconstruction of Life from the Skeleton, pp 53-93. New York: Alan R. Liss, Inc.

Siegel, S.

1956 Nonparametric Statistics for the Behavioural Sciences. Toronto: McGraw Hill Book Company.

Singh, S. and B. R. Potturi

1978 Greater Sciatic Notch in Sex Determination. Journal of Anatomy, 125: 619-624.

Sokal, R. P. and F. J. Rohlf

1981 Biometry (Second Edition). New York: W. H. Freeman and Company.

Steele, D. G. and C. A. Bramblett

1988 The Anatomy and Biology of the Human Skeleton. College Station: Texas A&M University Press.

Stewart, T. D.

1979 Essentials of Forensic Anthropology. Springfield: Charles C. Thomas.

1957 Distortion of the Pubic Symphyseal Surface in Females and its Effect on Age Determination. American Journal of Physical Anthropology, 15: 9-18.

1954 Sex Determination of the Skeleton By Guess and By Measurement. American Journal of Physical Anthropology, 12: 385-392.

1952 Hrdlicka's Practical Anthropometry, pp. (ed.) 127-134. Philadelphia: The Wistar Institute of Anatomy and Biology.

Straus, W. Jr.

1927 The Human Ilium: Sex and Stock. American Journal of Physical Anthropology, 11 (1): 1-28.

Suchey, J. M.

1979 Problems in the Aging of Females Using the Os Pubis. American Journal of Physical Anthropology, 51: 467-470.

Suchey, J. M., S. T. Brooks and D. Katz

1988 Instructions for Use of the Suchey-Brooks System for Age Determination of the Female Os Pubis. Instructional materials accompanying female pubic symphyseal models of the Suchey-Brooks system. Distributed by FRANCE CASTING (Diane France, 2190 West Drake Road, Suite 259, Fort Collins, Colorado 80526).

Suchey, J. M. and D. Katz

1986 Skeletal Age Standards Derived from an Extensive Sample of Modern Americans. Handout accompanying paper presented at the 55th Annual Meeting of the American Association of Physical Anthropologists, Albuquerque, New Mexico, 10 April 1986. American Journal of Physical Anthropology, 69: 269 (abstract).

Suchey, J. M., D. V. Wiseley and D. Katz

1986 Evaluation of the Todd and McKern-Stewart Methods for Aging the Male Os Pubis. IN Forensic Osteology. K.J. Reichs (ed) pp. 33-67. Springfield: Charles C. Thomas.

Sutherland L. D. and J. M. Suchey

1991 Use of the Ventral Arc in Sex Determination. Journal of Forensic Sciences, 36(2): 501-511.

1987 Use of the Ventral Arc in Sex Determination of the Os Pubis. Paper Presented at the 39th Annual Meeting of the American Academy of Forensic Sciences, San Diego, California.

Tague, R. G.

1989 Variation in Pelvic Size Between Males and Females. American Journal of Physical Anthropology, 80: 59-71.

1988 Bone resorption of the Pubis and Preauricular Area in Humans and Nonhuman Mammals. American Journal of Physical Anthropology, 76: 251-267.

Tallgren, A.

1974 Neurocranial Morphology and Ageing - A Longitudinal Roentgen Cephalometric Study of Adult Finnish Women. American Journal of Physical Anthropology, 41: 285-294.

Todd, T. W.

1920 Age Changes in the Pubic Bone. American Journal of Physical Anthropology, 3(3): 285-327.

Ubelaker, D. H.

1984 Human Skeletal Remains, pp. 41-68. Washington: Taraxacum.

Washburn, S.

1948 Sex Differences in the Pubic Bone. American Journal of Physical Anthropology, 7 (3): 425-432.

Weiss, K. M.

1972 On the Systematic Bias in Skeletal Sexing. American Journal of Physical Anthropology, 37: 239-250.

Wilkinson, L.

1988 Systat: The System for Statistics. Evanston, IL: Systat, Inc.

Wood, B. A. and A. T. Chamberlain

1986 The Primate Pelvis : Allometry or Sexual Dimorphism? Journal of Human Evolution, 15: 257-263.

Workshop of European Anthropologists

1980 Recommendations for Age and Sex Diagnoses of Skeletons. Journal of Human Evolution, 9: 517-549.

APPENDIX 1.

ST. THOMAS' CHURCH
BELLEVILLE, ONTARIO

YEAR OF EVENT	BAPTISMS	MARRIAGES	BURIAL	TOTAL
1821	8	9	4	21
1822	37	30	2	69
1823	38	36	8	82
1824	30	36	7	73
1825	57	30	11	98
1826	38	52	12	102
1827	42	30	9	81
1828	58	35	19	142
1829	59	34	11	104
1830	60	42	11	113
1831	38	20	9	67
1832	40	9	11	60
1833	42	16	24	82
1834	27	16	16	59
1835	28	10	14	52
1836	51	18	19	38
1837	34	15	16	65
1838	53	16	18	87
1839	71	19	17	107
1840	60	12	19	91
1841	103	50	31	184
1842	136	49	19	204
1843	122	45	28	195
1844	126	47	21	194
1845	122	36	14	172
1846	124	56	28	208
1847	116	34	46	196
1848	150	33	40	223
1849	83	31	34	148
1850	91	43	35	169
1851	116	38	33	197
1852	71	34	43	148
1853	77	30	49	156
1854	93	38	80	211
1855	86	37	48	171
1856	105	37	29	171
1857	101	33	39	173
1858	103	25	32	160
1859	114	32	38	184
1860	94	22	33	154
1861	100	29	38	166
1862	145	29	57	222
1863	138	27	70	235
1864	141	26	47	214
1865	58	17	49	124
1866	57	9	25	91
1867	66	19	38	123
1868	60	15	47	122

1869	71	13	48	132
1870	67	21	50	138
1871	68	31	60	159
1872	48	14	36	98
1873	58	18	30	106
1874	49	24	7	90

I would like to thank Ms. Carol Devito for the baptismal and marriage data.

APPENDIX 2.
EXPLANATION OF STATISTICS

BINOMIAL TEST

The binomial test indicates whether it is reasonable to believe that the observed frequencies could have been drawn from a population having a specified value of P. It is used when data are in two discrete categories, e.g. odd/even or male/female, and the design is the one-sample type (Siegel, 1956). It is calculated as follows:

$$z = \frac{(x \pm .5) - NP}{\sqrt{NPQ}}$$

where x = the number of objects in one of the two categories

N = the number of cases

P = the proportion of cases expected in one category

Q = the proportion of cases expected in the other category

If x<NP add .5
If x>NP subtract .5

Once z is obtained, refer to a z-table (normal distribution) to find the probability. For more information see Siegel (1956: 36-42; z-table on page 247).

CHI SQUARE

Chi square is used to test the hypothesis that two groups differ with respect to some characteristic and therefore with respect to the relative frequency with which members of the group fall in several categories (Siegel, 1956). The proportion of cases in one group is compared to the proportion of cases in the other group through the following formula:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(Oij - Eij)^2}{Eij}$$

where Oij = observed number of cases in the (i)th row of the (j)th column

Eij = number of cases expected under the null hypothesis to be in the (i)th row of the (j)th column

$\sum_{i=1}^{r} \sum_{j=1}^{k}$ directs one to sum over all rows (r) and all columns (k)

The critical values for X^2 can be found in tables such as those in Siegel (1956: 249). If a calculated value for X^2 is equal to or greater than the value in the table for a particular level of statistical significance and with a particular degree of freedom, than the null hypothesis may be rejected at that level of significance. Degrees of freedom reflect the number of observations that are free to vary after particular restrictions have been placed on the data.

df= (r-1)(k-1), where r= number of rows and k= number of columns in a contingency table

For more information see Siegel (1956: 104-111).

FISHER'S EXACT PROBABILITY TEST

Fisher's exact is a nonparametric test used to analyze discrete data (nominal or orcinal) when two independent samples are small in size and the scores fall into one or the other of two mutually excluseive categories, e.g. correct sex assignment/incorrect sex assignment (Siegel, 1956). This test determines whether the proportions in the categories differ between the two samples. The exact probability of the observed occurrence is calculated as follows:

$$p = \frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{N!\;A!\;B!\;C!\;D!}$$

where A,B,C,D are the individual cell frequencies of a 2x2 contingency table

N = the total number of cases

! = factorial which means the value is multiplied by each value lower than itself until (1) is reached $x!=x(x-1)(x-2)\ldots(2)(1)$

For more information see Siegel (1956: 96-104)

LOG-LINEAR TEST WITH CHI SQUARE RESIDUALS

A detailed explanation of log-linear models is beyond the scope of this paper. Readers are refered to Norusis (1985: 298-366) for more information.

Log-linear models provide a method of analyzing categorical data that is capable of revealing the nature of complex interrelationships among the variables in a multiway crosstabulation (Norusis, 1985). Instead of employing actual counts, log-linear tests use the natural logs of all cell frequencies in order to obtain a linear model. The observed values are compared to the expected values and standardized residuals are computed using the formula:

$$\frac{(\text{Observed} - \text{Expected})}{\sqrt{\text{Expected}}}$$

As a general rule, standard residuals that exceed 1.96 (the 95th percentile of the normal distribution) indicate that there is a significant difference between the expected and observed values. Positive residuals indicated an actual value in excess of the expected and negative residuals indicate an under-representation of the actual value according to expectations.

APPENDIX 3.

PHASE MEANS, STANDARD DEVIATIONS AND RANGES FOR PUBIC SYMPHYSIS AGE ESTIMATION SYSTEMS

Suchey-Katz (male)

Phase Range	Mean	S.D.	95% Range
I	18.5	2.1	15-23
II	23.4	3.6	19-34
III	28.7	6.5	21-46
IV	35.2	9.4	23-57
V	45.6	10.4	27-66
VI	61.2	12.2	34-86

Suchey-Brooks (female)

Phase	Mean	S.D.	95%
I	19.4	2.6	15-24
II	25.0	4.9	19-40
III	30.7	8.1	21-53
IV	38.2	10.9	26-70
V	48.1	14.6	25-83
VI	60.0	12.4	42-87

Source: Suchey et. al., 1988.

McKern-Stewart (male)

Score Range	Mean	S.D.	95% Range
0	17.29	.49	-17
1-2	19.04	.79	17-20
3	19.79	.85	18-21
4-5	20.84	1.13	18-23
6-7	22.42	.99	20-24
8-9	24.14	1.93	22-28
10	26.05	1.87	23-28
11-13	29.18	3.33	23-39
14	35.84	3.89	29+
15	41.00	6.22	36+

Source: McKern and Stewart, 1957

Gilbert-McKern (female)

Score	Mean	S.D.	95%
0	16.00	2.82	14-18
1	19.80	2.62	13-24
2	20.15	2.19	16-25
3	21.50	3.10	18-25
4-5	26.00	2.61	22-29
6	29.62	4.43	25-36
7-8	32.00	4.55	23-39
9	33.00	7.75	22-40
10-11	36.90	4.94	30-47
12	39.00	6.09	32-52
13	47.75	3.59	44-54
14-15	55.71	3.24	52-59

Gilbert and McKern, 1973