



Universiteit
Leiden
The Netherlands

Sexual dimorphism in the talus: The talus as a tool for sex estimation in a post-medieval Dutch population.

Piso, Nina

Citation

Piso, N. (2022). *Sexual dimorphism in the talus: The talus as a tool for sex estimation in a post-medieval Dutch population.*

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

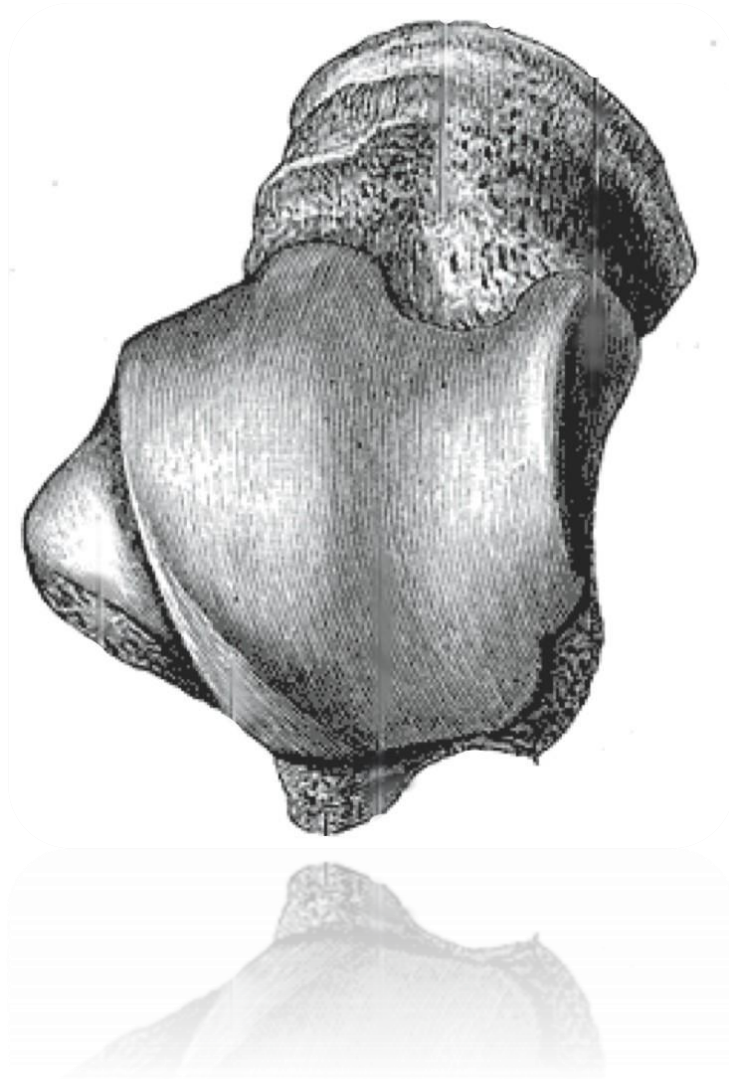
Downloaded from: <https://hdl.handle.net/1887/3448981>

Note: To cite this publication please use the final published version (if applicable).

Sexual dimorphism in the talus

The talus as a tool for sex estimation in a post-medieval Dutch population.

Nina Piso



Cover image: https://radiopaedia.org/images/53715196?case_id=83339, modified by
Piso, N.

Sexual dimorphism in the talus

The talus as a tool for sex estimation in a post-medieval
Dutch population.

Nina Piso

Thesis Archaeological Science (1084VTSY_2122_HS)

Supervisor: Dr. S.A. Schrader

Human Osteology and Funerary Archaeology

University of Leiden, Faculty of Archaeology

Leiden, 31-05-2022, final version

Table of Contents

1.	Introduction	8
1.1	Sex estimation methods.....	8
1.2	Talus	9
1.3	Research aims and questions.....	10
1.4	Thesis structure.....	11
2.	Background	12
2.1	Sexual dimorphism of the skeleton.....	12
2.2	Morphological sex estimation.....	13
2.2.1	<i>Pelvis</i>	13
2.2.2	<i>Skull</i>	16
2.3	Metric methods.....	18
2.3.1	<i>Skull</i>	18
2.3.2	<i>Pelvis</i>	19
2.3.3	<i>Humerus</i>	20
2.3.4	<i>Radius and ulna</i>	20
2.3.5	<i>Femur</i>	21
2.3.6	<i>Patella</i>	21
2.3.7	<i>Calcaneus</i>	21
2.4	Talus	22
2.4.1	<i>Previous studies</i>	23
3.	Materials and methods.....	31
3.1	Materials	31
3.1.1	<i>Historical context Middenbeemster</i>	31
3.1.2	<i>Archaeological context Middenbeemster</i>	32
3.2	Methods.....	34
3.2.1	<i>Data selection</i>	34
3.2.2	<i>Data collection</i>	34
3.2.3	<i>Measurements</i>	35
3.2.4	<i>Statistical analysis</i>	37
4.	Results.....	39
4.1	Intra-observer error	39
4.2	Outliers.....	40
4.3	Normal distribution.....	41

4.4	Descriptive statistics and Student's t-test.....	41
4.5	Univariate analysis	42
4.6	Multivariate analysis	43
4.6.1	<i>Direct discriminant functions</i>	43
4.6.2	<i>Stepwise discriminant function</i>	45
4.7	Comparison populations	45
5.	Discussion.....	47
5.1	Repeatability	47
5.2	Univariate analysis and sexual dimorphism.....	47
5.3	Multivariate analysis versus univariate analysis	49
5.4	Population specificity	51
5.4.1	<i>Peckmann et al. (2015)</i>	52
5.4.2	<i>Lee et al. (2011)</i>	52
5.4.3	<i>Bidmos and Dayal (2003)</i>	53
6.	Conclusion.....	55
6.1	Sub-questions.....	56
6.1.1	<i>Which measurements are the most sexually dimorphic and thus most reliable in sex estimation?</i>	56
6.1.2	<i>Does a combination of measurements perform better than a single measurement in sex estimation?</i>	56
6.1.3	<i>How accurate are formulas of Bidmos & Dayal (2003), Lee et al. (2012), and Peckmann et al. (2015) when using the data of the Dutch population? Is it indeed population specific?</i>	57
6.2	Main research question	57
6.2.1	<i>How effective is the talus metric method in estimating sex in a post-medieval Dutch population?</i>	57
6.3	Limitations.....	58
6.4	Future research	58
	Abstract.....	60
	Bibliography	61
	Other	64

List of figures

Figure 1.1: Position of the talus in the right foot, medial side (White & Folkens, 2005, p. 291. Modified by Piso, N.).	10
Figure 2.1: The Phenice traits in the pubis region per sex (Shook et al., 2019, p. 556).	14
Figure 2.2: Scores of the greater sciatic notch; (1) female, (2) probable female, (3) indeterminate, (4) probable male, (5) male (Buikstra & Ubelaker, 1994, p. 18).	16
Figure 2.3: Scores of the preauricular sulcus; (1) female, (2) probable female, (3) indeterminate, (4) probable male, (5) is absent but can be scored when preauricular surface is not visible scoring it as male (Buikstra & Ubelaker, 1994, p. 18).	16
Figure 2.4: Cranial and mandibular traits as defined by Buikstra and Ubelaker with scale (White & Folkens, 2005, pp. 390–391).	18
Figure 2.5: The tibia as situated in the foot (www.orthoinfo.aaos.org/en/diseases--conditions/talus-fractures).	23
Figure 2.6: The measurements of the talus. Talar length (TL, 1), talar width (TW), talar height (TH, 3), trochlear length (TrL, 4), trochlear breadth (TrB, 5), head-neck length (HNL, 6), head height (HH, 7), length of the posterior articular surface (LPAS, 8), breadth of the posterior articular surface (BPAS, 9). A, C: superior view, B: lateral view, D: inferior view (Lee et al., 2012, p. 167).	26
Figure 2.7: Measurements NASH (line 11) and NASW (line 12) (Abd-Elaleem et al., 2012, p. 71).	27
Figure 3.1: Map showing the division of the Beemster into square plots, 1658 (van Spelde & Hoogland, 2018, p. 309).	31
Figure 3.2: Location of the excavation in Middenbeemster (Hakvoort, 2013, p. 11).	33
Figure 3.3: Wooden coffin with the lid pressed inside of the coffin due to the weight of the soil (van Spelde & Hoogland, 2018, p. 319).	33
Figure 3.4: The measurements of the talus. Talar length (TL, 1), talar width (TW), talar height (TH, 3), trochlear length (TrL, 4), trochlear breadth (TrB, 5), head-neck length (HNL, 6), head height (HH, 7), length of the posterior articular surface (LPAS, 8), breadth of the posterior articular surface (BPAS, 9). A, C: superior view, B: lateral view, D: inferior view (Lee et al., 2012, p. 167).	36
Figure 4.1: Boxplot of outlier individual 95.	40

List of tables

Table 2.1: Traits and corresponding weights of the pelvis as recommended by the WEA (after WEA, 1980, p. 518).....	15
Table 2.2: Traits and corresponding weights of the cranium and mandible as recommended by the WEA (after WEA, 1980, p. 523).	17
Table 2.3: Abbreviations and corresponding measurements of all discussed studies (compiled by Piso, N.).	23
Table 2.4: Measurements and their definitions of all discussed studies (compiled by Piso, N.).....	24
Table 2.5: Studies and their best outcomes (compiled by Piso, N.).	30
Table 4.1: Intraclass correlation coefficient and the average difference between repeated measurements.....	39
Table 4.2: Individuals with their z-score.	40
Table 4.3: KS p-value = Kolmogorov-Smirnov p-value, SW p-value = Shapiro Wilk p-value, * = lower bound of the true significance.	41
Table 4.4: Descriptive statistics, Student's t-test and the significance (p-value).	42
Table 4.5: Accuracies and demarcation points resulting from univariate analysis.	43
Table 4.6: Direct multivariate discriminant function analysis. M = male, F = female, Avg = average.....	44
Table 4.7: Stepwise multivariate discriminant function analysis. M = male, F = female, Avg = average.	45
Table 4.8: Comparison of the accuracies of the original study versus the accuracies with the Dutch data.	46
Table 5.1: The highest scoring leave-one-out accuracies in descending order.	48
Table 5.2: Functions and their original and leave-one-out accuracies.	51

Intentional blank page

1. Introduction

Sex is a pivotal part of one's biological profile. Sex is not to be confused with gender, two concepts that are often used interchangeably but should not be. There is a clear distinction between the two, gender is a sociocultural construct while sex is determined on a biological basis (DiGangi & Moore, 2012, p. 92; Walker & Cook, 1998, p. 256; White & Folkens, 2005, p. 385). Besides sex, the main features of an osteoarchaeological biological profile are age-at-death, stature and ancestry. If skeletal elements show signs of pathological conditions or trauma the observed lesions can also be a subsidiary part of the biological profile (DiGangi & Moore, 2012, p. 63; Reichs, 1998, p. 163; Scheuer, 2002, pp. 298–299). To create the osteoarchaeological biological profile sex estimation is generally the first step that is taken. Without it, other elements of the biological profile (e.g. stature and age-at-death) cannot be estimated as they are based on sex-specific regression formulas (Reichs, 1998, p. 163).

In both archaeological contexts as in forensic contexts establishing the osteoarchaeological biological profile is vital. In the forensic field the profile can help identifying an individual when only their skeletal remains are available, for example when a mass disaster has occurred or when a clandestine grave is found (Blau & Ubelaker, 2016, p. 261; Scheuer, 2002, p. 297). Estimating sex narrows down a missing-person search by 50% (Blau & Ubelaker, 2016, p. 261; DiGangi & Moore, 2012, p. 91). In archaeology on the other hand, establishing an osteoarchaeological biological profile of an individual can give an understanding of the demography and social construction of the population in question (Blau & Ubelaker, 2016, p. 261; DiGangi & Moore, 2012, p. 91).

1.1 Sex estimation methods

Sex can be estimated with a multitude of methods. One of these methods is aDNA and has taken hold in the scientific community in recent years. The term aDNA is short for 'ancient deoxyribonucleic acid'. Ancient DNA cannot only determine sex, it can also inform us on the ancestry, disease status, and identity of an individual. To obtain and analyse DNA there are generally three stages. The first step is to extract and isolate the DNA, this is done by reducing bone or teeth to powder. Proteins and other compounds are then chemically removed from the powder to concentrate the DNA. Secondly,

through the use of polymerase chain reaction, a section of the DNA is amplified. The last step is to analyse the remaining DNA sample through its nucleotide sequence (White & Folkens, 2005, pp. 346–347). Although aDNA is one of the most exact methods to estimate sex it has a number of disadvantages. It can be time-consuming, there is a risk of cross-contamination and inhibitors, in some instances not enough DNA can be extracted, and it is a destructive method. However, the main reason aDNA is not the standard method is because it is still fairly expensive. Therefore, the more 'traditional' morphological and metric methods are still preferred by a majority of researchers (Jerković et al., 2018, p. 44).

The 'traditional' morphological methods are based on the visual assessment of certain features of bones. The skull and pelvis display differences in traits between males and females which can be observed and scored. Although fairly quick, this type of method requires a certain level of expertise, is comparatively quite a subjective method, and requires fairly complete skeletal elements. Nevertheless, the pelvis alone can obtain accuracy rates up to 96% (DiGangi & Moore, 2012, p. 96; Phenice, 1969, p. 300). When these bones are not available or are highly damaged one can opt for metric methods.

Metric methods can be explained through sexual dimorphism. In short sexual dimorphism is the variation between males and females. This variation can be seen in the size of various skeletal elements as females tend to be smaller in size than males. The size of the bones can be measured and through the use of statistical techniques the sex of an individual can be estimated. Advantages of metric methods are that a greater variety of bones can be used, expertise is less necessary, and it can be easily replicated in order to validate results. It does have one major disadvantage, it is population specific (Bidmos & Dayal, 2003, p. 322). This means that every bone that shows promise of sexual dimorphism needs to be tested per population. Both morphological and metric methods as well as sexual dimorphism will be further clarified in chapter two.

1.2 Talus

Over the years more bones are being tested for their validity in sex estimation. This is necessary because one cannot rely on the presence of a complete and perfectly preserved individual. In a worst case scenario only a few bone fragments are preserved and available for analysis (Navega et al., 2014, p. 651). The talus is a foot bone, also known as the ankle bone (figure 1.1), that is often fairly well preserved due to its

compact and robust nature. Therefore, the bone has a high recovery rate within archaeological and forensic contexts making it a perfect bone for sex estimation (Alonso-Llamazares & Pablos, 2019, p. 4928; Gualdi-Russo, 2007, p. 151; Indra et al., 2021, p. 556; Navega et al., 2014, p. 651).

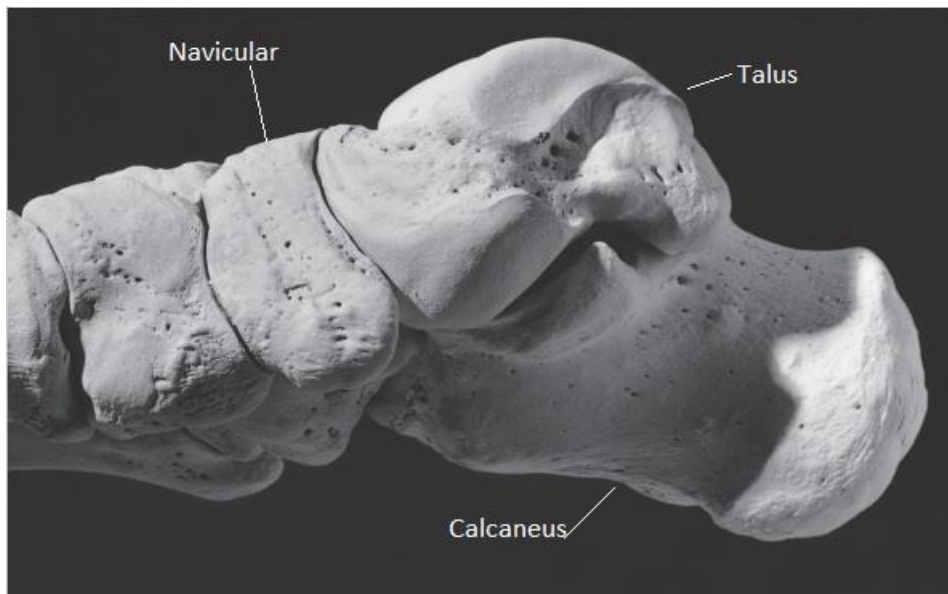


Figure 1.1: Position of the talus in the right foot, medial side (White & Folkens, 2005, p. 291. Modified by Pisco, N.).

Steele was the first to investigate if footbones could aid in the estimation of sex using the two biggest foot bones, the talus and calcaneus (Steele, 1976, p. 581). Steele used the American Terry Collection and was able to obtain accuracy rates of 79% to 89% (Alonso-Llamazares & Pablos, 2019, p. 4928; Steele, 1976, p. 581). After this study by Steele many other studies followed yielding accuracy rates of over 80% (Indra et al., 2021, p. 556). These studies will be further elaborated on in chapter two.

1.3 Research aims and questions

Measurements of the talus have been used on numerous populations. Studies have for example been done on a Thai population (Mahakkanukrauh et al., 2014), a Greek population (Peckmann et al., 2015), and an Italian population (Gualdi-Russo, 2007). To date no study has been done on a Dutch population. Therefore, the aim of this thesis is to research if measurements of the talus can be good indicators of sex in adults of a Dutch population. If so, it can prove useful if and when the usual sex estimation bones from an individual are missing or damaged or as supporting evidence. This research uses the collection of Middenbeemster currently housed at the Faculty of Archaeology at the University of Leiden in the Netherlands. This collection is dated to the post-medieval

period with burials ranging from 1615 to 1866 and it has accompanying archival data which is why this collection has been selected for this research (Hakvoort, 2013, p. 9).

Main research question:

- How effective is the talus metric method in estimating sex in a post-medieval Dutch population?

Sub-questions:

- Which measurements are the most sexually dimorphic and thus most reliable in sex estimation?
- Does a combination of measurements perform better than a single measurement in sex estimation?
- How accurate are formulas of Bidmos & Dayal (2003), Lee et al. (2012) and Peckmann et al. (2015) when using the data of the Dutch population? Is it indeed population specific?

1.4 Thesis structure

This thesis is divided into a total of six chapters. This first chapter, the introduction, has introduced the topic of research, its aims and its research questions. The second chapter will provide the background of this research. It will elaborate further on sexual dimorphism as well as on the sex estimation methods that have been briefly described in chapter one. A distinction will be made between metric and morphological (non-metric) methods and it will elaborate on the accuracy of these methods. It will also discuss studies that use the talus for sex estimation in other populations. In chapter three the materials and methods that are used in this study will be discussed. It will provide a background on the skeletal collection of Middenbeemster and the sample size that has been chosen for this research. Additionally, the measurements of the talus will be provided and discussed. Furthermore, the statistical methods that are utilised in this thesis will be outlined. Chapter four will present the results of the statistical methods that have been executed. These results will be interpreted and discussed in chapter five. Lastly, chapter six will be the final chapter in which a conclusion to this research will be given on the basis of the formulated sub-questions and main research question. In addition, it will discuss the limitations of this study and the potential it might have for future research.

2. Background

As previously stated, sex estimation is an important part of creating the osteoarchaeological biological profile of an individual and of identifying unidentified human remains (Colman et al., 2018, p. 268.e2). There are two types of 'traditional' methods that are most commonly applied to estimate sex. The visual, more subjective morphological methods and the statistics based metric methods (DiGangi & Moore, 2012, p. 92). Morphological methods are based on the pelvis and skull. Metric methods, on the other hand, are not only created for the pelvis and skull but also for many other skeletal elements of the human body. This chapter will discuss literature on both these methods providing examples for the metric studies. This is however not a conclusive list as many more studies on bones and populations exist. Separately, studies on the talus in several populations will be explored. Preceding these specified methods and studies, sexual dimorphism will be addressed as it is the reason sex can be estimated from skeletal elements.

2.1 Sexual dimorphism of the skeleton

Almost all species show a degree of sexual dimorphism. This means that males and females of the same species can differ in shape, size or traits (Isaac, 2005, p. 101). The human species also displays variation. This variation can be seen in both size and shape. Females tend to be smaller in both bone and tooth size in comparison to males and the shape can vary in certain elements of the skeleton (White & Folkens, 2005, p. 32).

This variation is not yet visible in birth but develops as the individual matures (White & Folkens, 2005, p. 385). Males and females grow and develop at different rates due to factors either inside or outside of the body (DiGangi & Moore, 2012, p. 93; Veldhuis et al., 2005, p. 114). Intrinsic factors such as hormone levels, in particular those put out by the pituitary gland and gonad, play a vital role in sexual dimorphism. As boys and girls reach puberty these hormone levels reach a high, initiating the last major growth spurt (DiGangi & Moore, 2012, p. 93). This causes certain features to become visible in the skeleton. An example of such a feature is the shape of the pelvic inlet accounting for a wider birth canal. If this shape were to be the same in both males and females it would have ramifications during childbirth (DiGangi & Moore, 2012, p. 94; Ubelaker & Degaglia, 2017, p. 407).

Extrinsic factors on the other hand come from outside of the body. These factors range from the environment and nutrition to activity and locomotion. Studies have shown that good nutrition accelerates the growth of humans while malnutrition tends to slow down the bodies maturation (DiGangi & Moore, 2012, p. 94). Biomechanical forces like locomotion and activity can also stunt growth due to the plasticity of the bones. The human skeletal system is designed to comply with an individual's weight, size, behaviours, and activities. If this plasticity was absent the bones would 'crack under the pressure' (DiGangi & Moore, 2012, p. 94).

These factors are not the same globally and differ between populations demonstrating that sexual dimorphism itself can be very population specific. This population specificity is particularly seen when applied to studies that are in correlation with growth and size (Ubelaker & Degaglia, 2017, p. 410). Therefore, it is the believe that metric methods can't be blindly used on every population but should be tested and personalised for every population under investigation.

2.2 Morphological sex estimation

Morphological or non-metric sex estimation is concerned with the visual assessment of the features that display sexual dimorphism. An advantage of this type of sex estimation is that it is a relatively quick way of estimating an individuals' sex (Bidmos & Dayal, 2003, p. 322; DiGangi & Moore, 2012, p. 95). It is also a reliable type of method due to accuracies that range from around 85% to 96% (DiGangi & Moore, 2012, p. 96; Phenice, 1969, p. 300; Walker, 2008, p. 46).

2.2.1 *Pelvis*

The pelvis is considered to be the most sexually dimorphic skeletal element of the human body. In both males and females it is a weight-bearing joint necessary for locomotion and for providing protection for the organs. The sexual dimorphism in the pelvis can be easily explained by the added function of childbirth for females. If a female had the same narrow pelvis as a male, the lives of both baby and female would be at risk with the highly likely result of death. This difference in function brings about numerous morphological traits that can be observed and consequently scored (Dirkmaat, 2012, p. 241). Phenice (1969) investigated three traits that occur in the pubic portion of the os coxae. These traits are the presence of the ventral arc, the subpubic concavity, and the

medial aspect of the ischiopubic ramus (figure 2.1). The presence of the ventral arc is indicative of a female. The subpubic concavity can either show a lateral recurve or it can be straight, the recurve is an indication for a female. Lastly, the medial aspect of the ischiopubic ramus has a ridge in females and a broad surface in males (Dirkmaat, 2012, p. 241; Phenice, 1969, pp. 298–300; White & Folkens, 2005, pp. 396–397). Phenice was able to obtain an accuracy of 96% when applied to the American Terry collection. No scale was developed for this method, nevertheless, Phenice did point out that not every os coxa will be perfectly male or perfectly female. One or two criteria might be ambiguous, but there is almost always one feature that will be characteristic of the sex (Phenice, 1969, p. 300). The author also notes that the ischiopubic ramus is the least reliable for sex estimation of the three. Therefore, heavy reliance should only be placed on this trait in the case the two other traits are absent (Dirkmaat, 2012, p. 242; Phenice, 1969, p. 300).

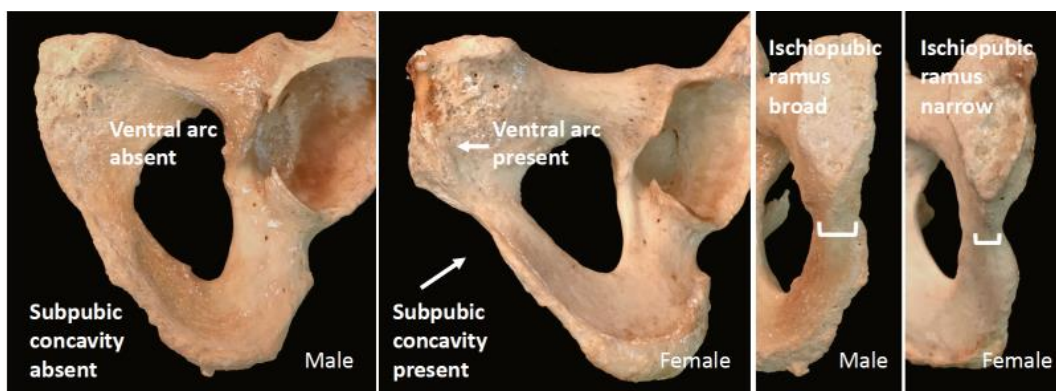


Figure 2.1: The Phenice traits in the pubis region per sex (Shook et al., 2019, p. 556).

The Phenice traits are not the only morphological differences in the pelvis. The Workshop of European Anthropologists (WEA) wrote up recommendations for both sex and age estimations in order to create some homogeneity within the community by combining studies on the subject. A total of ten traits were recommended for the pelvis (table 2.1). These traits are scored following the scale created by Acsádi and Nemeskéri (1970); hyperfeminine (-2), feminine (-1), neutral (0), masculine (+1), and hypermasculine (+2) (Walker, 2008, p. 40; WEA, 1980, p. 521). Examples of traits are the subpubic angle which is more narrow in males, the shape of the obturator foramen which is more triangular in females as opposed to oval in males, and the iliac fossa which is low and broad in females and high and narrow in males. The indication of weight as seen in table 2.1 is a reflection of the importance of the feature (DiGangi & Moore, 2012, p. 95; WEA, 1980, p. 518; White & Folkens, 2005, p. 394).

Table 2.1: Traits and corresponding weights of the pelvis as recommended by the WEA (after WEA, 1980, p. 518).

Pelvic traits	Weight
Preauricular sulcus	3
Greater sciatic notch	3
Sub-pubic angle	2
Arc composé	2
Innonimate bone	2
Obturator foramen shape	2
Ischial body	2
Iliac crest	1
Iliac fossa	1
Pelvic inlet	1

In 1994 Buikstra and Ubelaker published *Standards for data collection from human skeletal remains*. In addition to the Phenice traits the authors added two other features for sex estimation that are also presented in the recommendations of the WEA; the greater sciatic notch and the preauricular sulcus. However, instead of the -2 to +2 scale, the authors opted for a more gradual scale of 1 to 4/5, 1 representing the more soft and gracile forms of females and 5 representing the robusticity of males. One of the added features, the greater sciatic notch, is broader in females and more narrow in males (figure 2.2). In an American population this feature reached an accuracy of 80% (Walker, 2005, p. 385). The other added feature, the preauricular sulcus, tends to not be present in males. As seen in figure 2.3, the 5 on the scale is not visible for this particular trait, but the five can still be scored when the preauricular sulcus is completely absent (Buikstra & Ubelaker, 1994, pp. 18–19). This trait reached an accuracy of 75.8% when tested in an American population (Karsten, 2018, p. 606).

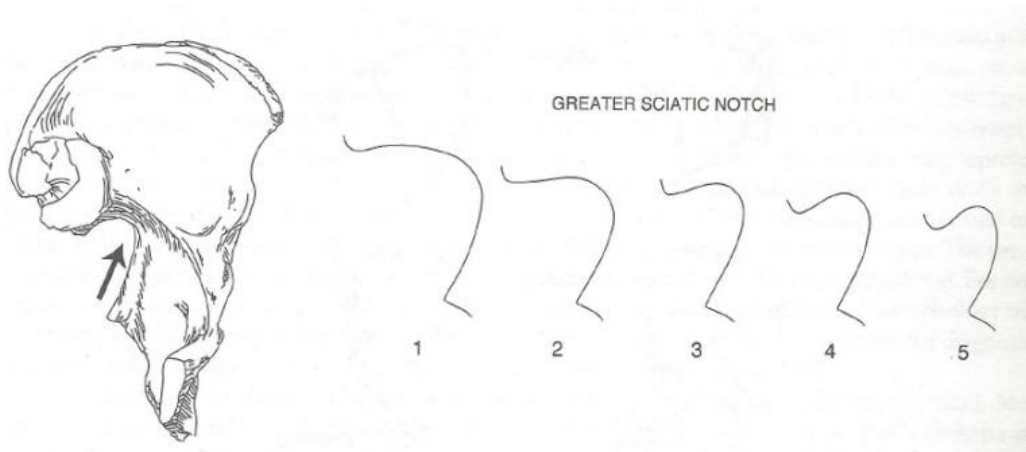


Figure 2.2: Scores of the greater sciatic notch; (1) female, (2) probable female, (3) indeterminate, (4) probable male, (5) male (Buikstra & Ubelaker, 1994, p. 18).

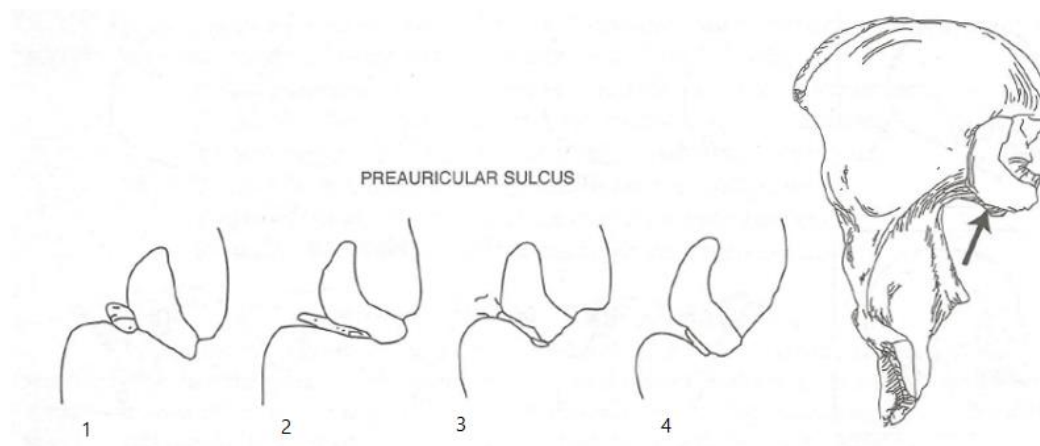


Figure 2.3: Scores of the preauricular sulcus; (1) female, (2) probable female, (3) indeterminate, (4) probable male, (5) is absent but can be scored when preauricular surface is not visible scoring it as male (Buikstra & Ubelaker, 1994, p. 18).

2.2.2 Skull

The skull is the second skeletal element that displays sexually dimorphic traits observable by eye and is traditionally used for sex estimation in conjunction with the pelvis or when the pelvis is not available (Spradley & Jantz, 2011, p. 289). The skull, however, comes with some problems when used for sex estimation. Skulls typically are more gracile in females and more robust in males, but, there is a certain degree of masculinization in the skull as females age and in younger males secondary sex characteristics are often times not yet developed. It is therefore recommended that the use of the skull should be limited to individuals between the ages of 20 and 55 years (DiGangi & Moore, 2012, p. 97; White & Folkens, 2005, p. 387). In addition to this, there is a certain degree of population specificity that can be displayed in some morphological features of the skull. For example, one population may have males with more gracile

brow ridges while another population has males with more robust brow ridges (DiGangi & Moore, 2012, p. 97). The WEA has recommended ten traits for the cranium and four for the mandible, again, the associated weight shows the importance of the feature (table 2.2).

Table 2.2: Traits and corresponding weights of the cranium and mandible as recommended by the WEA (after WEA, 1980, p. 523).

Cranial traits	Weight	Mandibular traits	Weight
Glabella	3	Total aspect	3
Mastoid process	3	Mental eminence	2
Nuchal plane	3	Gonial angle	1
Zygomatic process	3	Inferior margin	1
Superciliary arch	2		
Parietal/frontal bossing	2		
External occipital protuberance	2		
Zygomatic bone	2		
Frontal inclination	1		
Orbit shape	1		

Similar to the pelvis these traits were adapted and scored based on Acsádi and Nemeskéri (1970). The traits have been based on individuals from European descent making it more appropriate for European populations than, for instance, for African populations (WEA, 1980, p. 523). The difference can be seen in a lack of or less frontal and parietal bossing, a larger occipital protuberance, more square orbit shapes, and more gonial eversion in males in comparison to females (White & Folkens, 2005, p. 386).

In an effort to account for the human variation within the features of the skull, Buikstra and Ubelaker chose a total of five cranial traits (the nuchal crest, the mastoid process, the supraorbital margin, the glabella, and the mental eminence on the mandible) and used their previously described 1 to 5 scale (figure 2.4). According to Walker (2008), this type of scale 'generalizes the system and removes the implicit assumption that the morphological condition assigned a zero value represents the optimal cut-point for separating males from females compared to the scale of Ascádi and Nemeskéri (1975) (DiGangi & Moore, 2012, p. 97; Dirkmaat, 2012, p. 243; Walker, 2008, p. 40). Buikstra & Ubelakers (1997) traits and scoring system were tested on African American, European American, and English populations. Multivariate analysis showed an accuracy of 89% when using all five traits (Walker, 2008, p. 46).

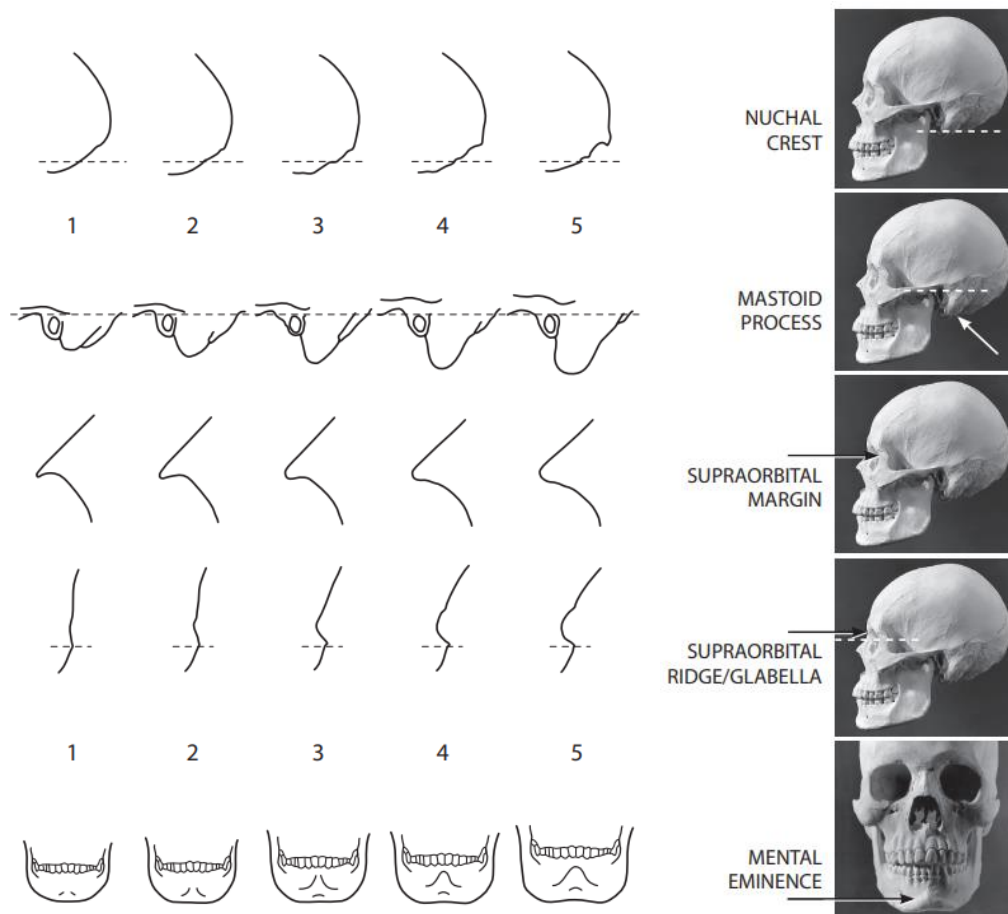


Figure 2.4: Cranial and mandibular traits as defined by Buikstra and Ubelaker with scale (White & Folkens, 2005, pp. 390–391).

2.3 Metric methods

Metric methods are based on the notion that males are relatively larger than females. This makes it possible to give an indication of sex through measurements on a variety of bones. These methods can be applied when morphological traits are either not available, produce indeterminate results or even as supporting evidence. Originally, metric methods were created for the pelvis and skull as they were already in use for morphological analysis, however, in recent decades more and more bones have been and are added to the list (DiGangi & Moore, 2012, p. 92).

2.3.1 Skull

Metric analysis on the skull involves a combination of both metric and observable descriptive traits, one example is the length of the mastoid process. There are, however, exceptions to the rule (DiGangi & Moore, 2012, p. 100). An early study by Giles and Elliot (1963) done on African and European Americans used nine measurements (e.g. mastoid

length, nasal breadth) in discriminant function analysis. Accuracies between 82% and 89% were obtained (Giles & Elliot, 1963, pp. 58; 67). Similarly, Kranioti and colleagues (2008) investigated a total of 16 measurements (e.g. cranial length, frontal breadth, bizygomatic breadth, and mastoid height) on a Cretan population. Multivariate analysis showed accuracies between 83% and 88%. The highest performing single measurement was the bizygomatic breadth with an accuracy of 82% (Kranioti et al., 2008, p. 110.e2). Spradley and Jantz (2011) did multivariate analysis on the crania of African Americans and European Americans using a total of 24 measurements. In African American individuals the authors saw the highest accuracy (90.64%) when using eight measurements (e.g. bizygomatic breadth, mastoid height, and nasal height). In European Americans a similar accuracy was reached (90.01%) using a total of 11 measurements, including the ones used in the African American sample. The bizygomatic breadth performed best with an accuracy of 78% in African Americans and 75% in European Americans (Spradley & Jantz, 2011, pp. 291–294).

2.3.2 *Pelvis*

As the pelvis is the most sexually dimorphic element of the human body metric methods were also created for the pelvis. Following the example of the skull, the pelvis also combines metrics with some observable traits. The subpubic angle, for example, was measured in 1979 by Stewart (1979) showing an obtuse angle in females and an acute angle in males. A later study used the same angle on an Ugandan population and measured it showing a mean of 93.86 degrees in males and of 116.11 degrees in females. The authors were only able to reach an accuracy of 71%, although this was most likely due to poor statistical analysis (DiGangi & Moore, 2012, p. 102). Singh and Potturi (1978) measured the greater sciatic notch (e.g. width, depth, length of posterior segment, and total/posterior angle). The authors found that the width and depth were useless for sex estimation, but the posterior angle on the other hand was accurately assigned in 75% of left and 88% of right male bones and scored an even higher accuracy of 92% in the left and 100% in the right hip bones of females (Singh et al., 1978, pp. 623–624). A study by Patriquin and colleagues (2005) on a black and white South African sample used a total of nine measurements e.g. iliac breadth, width and depth of greater sciatic notch, acetabulum diameter, and pubic length and height (Patriquin et al., 2004, pp. 120–121). A 86% accuracy was obtained for the ischial length exhibiting the most sexual dimorphism out of all the features. The diameter of the acetabulum followed with an accuracy of 84% (DiGangi & Moore, 2012, p. 103; Patriquin et al., 2004, p. 125).

Spradley and Jantz (2011) reached an accuracy of 85% measuring the height of the os coxa and of 83% measuring the length of the ischium when studying multiple postcranial elements of populations from African American and European American descent (Spradley & Jantz, 2011, pp. 293–294). Interestingly, multivariate analysis on the pelvis showed lesser accuracies than were obtained in many of the other postcranial elements researched in this study (DiGangi & Moore, 2012, p. 103).

2.3.3 *Humerus*

Metric data of the humerus (upper arm) is used in several studies over the world. In a modern Greek population Charisi and colleagues (2011) used three measurements (epicondylar width, maximum length, vertical head diameter) and reached an accuracy rate of 96.3%. Leaving the maximum length out, an accuracy of 95.1% was obtained (Charisi et al., 2011, pp. 14–15). The humerus was tested in a Thai collection by Duangto and Mahakkanukrauh (2019). These authors used the same measurements as Charisi et al. (2011), leaving out the diameter of the head and adding both the maximum and minimum diameter of the midshaft. Using all the variables the analysis yielded an accuracy of 97.8%. Univariate analysis saw the highest accuracy (91.2%) in the epicondylar width (Duangto & Mahakkanukrauh, 2019, p. 41). A fairly recent study by Bidmos and Mazengenya (2021) investigated a South African sample with individuals of African, Mixed Ancestry, and European descent. This study took a different route and focused their measurements around the nutrient foramen compiling a total of five measurements (Bidmos & Mazengenya, 2021, p. 1096). Via stepwise analysis the highest accuracy obtained was 82.4% (Bidmos & Mazengenya, 2021, p. 1098).

2.3.4 *Radius and ulna*

The same studies that applied measurements of the humerus also took the radius and ulna (under arm) under investigation. Charisi and colleagues (2011) used the maximum length, the maximum proximal width, and the maximum distal width of both the radius and ulna in a Greek population. Using all measurements of the radius an accuracy of 94.6% was reached whereas omitting the distal width led to a slightly higher accuracy of 95.1%. The same combinations in the ulna yielded accuracy rates of 92.4% and 93.0% (Charisi et al., 2011, p. 15). In the Thai population (Duangto & Mahakkanukrauh, 2019) three measurements of the radius (maximum length, antero-posterior diameter at midshaft, and medio-lateral diameter at midshaft) and five of the ulna (maximum length, antero-posterior diameter at midshaft, medio-lateral diameter at midshaft,

physiological length, minimum circumference) were used. All variables produced an accuracy of 98.2% in the radius and in the ulna it yielded an accuracy of 96.5%. The highest scoring accuracy in the radius was for the antero-posterior diameter (95.2%) and the physiological length scored best in the ulna with a percentage of 90.4% (Duangto & Mahakkanukrauh, 2019, p. 41). Bidmos and Mazengenya (2021) used the same measurements of the humerus for the radius and ulna in their South African sample. The radius reached an accuracy of 86.5% and the ulna an accuracy of 83.5% (Bidmos & Mazengenya, 2021, p. 1098).

2.3.5 *Femur*

Proceeding with the lower limbs, the femur (upper leg) has also proved to be a useful tool in sex estimation. Twelve measurements of the proximal femur were applied to a Dutch collection by Colman and colleagues (2018). Single variables yielded accuracies between 86% and 90% (Colman et al., 2018, p. 268.e5). Robinson and Bidmos (2011) tested three measurements on the proximal end of the femur (head diameter, distal breadth, and transverse diameter at midshaft) on South African collections. Accuracies were obtained ranging between 76% and 93.5% (Robinson & Bidmos, 2011, pp. 212.e2-212.e3). Carvallo and Retamal (2020) also put their focus on the proximal end of the femur using a total of eight measurements on a Chilean sample. Accuracies between 81.4% and 95.7% were obtained (Carvallo & Retamal, 2020, p.5).

2.3.6 *Patella*

Stepping away from the long bones, multiple studies have used the patella (knee cap) in sex estimation. A 2016 study by Peckmann and colleagues (2016) used six measurements of the patella (maximum height, breadth, thickness, and height, lateral breadth, medial breadth of the articular facet) in a Spanish population (Peckmann et al., 2016, p. 85). Direct discriminant function analysis yielded accuracies between 75.2% and 84.8% (Peckmann et al., 2016, p. 88). These same six measurements were applied to an African American population (Peckmann & Fisher 2018). In this study accuracies between 80.5% and 85% were obtained (Peckmann & Fisher, 2018, p. 4).

2.3.7 *Calcaneus*

The calcaneus (heel bone) is often times used in combination with the talus in sex estimation studies. Gualdi-Russo (2007) measured the maximum length, the medial breadth, and the height of calcanei of Northern-Italians. Stepwise analysis produced

accuracies between 87.9% and 90.7% (Gualdi-Russo, 2007, pp. 152–153). Five measurements were taken from an American population with both individuals of African American and European American descent (Alonso-Llamazares & Pablos, 2019, pp. 4929–4930). The length of the talar posterior articular surface produced the highest accuracy of 86.0%. Multivariate accuracies ranged between 86.0% and 87.7% (Alonso-Llamazares & Pablos, 2019, pp. 4933–4934). Curate and colleagues (2021) used a Portuguese sample and took a total of nine measurements reaching accuracies between 73.6% and 86.4% (Curate et al., 2021, pp. 74.e2–74.e4).

2.2.3 Talus

The talus, better known as the ankle bone, is the second largest bone of the tarsals. It is situated between the tibia and the fibula superiorly, lays atop of the calcaneus, and articulates with the navicular distally (figure 2.5). The talus is one of only two bones that has zero attachment with any of the muscles of the human body (White & Folkens, 2005, p. 292). Together with the calcaneus it forms the subtalar joint, which is of importance as it receives and distributes the entire load from the lower limbs and makes movement possible. Articular cartilage covers a large portion of the surface ensuring a smooth movement between the bones of the lower legs and feet (Mahato, 2011, p. 179). Tali make good candidates for sex estimation because they are comparatively well preserved due to their dense and thick structure. In forensic contexts the foot is also fairly well protected due to the presence of footwear. However, post-depositional processes and taphonomy can still do damage to the bone. Besides the preservation objectives, tali are weightbearing bones and weight is believed to be an influence for sexual dimorphism which means it can be used for sex estimation (Mahakkanukrauh et al., 2014, p. 152.e2).

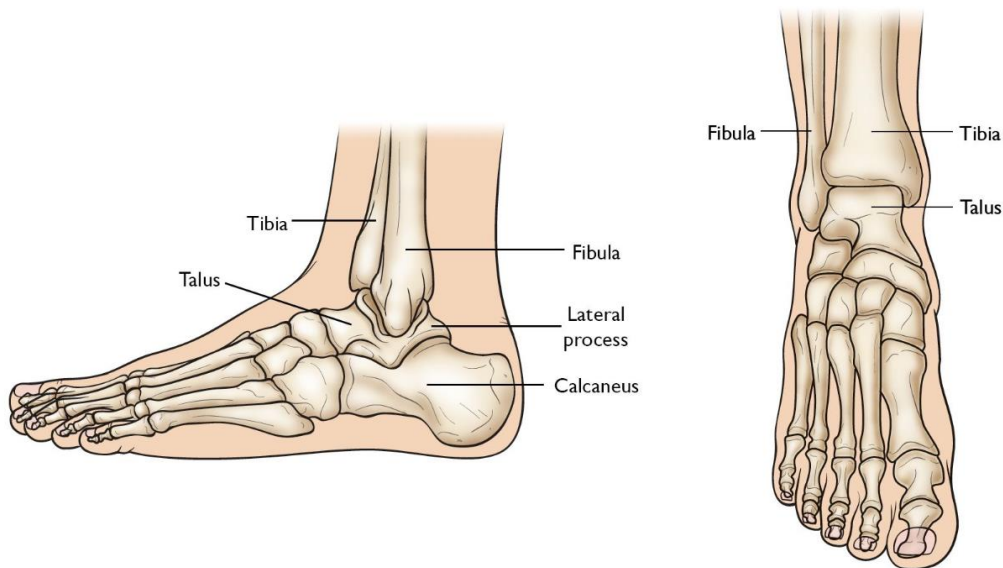


Figure 2.5: The tibia as situated in the foot (www.orthoinfo.aaos.org/en/diseases--conditions/talus-fractures).

2.4.1 Previous studies

The talus has been used for sex estimation in several populations, these studies are discussed below. Tables 2.3 and 2.4 show the abbreviations, names of the measurements, and definitions of the measurements that have been used in these studies.

Table 2.3: Abbreviations and corresponding measurements of all discussed studies (compiled by Piso, N.).

Abbreviation	Measurement
TL (=Le_T=TalLg=M1=TM1=Tal1)	Talar length
TW (=Wi_T=TalBrd=M2=TM2=Tal2)	Talar width
TH (=He_T=TalHt=TM3)	Talar height
TrL (=TRL=M4=Tal4)	Length trochlea
TrB (=TRW=M6=Tal5)	Breadth trochlea
HNL	Head-neck length
HH (=NASH)	Height of head
LPAS (=CASL=MaxIASLg=M12=TM4)	Length of posterior articular surface for calcaneus
BPAS (=CASW=MaxIASBr=M13=TM5)	Breadth of posterior articular surface for calcaneus
NL	Neck length
NW	Neck width
NH	Neck height
NASW	Navicular articular surface width
MinIID	Minimum inferior interarticular distance
MaxLMSHt	Maximum lateral malleolar surface height
MinIDNk	Minimum interarticular distance across the neck.
M1a	Total length of talus
M7	Lateral malleolar oblique height
M9	Length of head talus
TM3a	Talus maximum body height

Table 2.4: Measurements and their definitions of all discussed studies (compiled by Piso, N.).

Measurement	Definition
TL	The projected line from the groove for the flexor hallucis longus muscle at the posterior aspect of the talus to the most anterior point on the articular surface for the navicular.
TW	The maximum projected line laterally/medially perpendicular to the sagittal plane.
TH	The maximum height of the body in an inferior/superior plane.
TrL	The maximum length of the trochlear surface in an sagittal anterior/posterior plane.
TrB	The maximum width of the trochlear surface at the midline, perpendicular to the projected line for the maximum length of the trochlear surface.
HNL	The maximum length of the head and the neck in sagittal plane.
HH	The maximum height of the head of the talus.
LPAS	Maximum length of the calcaneal posterior articular surface, inferior view.
BPAS	Maximum width of the calcaneal posterior articular surface, inferior view.
NL	The length of the neck in superior view.
NW	The maximum projected line of the neck laterally/medially perpendicular to the sagittal plane.
NH	The maximum projected line of the neck superiorly/inferiorly.
NASW	Maximum breadth of the head of the talus.
MinIID	The minimum distance between the anterior edge of the subtalar joint surface and the posterior edge of the middle articular surface.
MaxLMS Ht	The maximum height of only the articular portion of the lateral malleolar surface.
MinIDNk	The minimal distance between the edge of the medial malleolar surface and the edge of the articular surface of the head talus.
M1a	Maximum length from the posterior tubercle to the most anterior point of the head.
M7	Direct distance from the inferior edge of the lateral process to the superior border of the trochlea.
M9	Maximum length of the navicular articular surface.
TM3a	The maximum height of the body in an inferior/superior plane.

Steele (1976) was a pioneer in the study of sexual dimorphism in the foot. He used several measurements, following definitions by Martin (1928), of both the calcaneus and talus from individuals of the American Terry Skeletal Collection (n=120). Steele used five measurements of the talus (TL, TW, TH, TrL, TrB) and two indices (TW/TL, TrB/TrL) in total. Initially, the author divided the sample not only according to sex but also to race, creating four groups; African American males, American white males, African American females and American white females. After descriptive statistics, however, Steele noticed minimal variation in race and combined the sample into two groups, males (n=60) and females (n=60) (Steele, 1976, pp. 582–583). Steele established sexual dimorphism in this population, as all measurements of males were significantly larger than females, however, he found that due to the large overlap of the ranges between males and females that single measurements are of no value in sex estimation. One exception to this rule was the length of the talus due to its accuracy over 80.0% (81.0%) (Steele, 1976, p. 584). Based on stepwise discriminant analysis, Steele developed three discriminant functions. The TL and TW reached an accuracy of 83%; the TL, TW/TL, and

TrB/TrL had an accuracy rate of 86%; and the TL, TW, TH, and TrB/TrL reached an accuracy of 88% (Steele, 1976, p. 585).

Murphy (2002) tested the sexual dimorphism of the talus on a prehistoric Polynesian collection from New Zealand. He used the same five measurements as Steele (TL, TW, TH, TrL, TrB) on a sample size of 51 individuals (24 males and 27 females). Univariate statistics showed that all measurements were significantly larger in males than in females and, in accordance with the study of Steele, the measurement that showed the greatest discrimination between males and females was the maximum length of the talus (88.2%) (Murphy, 2002, pp. 156–157). Stepwise discriminant analysis used two variables (TL, TrB) and yielded an accuracy of 91.3%.

Bidmos and Dayal (2003) added four more measurements to these aforementioned measurements in their study (TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS; figure 2.6). The measurements were taken from the left tali of a white South African sample (n=120, 60 males, 60 females) (Bidmos & Dayal, 2003, p. 323). Univariate discriminant function analysis showed the highest accuracies (81.7% and 80%) for the TL and BPAS and the lowest accuracy (57.5%) for the HH (Bidmos & Dayal, 2003, p. 324). Stepwise discriminate function analysis produced three functions, all nine measurements, all the length measurements and all the breadth measurements. This first function (TL, TrB, HH, LPAS) had an average accuracy of 85.0%. The function of length (TL, LPAS, HNL) yielded an accuracy of 87.5% which was higher than the average function of the breadth (TrB, BPAS) which had an accuracy of 77.5% (Bidmos & Dayal, 2003, p. 325). In total six direct functions were created using between two and nine variables yielding accuracies between 80.8% and 85.8% (Bidmos & Dayal, 2003, p. 326).

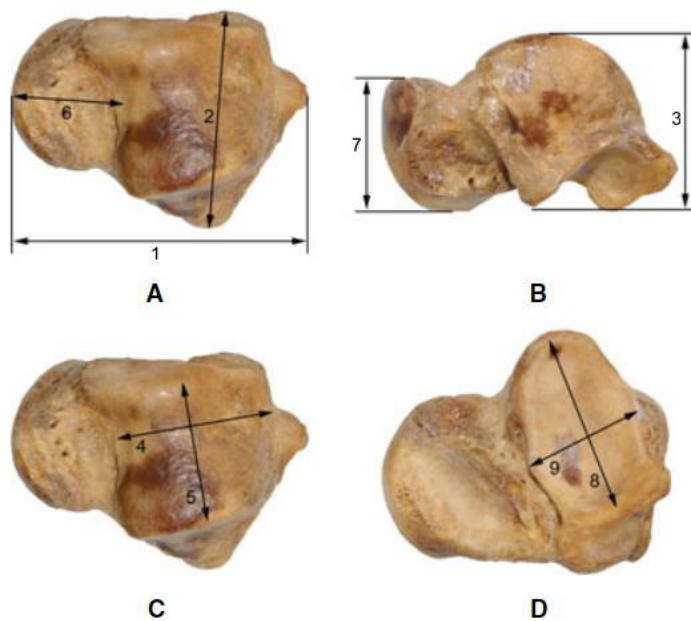


Figure 2.6: The measurements of the talus. Talar length (TL, 1), talar width (TW), talar height (TH, 3), trochlear length (TrL, 4), trochlear breadth (TrB, 5), head-neck length (HNL, 6), head height (HH, 7), length of the posterior articular surface (LPAS, 8), breadth of the posterior articular surface (BPAS, 9). A, C: superior view, B: lateral view, D: inferior view (Lee et al., 2012, p. 167).

A total of 118 adult skeletons (62 males, 56 females) of northern Italian origin were used in a study on talus and calcaneus sex estimation conducted by Gualdi-Russo (2007) (Gualdi-Russo, 2007, p. 151). The author used three measurements of the talus (TL, TW, TH), applied them on both left and right tali and accounted for asymmetry between left and right which led to a total of nine measurements. Asymmetry indices showed no statistically significant differences, with the exception of the talar height. This agrees with the notion of low asymmetry in the lower limbs of humans. Nonetheless, stepwise discriminant functions were created using these distinctions yielding accuracy rates between 90.7% and 95.7% (Gualdi-Russo, 2007, pp. 152–153). This study cross-validated their results with a study done on southern Italians. This produced an accuracy as low as 48.2% on males showing that even within one country there can be much biodiversity (Gualdi-Russo, 2007, p. 155).

In 2012 Lee and colleagues (2012) tested the sexual dimorphism of the talus on a Korean sample (n=140, 70 males, 70 females). The authors used nine measurements (TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS) and 8 indices made from a combination of length, width, and height variables on only the left tali. These indices, however, showed no statistically significant differences between sex and were therefore omitted from further

analysis. Univariate analysis produced accuracy rates from 67.1% to 82.9%, with the LPAS scoring the highest (Lee et al., 2012, pp. 166–167). The authors also computed bivariate equations, giving accuracies of 81.4% to 84.3%. Stepwise analysis chose the TL and BPAS which yielded an accuracy of 82.1%. Direct multivariate discriminant function equations combined the TH, TrL, and LPAS giving an accuracy of 86.4%, the TL, TW, TH produced an accuracy of 72.9%, and a combination of all nine measurements gave the highest accuracy of 87.9%. Population specificity was verified through the studies by Steele (1976), Bidmos and Dayal (2003), Gualdi-Russo (2007), and Murphy (2002). These cross-validated average accuracies ranged between 66.4% and 74.3%, which was lower than the accuracies obtained in the original studies (Lee et al., 2012, pp. 168–169).

That same year the talus was tested on an Egyptian population by Abd-Elaleem and colleagues (2012). Using a sample of 110 right tali (67 male, 43 female) the authors took a total of 12 measurements (TL, TW, NL, NW, TRL, TRW, CASL, CASW, NASH, NASW, TH, NH; figure 2.7) (Abd-Elaleem et al., 2012, p. 72). CASW and NASW showed no statistically significant differences between male and female and therefore got excluded from discriminant function analysis (Abd-Elaleem et al., 2012, p. 73). The measurement that yielded the highest accuracy was the TL (90.9%) and the lowest accuracy was reached with the NL (51.8%). The direct discriminant analysis showed the highest accuracy with the TL, TW, and NW with a percentage of 85.5%. The stepwise analysis on the other hand had an accuracy of 83.6% using the TL and CASL (Abd-Elaleem et al., 2012, p. 75).



Figure 2.7: Measurements NASH (line 11) and NASW (line 12) (Abd-Elaleem et al., 2012, p. 71).

Harris and Case (2012) conducted research on a modern American population of European American ancestry (n=160, 82 males, 78 females). The authors took a total of three measurements (TalLg, TalBrD, TalHt) on both the left and the right talus (Harris & Case, 2012, p. 296). Using logistic regression analysis, the TalLg got an accuracy of 86.7%, the TalBrD 87.4%, and the TalHt yielded an accuracy of 86.8% on the left side. The right side reached accuracies of 87.6%, 87.3%, and 88.0% respectively. The breadth of the talus proved to be the most sexual dimorphic based on a sexual dimorphism index (Harris & Case, 2012, pp. 303–304).

A total of 252 skeletons (126 males, 126 females) of Thai origin were investigated by Mahakkanukrauh and colleagues (2014) by taking a total of ten measurements (MaxLg, MaxBr, MaxHt, MaxTrLg, MaxTrBr, MaxIASLg, MaxIASBr, MinIID, MaxLMSHt, MinIDNk) on both left and right tali (Mahakkanukrauh et al., 2014, p. 152.e2). Logistic regression analysis showed that variables MaxTrLg on the left and MaxTrBr on the right reached the highest allocation accuracy of over 89%. MinIID, MaxLMSHt, and MinIDNk reached very low accuracy rates and were excluded from further investigation (Mahakkanukrauh et al., 2014, p. 152.e5). A combination of MaxTrLg and MaxTrBr on both the left and right side performed the best with accuracies of 91.3% and 91.4% respectively. All other created equations reached accuracy rates of at least 88.0% (Mahakkanukrauh et al., 2014, p. 152.e8).

Peckmann and colleagues (2015) tested nine measurements (TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS) on a Greek collection (n=182, 96 males, 86 females) (Peckmann et al., 2015, p. 15). The LPAS and TL showed the highest accuracies of 87.3% and 85.0% respectively with univariate analysis. Stepwise analysis selected the TW and TL and computed an accuracy of 84.7%. The direct functions were created by combining all nine variables yielding an accuracy of 86.9%, all variables of length (TL, HNL, TrL, LPAS) correctly assigning 85.4%, all breadth variables (TW, TrB, BPAS) computing an accuracy rate of 82.5%, and lastly all variables of height (HH, TH) with an accuracy of 82.4% (Peckmann et al., 2015, p. 16). This study also cross-validated their data with Bidmos and Dayal (2003) and Lee et al. (2011) producing accuracies of 83.2% and 78.7% which is lower than the original accuracies (Peckmann et al., 2015, p. 18).

Nine measurements (M1, M1a, M2, M4, M6, M7, M9, M12, M13) were taken by Alonso-Llamazares and Pablos (2019) on 114 individuals (57 males, 57 females) of an American collection comprising of individuals of both European American and African American

origin (Alonso-Llamazares & Pablos, 2019, p. 4929). Individually, the talar length (M1) reached the highest accuracy of 90.2% with the total breadth (M2) a close second with 88.4% (Alonso-Llamazares & Pablos, 2019, p. 4933). Stepwise analysis combined the talar length, the trochlear height, the trochlear length, and the breadth of the calcaneal posterior articular surface (M1, M6, M12, M13) and reached an accuracy of 93.8%. Multiple direct discriminant functions equations were created giving an accuracy range between 86.6% and 92.9% (Alonso-Llamazares & Pablos, 2019, p. 4934).

Curate and colleagues (2021) used 180 individuals (87 males, 93 females) from a Portuguese collection testing six measurements (TM1, TM2, TM3, TM3a, TM4, TM5) (Curate et al., 2021, p. 74.e2). Through logistic regression, the authors found that the variable that performed best individually was the maximum talar body height (TM3a) with an accuracy of 83.9% (Curate et al., 2021, p. 74.e5). A combination of the talar length, and length and breadth of the calcaneal posterior articular surface (TM1, TM4, TM5) gave the highest accuracy of 89.4 % (Curate et al., 2021, pp. 74.e6-74.e7).

The study by Indra and colleagues (2021) focused mainly on validating the population specificity of sex estimation equations created by Peckmann et al. (2015) by using a Swiss sample (n=234, 117 males, 117 females). The authors used tali primarily from the left side and took four measurements based on the highest accuracies that the study of Peckmann (2015) had obtained (TL, TW, TrL, TrB) (Indra et al., 2021, p. 556). The talar length performed best with an accuracy of 85.0%. Via stepwise analysis an accuracy of 88.0% was reached when using the variables TL, TW, TrB. Direct analysis used all four variables and yielded an accuracy of 87.6% (Indra et al., 2021, p. 561). Inserting their own data into the equations by Peckmann (2015) accuracies between 67.1% and 86.3% were reached in comparison to the original 79.1% to 85.0% when looking at both univariable and multivariable functions (Indra et al., 2021, p. 561).

Based on these studies it can be concluded that the talus is indeed sexually dimorphic. An accuracy of at least 80% is reached in the multivariate analysis of all studies, except for one function of Lee et al. (2012) and one of Bidmos & Dayal (2003). However, there is no homogeneity in the used measurements. Univariate analysis also lacks a form of homogeneity, although the talar length provided the highest accuracy in a total of six studies (table 2.5).

Table 2.5: Studies and their best outcomes (compiled by Piso, N.).

Study	N	Sample Type	Best Multiple Function (%)	Best Single Bone Function (%)	Most Dimorphic Measure
Steele (1976)	120	Mixed American	88.0	81.0	Talar length
Murphy (2002)	51	New Zealand	93.3	88.2	Talar length
Bidmos&Dayal (2003)	120	South African (white)	87.5	81.7	Talar length
Gualdi-Russo (2007)	118	Northern Italian	95.7	-	-
Lee et al. (2012)	140	Korean	87.9	82.9	Length of posterior articular surface for calcaneus
Abd-Elaleem (2012)	110	Egyptian	85.5	90.9	Talar length
Harris & Case (2012)	160	European American	-	88.0	Talar height
Mahakkanukrauh (2014)	252	Thai	91.4	89.0	Trochlear length/Trochlear breadth
Peckmann et al. (2015)	182	Greek	86.9	87.3	Length of posterior articular surface for calcaneus
Alonso-Llamazares & Pablos (2019)	114	Mixed American	93.3	90.2	Talar length
Curate et al. (2021)	180	Portuguese	89.4	83.9	Talar body height
Indra et al. (2021)	234	Swiss	88.0	85.0	Talar length

This chapter has provided an explanation as to why the sex of humans can be estimated on the basis of skeletal elements which involves sexual dimorphism. The differences that occur due to this phenomenon can be observed through various methods. This chapter has provided explanations and accuracies of morphological methods and of various bones using metric methods. This showed that morphological methods are only applicable to the pelvis and skull with accuracies that range from 85% to 96%. Metric methods on the other hand have been applied to a variety of bones. The examples provided in this study show accuracies ranging from 75% to as high as 98.2%. Lastly, the talus has been emphasized in the form of various previously performed studies. Here we see accuracies as high as 95.7% and as low as 57.7%. Some of the methods and measurements of these studies have been taken into consideration for this particular study. The exact measurements, types of analysis, and the origins of the material will be set out in chapter three.

3. Materials and methods

This chapter provides information on the skeletal collection that is used in this study to determine if the talus metric method is effective in estimating sex in a post-medieval Dutch population. It will first discuss the historical context of the site Middenbeemster and then it will provide information on the archaeological context of the site. Besides the materials that are used in this research this chapter will also discuss the methods that this study implements. This includes how the data is selected and collected, the measurements that are used and how they can be measured, and lastly, how the data for this research is analysed.

3.1 Materials

3.1.1 *Historical context Middenbeemster*

The Beemster municipality is a polder located in the province of Noord-Holland. Settlers drained the marshy lake and elevated the land between 1609 and 1613 making the polder habitable (Palmer, 2019, p. 17). The polder was divided into square plots making it unique for the time period of the 17th century (figure 3.1). This design is one of the reasons UNESCO declared the Beemsterpolder a world heritage site in 1999 (van Spelde & Hoogland, 2018, p. 307). At the end of the drainage in AD 1613 the village of Middenbeemster was founded in the centre of the polder (Palmer, 2019, p. 17).

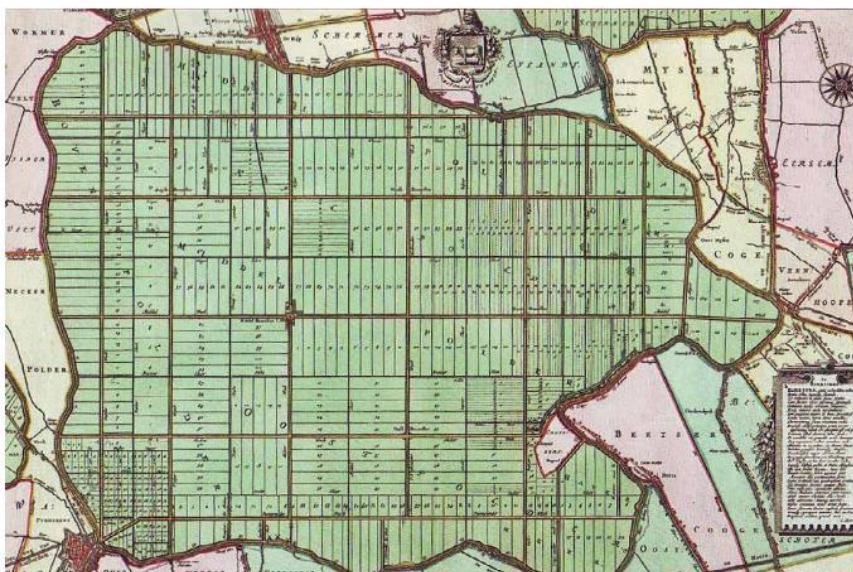


Figure 3.1: Map showing the division of the Beemster into square plots, 1658 (van Spelde & Hoogland, 2018, p. 309).

Initially, the settlers decided on building a total of five churches within the municipality. For this purpose five slots were assigned and elevated using the ground recovered from the surrounding ditch they created. Ultimately only one of those churches was realised, the Keyserkerk in Middenbeemster. The building of the church started in 1618 and it was consecrated a few years later in 1623 (Hakvoort, 2013, p. 13; Palmer, 2019, p. 17; van Spelde & Hoogland, 2018, p. 308). Immediately after its consecration the cemetery located on the southside of the church came into use. As the other churches were never realised this cemetery became the only place for burials in the whole Beemster municipality. It remained in use until 1867 when new laws called for cemeteries outside of town limits (Hakvoort, 2013, p. 29; van Spelde & Hoogland, 2018, pp. 308–309).

3.1.2 Archaeological context Middenbeemster

In 2011 archaeological excavations commenced after plans were made for construction adjacent to the church (figure 3.2). The excavation was executed in eight weeks by a collaboration of the University of Leiden and the commercial company Hollandia Archeologen. A total of 488 primary graves and a small amount of secondary ossuary's were excavated (van Spelde & Hoogland, 2018, p. 309). The cemetery was organized into rectangular plots slightly larger than the average adult coffin. Every individual was buried in a wooden coffin (figure 3.3). These coffins were stacked on top of each other generally three or four deep in a plot (van Spelde & Hoogland, 2018, p. 310). A large portion of the excavated individuals date to a later period between 1829 and 1866. There are also a few that date back to 1615 meaning that individuals were already buried on the church ground before the church was even build (Hakvoort, 2013, p. 35).

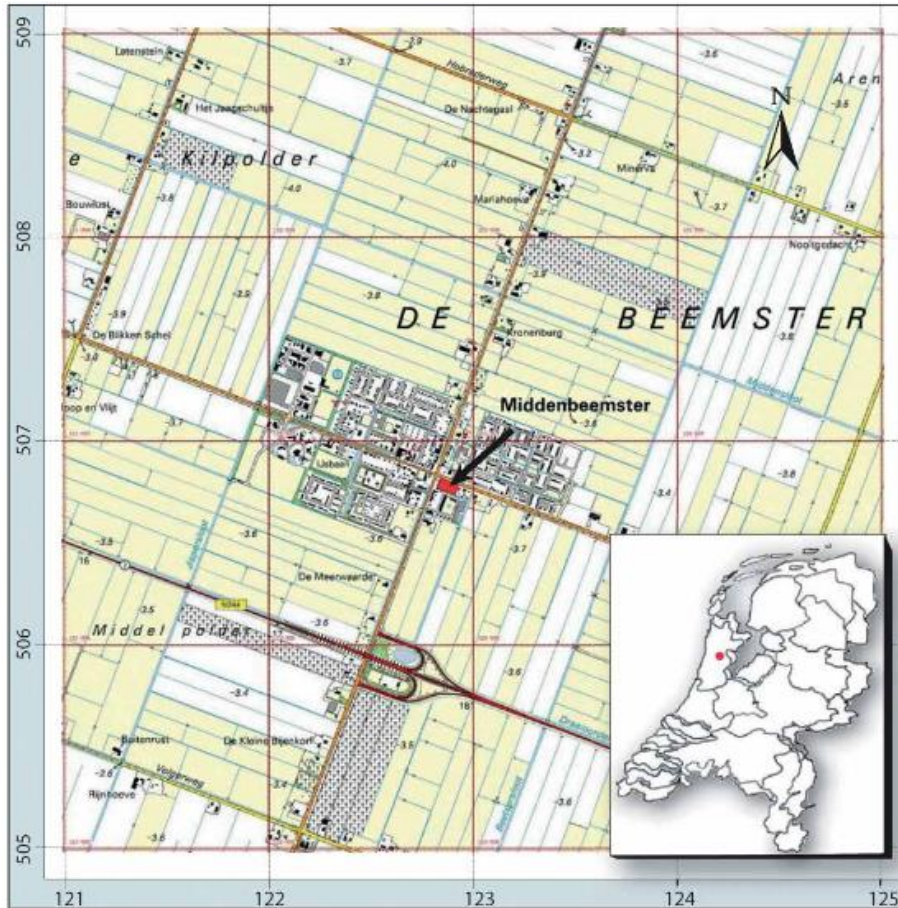


Figure 3.2: Location of the excavation in Middenbeemster (Hakvoort, 2013, p. 11).



Figure 3.3: Wooden coffin with the lid pressed inside of the coffin due to the weight of the soil (van Spelde & Hoogland, 2018, p. 319).

This post-medieval collection is very valuable for osteoarchaeological research not only due to its rich historical background but also because of existing archival records of the cemetery. The archaeologists, in collaboration with the Middenbeemster Historical Society, were able to link burial plots to entries in the burial register. They were able to

identify a total of 120 excavated individuals of whom the age-at-death, sex, and name is now known (Palmer, 2019, p. 18). For this study in particular knowledge of the sex of the individual is vital.

3.2 Methods

3.2.1 Data selection

For this particular research tali of adult individuals of known sex are required. Therefore, the 120 individuals identified through the archival records of the Middenbeemster collection were investigated. The studies by Gualdi-Russo (2007), Peckmann et al. (2015), Curate et al. (2021) showed no to negligible differences between the left and right sides (Curate et al., 2021, p. 74.e2; Gualdi-Russo, 2007, p. 152; Peckmann et al., 2015, p. 15). In accordance with those discoveries and the studies done by Steele (1976), Murphy (2002), Bidmos & Dayal (2003), Abd-Elaleem et al. (2012), Lee et al. (2012), Navega et al. (2014), Alonso-Llamazares & Pablos (2019), Indra et al. (2021) that only investigated either the left or the right-side talus (Abd-Elaleem et al., 2012, p. 72; Alonso-Llamazares & Pablos, 2019, p. 4929; Bidmos & Dayal, 2003, p. 323; Indra et al., 2021, p. 556; Lee et al., 2012, p. 166; Murphy, 2002, p. 156; Navega et al., 2014, p. 653; Steele, 1976, p. 582) this research opted to not account for asymmetry as well. Therefore, only the left or the right talus is needed. Inventory of the presence or absence of a talus of an individual yielded a total of 109 left tali and 107 right tali. In combination with this higher number of individuals the left side appeared to have more intact tali than the right side. Therefore, the decision was made to investigate primarily the left tali. In the event that the left side was not present or in a bad condition, the right talus was substituted. In total 111 tali of the 120 individuals (47 males, 64 females) were present and suitable for this research.

3.2.2 Data collection

During data collection there was no prior knowledge of the sex of the individuals. A total of nine measurements were taken with a digital sliding calliper and added into *Microsoft Excel*. Around 50% of the talus had to be present in order to be included in this research. If the condition of the talus prevented certain measurements to be taken these measurements were then omitted from the data. In order to check for repeatability every fifth individual was measured twice with at least a week in between measuring. In the cases that not every measurement could be taken from the fifth individual either the

next or the previous complete talus in the sequence was measured (Harris & Case, 2012, p. 296). The measuring process involved only one single observer, therefore, no inter-observer reliability could be tested (Lee et al., 2012, p. 167).

3.2.3 *Measurements*

This research used a total of nine measurements, the talar length (TL), talar width (TW), talar height (TH), trochlear length (TrL), trochlear breadth (TrB), head-neck length (HNL), head height (HH), length of the posterior articular surface (LPAS), and the breadth of the posterior articular surface (BPAS) (figure 3.4). The first five of these measurements are introduced in the original study by Steele (Steele, 1976, p. 582-583). Bidmos and Dayal (2003) added the latter four measurements to their study. Many other studies have opted for the same nine measurements or a variation of them. The decision to use the nine measurements (TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS) is primarily based on the high occurrence and accuracies of them in the other studies. A few studies, Gualdi-Russo (2007), Harris & Case (2012), and Navega et al. (2015) opted to only use three measurements in total, the TL, TW, and TH (Gualdi-Russo, 2007, p. 152; Harris & Case, 2012, p. 296; Navega et al., 2014, p. 653). All the other studies have also included these three into their research. As they have the highest occurrence of all measurements, these three measurements have been included in this study. The TrL and TrB were researched in all of the studies that included more than the three measurements (Curate et al., 2021, p. 74; Murphy, 2002, p. 156; Steele, 1976, p. 582) and more often than not these studies also included the LPAS and BPAS. These measurements produce fairly high accuracies (Abd-Elaleem et al., 2012, p. 72; Alonso-Llamazares & Pablos, 2019, p. 4930; Bidmos & Dayal, 2003, p. 324; Lee et al., 2012, p. 167; Mahakkanukrauh et al., 2014, p. 152.e2; Peckmann et al., 2015, p. 15), therefore, this study chose to include these measurements as well. Lastly, the HNL and HH measurements have a lesser occurrence in previous studies and produce lower accuracy rates (Abd-Elaleem et al., 2012, p. 72; Bidmos & Dayal, 2003, p. 324; Lee et al., 2012, p. 167; Peckmann et al., 2015, p. 15). Nevertheless, these measurements are included in this study so that a comparison can be made with other studies in order to test for population specificity.

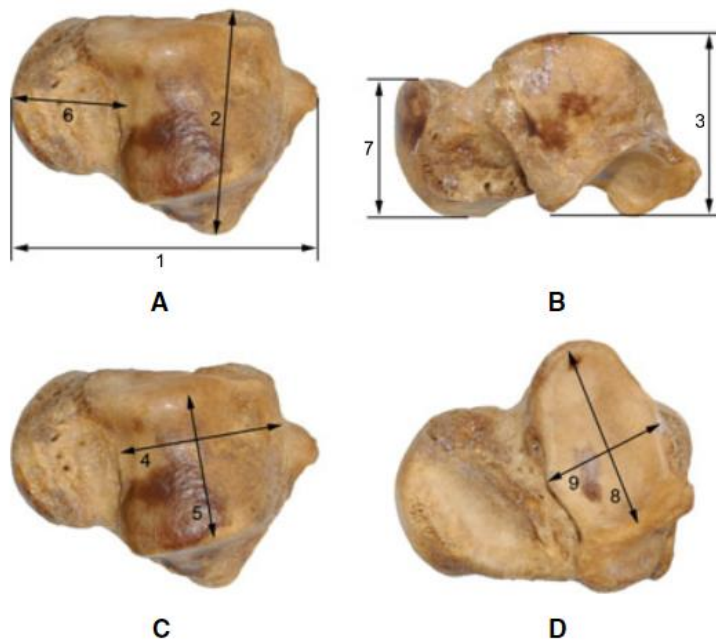


Figure 3.4: The measurements of the talus. Talar length (TL, 1), talar width (TW), talar height (TH, 3), trochlear length (TrL, 4), trochlear breadth (TrB, 5), head-neck length (HNL, 6), head height (HH, 7), length of the posterior articular surface (LPAS, 8), breadth of the posterior articular surface (BPAS, 9). A, C: superior view, B: lateral view, D: inferior view (Lee et al., 2012, p. 167).

The measurements (figure 3.4) follow definitions given by Martin and Knussman (1988) (Bidmos & Dayal, 2003, p. 324). The talar length (TL) is the projected line from the groove for the flexor hallucis longus muscle at the posterior aspect of the talus to the most anterior point on the articular surface for the navicular. It is measured by placing the fixed arm calliper on the most posterior point of the groove for the flexor hallucis longus tendon. The movable arm is placed on the most anterior point of the articular surface for the navicular i.e. the head of the talus. The talar width (TW) is the maximum projected line from the lateral side to the medial side perpendicular to the sagittal plane. The fixed arm of the calliper must be placed on the most lateral point of the articular facet of the lateral malleolus. The movable arm must be placed on the medial surface of the talus. The height of the talus (TH) is the maximum height of the body from the inferior to the superior plane. To measure this the talus needs to be placed on a flat surface. The fixed arm is placed on the most superior point on the trochlear articular surface while the moveable arm goes to the inferior surface of the talus (Martin & Knussmann, 1988, p. 223). The length of the trochlea (TrL) is the maximum length of the trochlear surface in an anterior to posterior plane. The trochlea is a superiorly positioned surface that articulates with both the tibia and the fibula. Its length is

measured by placing the fixed arm on the most posterior point and the movable arm on the most anterior point of the trochlear articular surface in a sagittal plane. Its breadth (TrB) is the maximum breadth of the surface perpendicular to the maximum length. Therefore, it is measured from the lateral to the medial point perpendicular to the TrL measurement. The head-neck length (HNL) is the maximum length of the head and the neck within the sagittal plane. In order to measure it the fixed arm is placed on the most anterior point of the trochlear surface and the moveable arm goes to the most anterior point of the edge of the articular surface for the navicular. The height of the head (HH) is the maximum height of the head of the talus, it is measured by placing the fixed arm on the most superior point of the head of the talus while the moveable arm is placed on the most inferior point of the head of the talus. Lastly, the measurements of the length (LPAS) and breadth (BPAS) of the posterior articular surface articulate with the calcaneus. The former is measured by placing the fixed arm on the most lateral point of the calcaneal posterior articular surface and the moveable arm goes to the most medial point of the same surface. The breadth is measured in the same manner but in a posterior to anterior plane on this surface (Martin & Knussmann, 1988, p. 224).

3.2.4 Statistical analysis

Data analysis was performed using the statistical program *IBM Statistics SPSS version 28*. All statistical significance was set at $p \leq 0.05$ (Abd-Elaleem et al., 2012, p. 72; Bidmos & Dayal, 2003, p. 323; Gualdi-Russo, 2007, p. 152). Intra-observer error was done using the intraclass correlation coefficient and by observing the average difference between the measurements. Data processing then started with observing any outliers in the dataset on the basis of z-scores and boxplots. It was then tested for normal distribution using the one-sample Kolmogorov-Smirnov test and the Shapiro Wilk test (Alonso-Llamazares & Pablos, 2019, p. 4932). Following this, the descriptive statistics of each measurement including the mean, the standard deviation, the minimum, and the maximum were calculated for each sex. This research is based on the assumption that there is a significant difference between males and females in certain features. In order to test this, independent Student's *t*-tests were performed for each variable (Lee et al., 2012, p. 167). Next, discriminant functions were generated if and when the Student's *t*-test of the variable showed a significant difference. These functions were generated with both single and multiple variables. As a result of the univariate analysis, demarcation points can be calculated. If a measured value is higher than the demarcation point the individual is classified as male and if it is lower the individual can be classified as female.

It is calculated by taking the average of the male and female mean (Peckmann et al., 2015, p. 15). Multivariate functions were created using both the direct and the stepwise method. With the direct method functions are created by combining certain variables that are entered simultaneously and *IBM Statistics SPSS* will calculate an accuracy rate. For the stepwise method on the other hand all the variables are entered and *IBM Statistics SPSS* will decide which variable will provide the best accuracy (Bidmos & Dayal, 2003, p. 324). Simultaneously, a leave-one-out approach was taken for every discriminant analysis in order to cross-validate the function. This ensures that the function retains a similar accuracy rate even when one measurement is missing from the dataset (Bidmos & Dayal, 2003, p. 325). Lastly, the metric data of this Dutch population has been entered into various functions of three previously discussed studies. In all studies the sectioning point was set at 0.000. A discriminant function score greater than the sectioning point indicates a male and lesser or equal to the sectioning point indicates a female (Bidmos & Dayal, 2003, p. 325; Lee et al., 2012, p. 167; Peckmann et al., 2015, p. 17). On the basis of those outcomes accuracy rates are calculated.

This chapter has provided information on the history of the burial site where the observed skeletal remains come from. It has also provided context on the archaeological excavations that took place to obtain these remains. Middenbeemster is a useful collection for research due to its archival records. This chapter has also given an insight into how the data has been collected, measured, and processed, substantiated with other studies on the talus metric method. The following chapter will provide the output of the performed statistical analysis.

4. Results

This chapter shows the results of the statistical analysis performed on the 111 individuals from the Middenbeemster collection. The data was inserted into *Microsoft Excel*, transformed into a csv. file and imported into the program *IBM Statistics SPSS version 28*. All statistics were performed using this program, the tables, however, were created in *Microsoft Excel*. Each test and its results are showed separately.

Interpretation of these results will be discussed in chapter five.

4.1 Intra-observer error

To test for the intra-observer error in this research every fifth individual was measured twice by the same observer. This led to measuring a total of 22 tali. To test for the observer error all the measurements were investigated separately using the intraclass correlation coefficient (Lee et al., 2012, p. 167). An intraclass correlation coefficient is a value between 0 and 1. A value over 0.9 presents excellent reliability, between 0.75 and 0.9 indicates a good reliability, between 0.5 and 0.75 it is of moderate reliability, and every value under 0.5 indicates a poor reliability (Bobak et al., 2018, p. 2).

Table 4.1 shows the intraclass correlation coefficient (ICC) and the average difference between repeated measurements. The LPAS shows the highest ICC (0.991), while the lowest ICC (0.946) is seen in the TH. This is in concordance with the lowest and highest average differences between repeated measures. This means that all ICC's are acceptable and of excellent reliability.

Table 4.1: Intraclass correlation coefficient and the average difference between repeated measurements.

Measurement	ICC	Average difference between repeat measurements (mm)
TL	0.987	0.39
TW	0.993	0.22
TH	0.946	0.52
TrL	0.955	0.37
TrB	0.965	0.38
HNL	0.975	0.40
HH	0.951	0.42
LPAS	0.991	0.20
BPAS	0.970	0.29

4.2 Outliers

To test for any outliers in the dataset boxplots were created per variable. If any outliers existed the matching z-score was evaluated. On the basis of the boxplots, z-scores, and comparison to the other measurements of the individual a decision was made whether it was indeed an outlier based on data entry error or whether it was in line with the large or small size of the individual. In case of the former the decision was made to treat the outlier as a missing value. Three cases were treated as such. The z-score should fall into a range of -3.0 and 3.0 (Fletcher & Lock, 2005, p. 50). As can be observed in table 4.2 individual 95 falls within this range, however, the boxplot (figure 4.1) depicts the HNL measurement of the individual as an outlier and is therefore treated as such. Individuals 30 and 12 have z-scores out of the norm and are therefore treated as outliers.

Table 4.2: Individuals with their z-score.

Individual	Sex	Measurement	Z-score
95	M	HNL	2,8369
30	F	HH	-3,83776
12	F	BPAS	6,38471

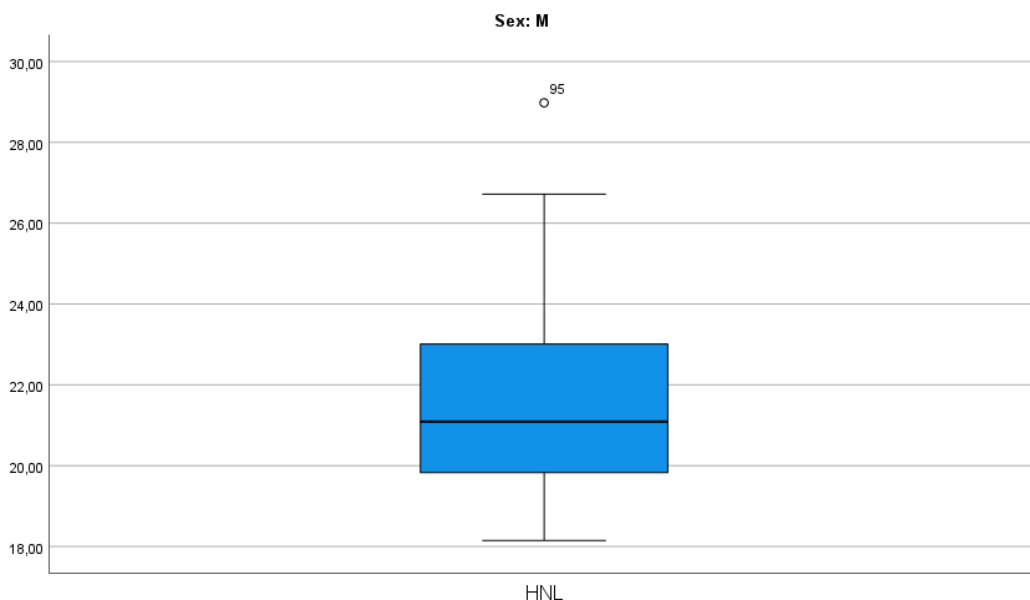


Figure 4.1: Boxplot of outlier individual 95.

4.3 Normal distribution

In order to see if the data is normally distributed a Kolmogorov-Smirnov and a Shapiro Wilk test was performed. Both significance values are set at $p \geq 0.05$. As can be seen in table 4.3 all measurements for both tests have a significance level over 0.05. Therefore, it can be accepted that this data is normally distributed.

Table 4.3: KS p -value = Kolmogorov-Smirnov p -value, SW p -value = Shapiro Wilk p -value, * = lower bound of the true significance.

Variable	Sex	Skewness	Kurtosis	KS p -value	SW p -value
TL	F	-0.208	0.122	0.200*	0.896
	M	-0.071	0.058	0.200*	0.510
TW	F	-0.103	-0.145	0.168	0.714
	M	0.197	-0.408	0.200*	0.732
TH	F	0.054	0.333	0.200*	0.992
	M	0.300	0.611	0.200*	0.409
TrL	F	0.062	1.208	0.200*	0.659
	M	0.242	-0.325	0.200*	0.473
TrB	F	-0.409	-0.735	0.071	0.073
	M	0.302	0.192	0.200*	0.921
HNL	F	0.518	0.075	0.200*	0.189
	M	0.799	-0.116	0.122	0.011
HH	F	-0.200	-0.079	0.200*	0.462
	M	-0.082	-0.002	0.200*	0.557
LPAS	F	-0.579	1.788	0.200*	0.143
	M	-0.129	-0.426	0.199	0.396
BPAS	F	-0.504	1.724	0.200*	0.185
	M	0.594	1.078	0.102	0.159

4.4 Descriptive statistics and Student's t -test

In table 4.4 the results of the descriptive statistics and the Student's t -tests per variable are presented. The descriptive statistics show the mean, the standard deviation, and the range (minimum and maximum) of the variables. It can be observed that the mean of the males is higher than the mean of the females in all nine variables. It can also be observed that the maximum of the females lies close to or slightly above the mean of the males. Likewise, the minimum of the males lies around the same measure of the mean of the females.

To test if the difference between the means of the males and females is statistically significant, a Student's t -test was performed. Table 4.4, shows a p -value of <0.001 for all

nine variables. Significance was set for a p -value lower than 0.05. All nine variables are below 0.05 meaning that the difference between males and females is statistically significant for the talar length (TL), the talar width (TW), the talar height (TH), the trochlear length (TrL), the trochlear breadth (TrB), the head-neck length (HNL), the head height (HH), the length of the posterior articular surface (LPAS), and the breadth of the posterior articular surface (BPAS).

Table 4.4: Descriptive statistics, Student's t -test and the significance (p -value).

Variable	Male					Female					T-value	P-value
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max		
TL	47	60.77	3.16	52.55	67.25	63	53.59	3.18	45.09	60.36	-11,73	<0,001
TW	45	44.25	3.12	37.94	51.15	59	39.03	2.28	32.98	44.41	-9,88	<0,001
TH	47	33.21	2.39	27.78	39.54	64	29.53	1.83	24.90	34.77	-9,17	<0,001
TrL	46	35.91	2.16	31.60	40.29	56	31.81	1.96	26.32	37.82	-10,05	<0,001
TrB	47	28.63	2.03	24.05	33.95	61	25.28	1.84	21.12	28.60	-8,96	<0,001
HNL	40	21.46	2.33	18.15	26.72	48	19.65	2.75	14.52	26.58	-3,29	<0,001
HH	47	29.90	2.82	23.66	35.98	58	26.97	2.32	21.25	31.26	-5,85	<0,001
LPAS	46	34.19	2.16	29.45	38.24	59	29.52	1.98	22.71	34.71	-11,52	<0,001
BPAS	47	23.37	1.69	20.01	28.21	62	20.43	1.49	15.07	23.83	-9,62	<0,001

4.5 Univariate analysis

In table 4.5 the accuracies and the demarcation points resulting from the univariate analysis per variable are shown. The demarcation points are created by taken the average of the means of both the males and females (Peckmann et al., 2015, p. 16). If a measured value is below the demarcation point the individual can be classified as female, if it is over the demarcation point the individual can be classified as male. Besides the demarcation points the original accuracies and the leave-one-out accuracies are presented. The highest average original accuracy (88.8%) as well as the highest female accuracy (94.9%) is seen in the LPAS, while the highest original accuracy for the males is seen in the TL (83.0%). The leave-one-out accuracies show the same results, except for the males (82.6%) where the highest accuracy is also found in the LPAS. The lowest accuracies can be found in the HNL for the males (47.5%; 47.5%) and the average (62.3%; 62.3%) while it is lowest for the females in the HH value (75.9%; 74.1%).

Table 4.5: Accuracies and demarcation points resulting from univariate analysis.

Variables	Original accuracy (%)			Leave-one-out accuracy (%)			Demarcation points
	Male	Female	Average	Male	Female	Average	
TL	83.0	88.9	86.0	80.9	88.9	84.9	F < 57.18 < M
TW	71.1	91.5	81.3	71.1	91.5	81.3	F < 41.64 < M
TH	72.3	92.2	82.3	72.3	90.6	81.5	F < 31.37 < M
TrL	82.6	89.3	86.0	80.4	89.3	84.9	F < 34.78 < M
TrB	70.2	85.2	77.7	68.1	85.2	76.7	F < 26.96 < M
HNL	47.5	77.1	62.3	47.5	77.1	62.3	F < 20.56 < M
HH	72.3	75.9	74.1	72.3	74.1	73.2	F < 28.43 < M
LPAS	82.6	94.9	88.8	82.6	94.9	88.8	F < 31.86 < M
BPAS	80.9	88.7	84.8	78.7	88.7	83.7	F < 21.90 < M

4.6 Multivariate analysis

4.6.1 Direct discriminant functions

Table 4.6 shows the unstandardized discriminant function coefficients, the constants, and the original and leave-one-out accuracies of different functions. These functions were created by different combinations of the variables. Function 1 is a combination of all nine variables, function 2 combines all the variables that measure a length which are the TL, the TrL, the HNL, and the LPAS. Function 3 and 4 are similar to function 2 but are combinations of the width (TW, TrB, BPAS) and height (TH, HH) variables respectively. Function 5 combines the talar variables (TL, TW, TH), function 6 is a combination of the trochlear variables (TrL, TrB), and function 7 combines the posterior articular surface variables (LPAS, BPAS). Lastly, function 8 was created by combining the three variables that presented the highest accuracies in the univariate analysis (TL, TrL, LPAS). Original accuracies range between 77.8% and 97.3% for males and between 86.2% and 98.0% for females. Overall, the best original accuracy is seen in function 1 (97.4%) which is a combination of all nine variables. This function also performs best in the leave-one-out accuracies (93.6%). The length function performs better than the width and height functions in both the original group and the leave-one-out group.

Table 4.6: Direct multivariate discriminant function analysis. M = male, F = female, Avg = average.

Function	Variables	Unstandardized coefficient	Wilks' Lambda	Centroids		Original accuracy (%)			Leave-one-out accuracy (%)		
				M	F	M	F	Avg	M	F	Avg
1. All nine variables	TL	0.231	0.316	-1.396	1.509	97.3	97.5	97.4	94.6	92.5	93.6
	TW	-0.122									
	TH	0.089									
	TrL	0.060									
	TrB	-0.030									
	HNL	-0.054									
	HH	-0.087									
	LPAS	0.250									
	BPAS	0.224									
	Constant	-21.450									
2. All length variables	TL	0.198	0.328	1.500	-1.330	87.2	97.7	92.5	87.2	97.7	92.5
	TrL	0.062									
	HNL	-0.039									
	LPAS	0.252									
	Constant	-20.655									
3. All width variables	TW	0.203	0.455	1.187	-0.990	86.7	94.4	90.6	86.7	92.6	89.7
	TrB	0.121									
	BPAS	0.247									
	Constant	-17.006									
4. All height variables	TH	0.422	0.562	0.972	-0.788	78.7	86.2	82.5	76.6	84.5	80.6
	HH	0.092									
	Constant	-15.767									
5. All talar variables	TL	0.208	0.434	1.283	-0.995	77.8	89.7	83.8	77.8	89.7	83.8
	TW	0.107									
	TH	0.074									
	Constant	-18.496									
6. All trochlear variables	TrL	0.345	0.465	1.150	-0.979	89.1	87.0	88.1	87.0	87.0	87.0
	TrB	0.220									
	Constant	-17.541									
7. All posterior articular surface variables	LPAS	0.372	0.413	1.325	-1.051	89.1	94.8	92.0	89.1	94.8	92.0
	BPAS	0.213									
	Constant	-16.372									
8. Highest single variables	TL	0.238	0.380	1.345	-1.187	86.7	98.0	92.4	84.4	96.1	90.3
	TrL	0.138									
	LPAS	0.109									
	Constant	-19.126									

4.6.2 Stepwise discriminant function

Table 4.7 shows the unstandardized discriminant function coefficients, the constant, and the original and leave-one-out accuracies of a function that was created following a stepwise analysis. All nine variables were entered into the stepwise discriminant function analysis. Through this method the TL and LPAS were selected to create the best function. The function that is created is the following:

$$y = (0.211 \times TL) + (0.246 \times LPAS) - 20.500.$$

The y is the score of the discriminant function. When this score is lower than the sectioning point of 0.000 the individual can be identified as female, when the score is higher than the sectioning point the individual can be identified as male. The original accuracy and the leave-one-out accuracy show that the females (96.6%; 94.8%) score higher than the accuracies of the males (80.4%; 80.4%).

Table 4.7: Stepwise multivariate discriminant function analysis. M = male, F = female, Avg = average.

Function	Variables	Unstandardized coefficient	Wilks' Lambda	Centroids		Original accuracy (%)			Leave-one-out accuracy (%)		
				M	F	M	F	Avg	M	F	Avg
1.	TL	0.211	0.351	1.397	-1.292	80.4	96.6	88.5	80.4	94.8	87.6
	LPAS	0.264									
	Constant	-20.500									

4.7 Comparison populations

Table 4.8 shows a comparison with three other populations that have been previously studied. This data is collected from studies on Korean, white South African, and Greek populations (Bidmos & Dayal, 2003; Lee et al., 2012; Peckmann et al., 2015). The data of this Dutch population has been entered into the function consisting of all nine variables and the function that provided the highest accuracy in their respective studies. The Korean and Greek groups showed lower overall accuracy rates with the Dutch data than their original accuracies. Interestingly, the cross-validated accuracy is slightly higher than the original accuracy in the white South African population group when it comes to the function with length variables.

Table 4.8: Comparison of the accuracies of the original study versus the accuracies with the Dutch data.

Populations	Authors	Variables of functions	Original average accuracy (%)	Cross validation of Dutch sample			
				Male (%)	Female (%)	Average (%)	Difference (%)
Korean	Lee et al. (2012)	All 9 variables	87.9	100.0	45.0	71.4	-16.5
Korean	Lee et al. (2012)	TH – TrL – LPAS	86.4	97.1	50.0	73.6	-12.8
South African whites	Bidmos & Dayal (2003)	TL – LPAS – TrL – HNL	87.5	87.2	88.9	88.1	+0.6
South African whites	Bidmos & Dayal (2003)	All 9 variables	84.2	54.0	100.0	77.0	-7.2
Greek	Peckmann et al. (2015)	All 9 variables	86.9	10.8	100.0	57.1	-29.8
Greek	Peckmann et al. (2015)	TL – LPAS – TrL – HNL	85.4	69.2	97.7	83.5	-1.9

This chapter has provided the results of the statistical analysis performed on the collected data. It is observed that the intraclass correlation coefficients are of excellent reliability for every measurement. Furthermore, the data proved to be normally distributed and Student’s *t*-tests showed *p*-values lower than 0.05 making the differences between male and female means statistically significant. Univariate analysis showed the highest accuracy for the LPAS measurement and the direct discriminant function consisting of all nine variables scored the best in the multivariate analysis with an accuracy of 97.4%. The comparison between populations showed a lesser accuracy when implementing the Dutch data into the functions of other populations than their original accuracies. The next chapter will analyse and interpret these results.

5. Discussion

This chapter will analyse, interpret, and discuss the results of the statistical tests that have been performed on the Dutch post-medieval collection of Middenbeemster. These results have been provided in the previous chapter. Subsequently, this will lead to an answer to the sub-questions and the overarching research question as to how effective the talus metric method is when estimating sex in a post-medieval Dutch population in chapter six.

5.1 Repeatability

This research is only as good as its reproducibility. In both the archaeological and the forensic field it is imperative that the measurements are well defined so that they can be easily reproduced by other researchers (Peckmann et al., 2015, p. 17). In order to test for repeatability this research observed the intra-observer error through the use of the intraclass correlation coefficient. All coefficients in this research are over 0.9 meaning that the measurements are of excellent repeatability. The lowest value is found for the talar height (0.946; TH) and the highest value is seen in the length of the posterior articular surface (0.991; LPAS). The intraclass correlation coefficient was used following the example by Lee et al. (2012) which produced similar values going into the range of excellent repeatability (Lee et al., 2012, p. 167). Although other studies have tested intra-observer errors in various other ways, these studies have come to similar conclusions: the parameters of the measurements are defined well enough in order for this type of research to be easily reproduced. This can either be for new discriminant analysis research on other populations or for this research to be applied to other Dutch collections.

5.2 Univariate analysis and sexual dimorphism

The descriptive statistics show that the mean of the males lies higher than the mean of the females in every measurement. The same can be observed in all of the studies that have been discussed in chapter two of this research (Abd-Elaleem et al., 2012; Alonso-Llamazares & Pablos, 2019; Bidmos & Dayal, 2003; Curate et al., 2021; Gualdi-Russo, 2007; Harris & Case, 2012; Indra et al., 2021; Lee et al., 2012; Mahakkanukrauh et al., 2014; Murphy, 2002; Navega et al., 2014; Peckmann et al., 2015; Steele, 1976). This is not a surprising result as differences in size is one of the major elements in sexual dimorphism (White & Folkens, 2005, p. 32). The biggest discrepancy between means can

be seen in the TL (7.18 mm) while the smallest difference is seen in the HNL (1.81 mm). The Student's *t*-tests that have been executed for every measurement show a *p*-value of <0.001 while the significance level was set at 0.05. This means that even though there is a bigger discrepancy between the means of the TL than between the means of the HNL, all differences between the means of males and females are in fact statistically significant. Subsequently, this means that every measurement and therefore the talus of the individuals of this Dutch population are indeed sexually dimorphic.

Steele (1976) stated in his research that individual measurements are of little value in sex estimation due to the large overlap between ranges of the males and females. This overlap essentially means that the largest female can also be estimated as a small male (Steele, 1976, p. 584). A similar overlap between ranges can also be seen in this Dutch population, however, the obtained accuracies for the individual measurements of this study are in stark contrast with Steele (1976). Steele (1976) was only able to obtain an accuracy over 80% for one measurement, the length of the talus, which made this measurement an exception to his rule (Steele, 1976, p. 584). This research has produced average accuracies over 80% for a total of six measurements: the talar length (TL), talar width (TW), talar height (TH), trochlear length (TrL), length of the posterior articular surface (LPAS), and the breadth of the posterior articular surface (BPAS) (table 5.1). However, it can be observed that even though the average accuracies of the talar width (TW) and the talar height (TH) are over 80% there is around a 20% difference between male and female accuracies. These measurements perform better on females than on males. Therefore, these two measurements should be used with caution.

Table 5.1: The highest scoring leave-one-out accuracies in descending order.

Variable	Leave-one-out accuracy (%)		
	Male	Female	Average
LPAS	82.6	94.9	88.8
TL	80.9	88.9	84.9
TrL	80.4	89.3	84.9
TW	71.1	91.5	84.9
BPAS	78.7	88.7	83.7
TH	72.3	90.6	81.5

Overall, it can be observed that the variables involving length measurements (TL, 84.9%; TrL, 84.9%; LPAS, 88.8%) performed best and are the most sexually dimorphic. The length of the posterior articular surface (LPAS) can be considered to be the overall best sexual dimorphism discriminator in this Dutch post-medieval population and therefore the most reliable. A Korean sample also obtained the length of the posterior articular surface as the most sexually dimorphic measure (Lee et al., 2012). This is, on the other hand, in disagreement with studies on Mixed-American, Prehistoric New Zealander, white South African, Egyptian, Greek, and Swiss populations whose best sexual dimorphism discriminator is the talar length (Abd-Elaleem et al., 2012; Alonso-Llamazares & Pablos, 2019; Bidmos & Dayal, 2003; Indra et al., 2021; Murphy, 2002; Peckmann et al., 2015; Steele, 1976). This dissimilarity can likely be explained by the differences that occur in various intrinsic and extrinsic factors like genetics and physical labour that affect the sexual dimorphism in populations (Peckmann et al., 2015, p. 18).

5.3 Multivariate analysis versus univariate analysis

For the direct and stepwise discriminant function analysis multiple combinations of measurements were created. The overall range of accuracies lies between 82.5% - 97.4% for the original group and between 80.6% - 93.6% for the leave-one-out cross-validation group. These results are similar to the accuracies that researchers obtained on Korean (72.1% - 87.1%), Greek (86.7% - 96.5%), white South African (80.8% - 85.8%), Thai (86.7% - 95.9%), and Egyptian (83.6% - 85.5%) populations (Abd-Elaleem et al., 2012, p. 75; Bidmos & Dayal, 2003, p. 326; Lee et al., 2012, p. 168; Mahakkanukrauh et al., 2014, p. 152.e5; Peckmann et al., 2015, p. 17). Comparing these multivariate ranges to those of the univariate analysis it can be observed that the range of the single measurements lies between 62.3% and 88.8% for both the original as for the cross-validated group. This shows that the range of the direct and stepwise functions is not only higher but also less broad. All functions pass the preferred 80% accuracy mark while this is not the case for the single measurements.

The function that performed best with a cross-validated average accuracy of 93.6% is the function consisting of all nine variables. However, notwithstanding that this function performs best statistically, it might not be the preferred function in practice. In archaeological and forensic contexts there is the possibility that the talus in question is either damaged, incomplete, or both due to post-depositional processes and taphonomy, despite its high recovery rate. This might prevent one or two

measurements from being taken and therefore will render this function useless in these situations as it requires all nine measurements to be available (Peckmann et al., 2015, p. 18). Therefore, functions 2 and 7 might be more useful when applied in practice (table 4.2).

Univariate analysis showed that three length variables performed best, therefore, the decision was made to enter these three top performing length measurements (TL, TrL, LPAS) as a function (8) in direct discriminate function analysis. This performed fairly well with an average accuracy of 90.3%. Interestingly, the function (2) that consists of all four length measurements (TL, TrL, HNL, LPAS) performed slightly better (92.5%) despite the fact that the HNL measurement (62.3%) is the worst performing measurement in the univariate analysis. Both these length functions performed better than all other functions. A similar observation has also been made in the studies on Greek and white South African populations (Bidmos & Dayal, 2003, p. 326; Peckmann et al., 2015, p. 18). This once again suggests that length variables are the best discriminators of sex within this Dutch population.

Besides functions of all nine variables and length variables, combinations were made for the other two dimensions (width, height) and the three elements of the talus. The two dimension functions of width (3) and height (4) contain, similarly to the length functions, two measurements that did not perform well in univariate analysis. The TrB reached an accuracy of 76.7% individually and an accuracy of 89.7% when combined with the other width variables (TW, BPAS). The combination of the TH and the HH in the height function reached an accuracy of 80.6% while the HH alone reached an accuracy as low as 73.2%. A similar occurrence is seen in the talar (5), trochlear (6), and articular surface (7) functions. All combinations have a higher accuracy rate than the measurements reach individually. The only exception is the TL measurement in the talar function. The TL reaches an accuracy of 84.9% when analysed alone and an accuracy of 83.8% when combined with the TW and TH.

Table 5.2: Functions and their original and leave-one-out accuracies.

Function	Original accuracy (%)			Leave-one-out accuracy (%)		
	Male	Female	Average	Male	Female	Average
1	97.3	97.5	97.4	94.6	92.5	93.6
2	87.2	97.7	92.5	87.2	97.7	92.5
3	86.7	94.4	90.6	86.7	92.6	89.7
4	78.7	86.2	82.5	76.6	84.5	80.6
5	77.8	89.7	83.8	77.8	89.7	83.8
6	89.1	87.0	88.1	87.0	87.0	87.0
7	89.1	94.8	92.0	89.1	94.8	92.0
8	86.7	98.0	92.4	84.4	96.1	90.3

1 = TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS; 2 = TL, TrL, HNL, LPAS; 3 = TW, TrB, BPAS; 4 = TH, HH; 5 = TL, TW, TH; 6 = TrL, TrB; 7 = LPAS, BPAS; 8 = TL, TrL, LPAS.

The stepwise created function combines the two best performing single measurements, the TL and LPAS. Generally, with the exception of functions 4, 5, and 6, the direct functions generated a better accuracy rate than the stepwise function (87.6%). This still outperforms all single measurements apart from the LPAS.

Overall, combinations of measurements outperform individual measurements. However, individual measurements should not be completely disregarded. As the study by Lee and colleagues (2012) mentions there might be some merit to using the demarcation points of single measurements when in the field for a quick observation. Later, during full investigation in the lab, the functions can be used to provide more certainty with a higher percentage of accuracy (Lee et al., 2012, p. 170).

5.4 Population specificity

Metric methods are generally population specific due to the population specificity of sexual dimorphism. This differentiation of size and shape between the sexes is influenced by intrinsic and extrinsic factors during growth. These factors can differ per population and can be differences in genetics, physical labour, and diet (DiGangi & Moore, 2012, p. 93; Peckmann et al., 2015, p. 18). Therefore, there is the assumption that the discriminant function analysis of one population does not work on another population (Alonso-Llamazares & Pablos, 2019, p. 4927; White & Folkens, 2005, p. 32). However, the study by Alonso-Llamazares and Pablos (2019) on fossil remains discusses the possibility that high accuracy rates can still be obtained when an already investigated population displays similar proportions in the bone as the population that is under investigation (Alonso-Llamazares & Pablos, 2019, p. 4929; Indra et al., 2021, p. 556). This study has made a comparison of the obtained data of this post-medieval

Dutch population and three other previously researched populations by entering Dutch data into two of their highest performing functions.

5.4.1 Peckmann et al. (2015)

This research compared its data with a Greek population from a modern collection. This study by Peckmann and colleagues (2015) has been able to obtain accuracy rates between 82.4% and 86.9%. The Dutch data was entered into the function that showed the best accuracy in this Greek population. In this case this was the function consisting of all nine variables (Peckmann et al., 2015, p. 17). The average accuracy that has been obtained in this cross-validation is 57.1%. This is a difference of 29.8% when compared to the 86.9% that was obtained in the original study. The Dutch males in particular do not share similar proportions in all nine measurements as the Greek males which can be observed by the low accuracy rate in males (10.8%). Despite the fact that this function performs best, Peckmann and colleagues suggest that it might not be the most preferable function as it requires all nine variables. The study suggests that the function consisting of all length variables might prove more useful in practice. Therefore, this function has also been cross-validated with the Dutch data. This function shows less of a difference between the original accuracy (85.4%) and the cross-validated accuracy (83.5%) in comparison to the first function with only a small difference of 1.9%.

5.4.2 Lee et al. (2011)

The study by Lee et al. (2011) on a Korean population has also been cross-validated with the Dutch data with the highest grossing function consisting of all nine measurements. Despite the fact that the cross-validated accuracy is higher (71.4%) than the accuracy that was obtained with the Greek population, it is still lower than the accuracy that was obtained with the original Korean data (87.9%). This is a difference of 16.5%. There is also a big discrepancy between males and females. This is most likely due to a lower mean of Korean males in relation to the higher mean of Dutch males. The 100% accuracy displays that every male in the Dutch collection is larger than the cut-off point of Korean males. The lower accuracy in females is most likely because Dutch females are also larger than their Korean counterparts. So much so that Dutch females already reach the size of Korean males and are therefore estimated as males in the analysis. The same can be observed in the second best scoring function consisting of the talar height (TH), the trochlear length (TrL), and the length of the posterior articular facet (LPAS). In this function the females perform slightly better and the males slightly worse than in the

previous function, however, there is still a large discrepancy between males (97.1%) and females (50.0%) as well as a large difference (12.8%) between the original accuracy and the Dutch accuracy. This can again be explained through the larger size of the males and females of the Dutch population.

5.4.3 *Bidmos and Dayal (2003)*

Lastly, a white South African population researched by Bidmos & Dayal (2003) has been cross-validated with the Dutch population of this research. The function that reached the highest accuracy in this South African research consists of the talar length (TL), the length of the posterior articular facet (LPAS), and the head-neck length (HNL) (Bidmos & Dayal, 2003, p. 325). Interestingly, this cross-validated accuracy (88.1%) is very similar to the original accuracy (87.5%) obtained in the South African research. There is also less discrepancy between males (87.2%) and females (88.9%) in comparison to the other two tested populations. Both males and females of Dutch descent most likely share similar proportions in the TL, LPAS, and HNL of the talus with the males and females of South African descent. One reason for this might be the migratory history of this particular white South African population. Bidmos and Dayal (2003) state that this population is made up of migrants from the Netherlands, Germany, France, and other European countries (Bidmos & Dayal, 2003, p. 323). In this study the function with all nine variables did not reach the best accuracy, however, the function is still cross-validated with Dutch data as it has also been investigated in the other two populations. It can be observed that this function performs less with the Dutch data than the function consisting of length variables. The cross-validated accuracy (77.0%) is less than the original accuracy (84.2%) as well. It does however score better than the accuracies that were obtained in the other two comparisons.

Overall, the accuracies reached in this research (82.5% - 97.4%) are higher than the accuracies that have been obtained in the cross-validation analysis (57.1% - 88.1%). This shows that Dutch data does not work as well in discriminant function equations from Greek, Korean, and white South African populations as it does in Dutch-specific equations. Due to the nature of sexual dimorphism this is not a surprising result. The intrinsic and extrinsic factors that influence the development of the human body differ across continents and countries (DiGangi & Moore, 2012, p. 93). However, the functions consisting solely of length variables performed similarly in both the Greek and South African populations. This could suggest that length measurements might be a common

denominator of sex and can be applied among different populations. Unfortunately, the Korean study did not execute an all length variable function which is why this hypothesis could not be tested with the Korean population.

6. Conclusion

There are various methods available that can provide estimations of sex. These methods, however, come with disadvantages. For example, aDNA analysis is still rather expensive and morphological methods of the skull and pelvis are not applicable to damaged or fragmented skeletal elements (Bidmos & Dayal, 2003, p. 322; DiGangi & Moore, 2012, p. 95; Jerković et al., 2018, p. 44). Therefore, it is necessary to consider methods that are equipped to handle these drawbacks. Metric methods can prove a useful solution as it can be applied to a greater variety of bones, has limited costs, and requires less expertise. This type of method, however, also has its flaws as it is generally population specific (Bidmos & Dayal, 2003, p. 322). This means that every bone that is considered for sex estimation needs its own population specific equations. Over the years research has been done on a wide variety of bones in numerous populations, the talus being one of them. Due to its compact and robust nature this foot bone is often recovered while fairly intact in archaeological and forensic contexts and has therefore been studied in numerous populations (Alonso-Llamazares & Pablos, 2019, p. 4928; Gualdi-Russo, 2007, p. 151; Indra et al., 2021, p. 556; Navega et al., 2014, p. 651).

The present study has extended current metric sex estimation methods in the Netherlands by researching the talus and creating population specific equations. It has done so by researching the post-medieval Middenbeemster collection with a sample of 111 individuals of known sex (47 males, 64 females). Based on previous studies on the talus a total of nine measurements (TL, TW, TH, TrL, TrB, HNL, HH, LPAS, BPAS) were taken of primarily the left tali. This chapter will synopsise the main outcomes and results of this research. This will provide an answer to the main research question of this study: how effective is the talus metric method in estimating sex in a post-medieval Dutch population? It will also provide the limitations of this study and suggestions for research in the future.

6.1 Sub-questions

6.1.1 Which measurements are the most sexually dimorphic and thus most reliable in sex estimation?

Sexual dimorphism of the talus in this Dutch population was established for all nine measurements: the talar length (TL), the talar width (TW), the talar height (TH), the trochlear length (TrL), the trochlear breadth (TrB), the head-neck length (HNL), the head height (HH), the length of the posterior articular surface (LPAS), and the breadth of the posterior articular surface (BPAS). This means that all nine variables are viable indicators of sex in this Dutch population. The LPAS in particular can be considered as the best discriminator of sexual dimorphism as it reached an average accuracy as high as 88.8% and thus it is the most reliable in sex estimation. The other measurements also reached accuracy levels over 80% with the exception of the TrB, HNL, and HH. Therefore, these three measurements are not as effective in estimating sex as the other six measurements (TL, TW, TH, TrL, LPAS, BPAS) when used individually. Another observation in this individual measurement assessment is that the length measurements, with the exception of the HNL, scored the highest accuracy rates in both males and females. Therefore, one should opt for the length variables when using individual measurements of the talus to estimate sex in this Dutch population.

6.1.2 Does a combination of measurements perform better than a single measurement in sex estimation?

Six of the nine individual measurements reached relatively high accuracy rates, however, analysis showed that combinations of the measurements perform better than individual measurements. This can be quickly observed in the ranges of accuracy both types obtained. Individual measurements have a range of 62.3% to 88.8% while combined measurements have a range of 80.6% and 93.6%. Not only is this a higher and less broad range, it also displays that all functions have an accuracy over 80% which is preferred when estimating sex. Analysis shows that combining all nine variables is most accurate when using the talus to estimate sex, however, this requires the talus to be completely intact meaning that not a single measurement can be absent when using this function. Despite the fact that the talus is often recovered and generally fairly intact due to its dense and compact structure it is not immune to damages. Therefore, it is more likely that functions using less variables will be preferred in practice (Peckmann et al.,

2015, p. 18). The function that proved to be the second most effective is the function consisting of all the length variables (TL, TrL, HNL, LPAS) reaching an accuracy of 92.5%. This shows that length measurements are the best discriminators of sex in a Dutch population.

6.1.3 How accurate are formulas of Bidmos & Dayal (2003), Lee et al. (2012), and Peckmann et al. (2015) when using the data of the Dutch population? Is it indeed population specific?

This research has demonstrated the need for population specific discriminant function equations for the estimation of sex by entering Dutch data into functions of three other populations that have used a metric method on the talus: Greek, Korean, and South African. Accuracies that were reached when entering Dutch data into the population specific equations were generally lower than the accuracies the studies obtained with their original data. However, the length functions of the South African and Greek populations did not perform poorly with the Dutch data showing less population specificity. This could potentially mean that length variables are a common denominator of sex and can be used among different populations.

6.2 Main research question

6.2.1 How effective is the talus metric method in estimating sex in a post-medieval Dutch population?

This study has explored both morphological and metric methods that have been developed to estimate the sex of individuals in order to create a osteoarchaeological biological profile. This profile can aid in understanding the demography and social construction of a population within an archaeological context and in identifying unknown individuals in a forensic context (Blau & Ubelaker, 2016, p. 261; DiGangi & Moore, 2012, p. 91; Scheuer, 2002, p. 297). These morphological methods can be applied to the pelvis and the skull of an individual and provide accuracies between 85% and 96% (DiGangi & Moore, 2012, p. 96; Phenice, 1969, p. 300; Walker, 2008, p. 46). Metric methods on the other hand can and have been applied to a large number of bones. Metric studies on the humerus have reached accuracies as high as 97.8% and studies on the radius and ulna have reached accuracies of 98.2% and 96.5% respectively (Duangto & Mahakkanukrauh, 2019, p. 41). Studies on the femur and patella were able to reach accuracies of 95.7% (Carvallo & Retamal, 2020, p. 5) and 85% (Peckmann & Fisher, 2018, p. 4). These population specific metric methods are not inferior to the

accuracies reached in morphological methods and are quite effective in the estimation of sex either as supporting evidence or as replacement when morphological traits are not available. Studies on the talus have also shown their effectiveness in sex estimation by reaching accuracies of at least 80% in their multivariate functions (Abd-Elaleem et al., 2012; Alonso-Llamazares & Pablos, 2019; Bidmos & Dayal, 2003; Curate et al., 2021; Gualdi-Russo, 2007; Harris & Case, 2012; Indra et al., 2021; Lee et al., 2012; Mahakkanukrauh et al., 2014; Murphy, 2002; Peckmann et al., 2015; Steele, 1976). This study has proofed through high accuracy rates that within this post-medieval Dutch population the metric method of the talus either in complete or fragmentary state is highly effective in estimating sex. Both individual measurements as combinations of measurements can be used, but a combination of measurements is preferred as it provides higher accuracies proofing a better ability to estimate sex.

6.3 Limitations

A limitation of this research is the unavailability of the raw data of the studies by Bidmos & Dayal (2003), Lee et al. (2012), and Peckmann et al. (2015). Due to the lack of raw data this research was only able to implement the Dutch data into the functions created by the authors to test for accuracy rates. It would however be beneficial to test the means of the Dutch population and the white South African, Korean, and Greek populations for statistically significant differences. This could potentially support the observations made from the accuracy rates. Without the raw data, however, this was not possible for this research.

6.4 Future research

This research has provided evidence that the talus can be used for sex estimation in a post-medieval Dutch population. Collections of this time period might differ from collections reaching the modern era. It is therefore suggested that the functions created in this research are cross-validated with a modern Dutch collection. It would also be interesting to test the functions on older collections, for example a Roman collection. Comparing multiple Dutch populations can perhaps showcase if there are differences in the size of tali through time and therefore its ability to estimate sex. This should pose no problem as this research proved to have excellent repeatability based on the intra-observer error and can therefore be easily reproduced on other Dutch populations.

In this research it has also become clear that the measurements of length performed best in both univariate and multivariate analysis. Dutch data in the length functions of other populations also performed rather well. This raises the question if length measurements might be a common sex denominator widely applicable amongst populations. Therefore, it might be interesting to research if there is a relationship between the talus and stature for the estimation of sex.

Abstract

The estimation of sex is very important in the analysis of skeletal remains in both archaeological and forensic contexts. It is generally the first step that is taken to establish the osteoarchaeological biological profile as the other elements of the profile (e.g. stature and age-at-death) are sex specific. Morphological sex estimation methods require a visual assessment of various features of the pelvis and skull, however, in practice the pelvis and skull are not always assessable when the skeletal elements are in a fragmentary state or are completely absent from the context. Therefore, it is useful to create methods that can be applied to a wider variety of bones and can handle incomplete skeletal elements. Metric methods have shown high accuracy rates on bones like the radius, ulna, femur, calcaneus, and many more. Previous research has shown that the talus is also a sexually dimorphic bone that can be used for sex estimation, however, the discriminant function equations that are created in previous research are population specific. This study extends the line of research done on the talus by testing its sex estimation ability in a post-medieval Dutch population using the Middenbeemster collection. A total of nine measurements were obtained from 111 individuals of known sex (47 males and 64 females). These measurements are based on previous studies and are the talar length, talar width, talar height, trochlear length, trochlear breadth, head-neck length, head height, length of the posterior articular surface, and the breadth of the posterior articular surface. Descriptive statistics and discriminant function analysis was applied to the acquired data. Basic statistics showed the sex discriminating potential of the talus in this Dutch post-medieval population. Univariate analysis reached accuracies between 62.3% and 88.8% while multivariate analysis reached even higher accuracies of between 82.5% and 97.4%. To further investigate the need for population specific equations, the Dutch data has been entered into multiple functions obtained from three previously researched populations. The accuracies obtained here proved to be less than the accuracies obtained when using their own data suggesting the need for population specific equations. In conclusion, this study has established that the talus of Dutch individuals is effective in sex estimation.

Bibliography

- Alonso-Llamazares, C., & Pablos, A. (2019). Sex estimation from the calcaneus and talus using discriminant function analysis and its possible application in fossil remains. *Archaeological and Anthropological Sciences*, *11*(9), 4927-4946. doi:10.1007/S12520-019-00855-Y
- Bidmos, M. A., & Dayal, M. R. (2003). Sex Determination From the Talus of South African Whites by Discriminant Function Analysis. *The American Journal of Forensic Medicine and Pathology*, *24*(4), 322-328. doi:10.1097/01.paf.0000098507.78553.4a
- Bidmos, M. A., & Mazengenya, P. (2021). Accuracies of discriminant function equations for sex estimation using long bones of upper extremities. *International Journal of Legal Medicine*, *135*, 1095-1102. doi:10.1007/s00414-020-02458-y
- Blau, S., & Ubelaker, D. H. (2016). *Handbook of Forensic Anthropology and Archaeology*. Routledge .
- Bobak, C. A., Barr, P. J., & O'Malley, A. J. (2018). Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Medical research Methodology*, *18*, 1-11. doi:10.1186/s12874-018-0550-6
- Carvalho, D., & Retamal, R. (2020). Sex estimation using the proximal end of the femur on a modern Chilean sample. *Forensic Science International: Reports*, *2*, 1-7. doi:10.1016/j.fsir.2020.100077
- Charisi, D., Eliopoulos, C., Vanna, V., Koiliias, C. G., & Manolis, S. K. (2001). Sexual Dimorphism of the Arm Bones in a Modern Greek Population. *Journal of Forensic Science*, *56*(1), 10-18. doi:10.1111/J.1556-4029.2010.01538.X
- Colman, K. L., Janssen, M. C., Stull, K. E., van Rijn, R. R., Oostra, R. J., de Boer, H. H., & van der Merwe, A. E. (2018). Dutch population specific sex estimation formulae using the proximal femur. *Forensic Science International*, *286*, 268e1-268e8. doi:10.1016/j.forsciint.2017.12.029
- Curate, F., d'Oliveira Coelho, J., & Silva, A. M. (2012). CalcTalus: an online decision support system for the estimation of sex with the calcaneus and talus. *Archaeological and Anthropological Sciences*, *13*(5), 74.e1-74.e10. doi:10.1007/S12520-021-01327-Y
- DiGangi, E., & Moore, M. (2012). *Research Methods in Human Skeletal Biology*. Academic Press. doi:10.1016/b978-0-12-385189-5.00004-2
- Dirkmaat, D. C. (2012). *A companion to Forensic Anthropology*. John Wiley & Sons.
- Duangto, P., & Mahakkanukrauh, P. (2019). Sex estimation from upper limb bones in a Thai population. *Anatomy & Cell Biology*, *53*, 36-43. doi:10.5115/acb.19.179
- Fletcher, M., & Lock, G. R. (2005). *Digging numbers - Elementary Statistics for Archaeologists*. Oxford University School of Archaeology .

- Giles, E., & Elliot, O. (1963). Sex Determination by Discriminant Function Analysis of Crania. *American Journal of Physical Anthropology*, 21(1), 53-68. doi:10.1002/ajpa.1330210108
- Gualdi-Russo, E. (2007). Sex determination from the talus and calcaneus measurements. *Forensic Science International*, 171, 151-156. doi:10.1016/J.FORSCIINT.2006.10.014
- Hakvoort, A. (2013). *De begravingen bij de Keyserkerk te Middenbeemster*. Hollandia (reeks 464).
- Harris, S. M., & Case, D. T. (2012). Sexual Dimorphism in the Tarsal Bones: Implications for Sex Determination. *Journal of Forensic Sciences*, 57(2), 295-305. doi:10.1111/j.1556-4029.2011.02004.x
- Indra, L., Vach, W., Desideri, J., Besse, M., & Pichler, S. L. (2021). Testing the validity of population-specific sex estimation equations: An evaluation based on talus and patella measurements. *Science & Justice*, 61, 555-563. doi:10.1016/J.SCIJUS.2021.06.011
- Isaac, J. L. (2005). Potential causes and life-history consequences of sexual size dimorphism in mammals. *Mammal Review*, 35(1), 101-115. doi:https://doi-org.ezproxy.leidenuniv.nl/10.1111/j.1365-2907.2005.00045.x
- Jerković, I., Bašić, Ž., Kružić, I., & Anđelinović, Š. (2018). Creating reference data on sex for ancient populations using the Probabilistic Sex Diagnosis method: A validation test using the results of aDNA analysis. *Journal of Archaeological Science*, 94, 44-50. doi:10.1016/j.jas.2018.04.007
- Karsten, J. (2018). A test of the preauricular sulcus as an indicator of sex. *American Journal of Physical Anthropology*, 165(3), 604-608. doi:10.1002/AJPA.23372
- Kranioti, E. F., Işcan, M. Y., & Michalodimitrakis, M. (2008). Craniometric analysis of the modern Cretan population. *Forensic Science International*, 180, 110.e1-110.e5. doi:10.1016/J.FORSCIINT.2008.06.018
- Lee, U., Han, S., Park, D., Kim, Y., Kim, D., Chung, I., & Chun, M. (2012). Sex Determination from the Talus of Koreans by Discriminant Function Analysis. *Journal of Forensic Sciences*, 57(1), 166-171. doi:10.1111/J.1556-4029.2011.01914.X
- Mahakkanukrauh, P., Praneatpolgrang, S., Ruengdit, S., Singsuwan, P., Duangto, P., & Case, T. D. (2014). Sex estimation from the talus in a Thai population. *Forensic Science International*, 240, 152e1-152e8. doi:10.1016/J.FORSCIINT.2014.04.001
- Mahato, N. (2011). Morphology of sustentaculum tali: Biomechanical importance and correlation with angular dimensions of the talus. *The Foot*, 21(4), 179-183. doi:10.1016/J.FOOT.2011.06.001
- Martin, R., & Knussmann, R. (1988). *Anthropologie: Handbuch der Vergleichenden Biologie des Menschen*. Gustav Fischer.

- Murphy, A. M. (2002). The talus: sex assessment of prehistoric New Zealand Polynesian skeletal remains. *Forensic Science International*, 128, 155-158. doi:10.1016/S0379-0738(02)00189-5
- Navega, D., Vicente, R., Vieira, D. N., Ross, A. H., & Cunha, E. (2014). Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. *International Journal of Legal Medicine*, 129, 651-659. doi:10.1007/S00414-014-1070-5
- Palmer, J. L. (2019). *A sense of society: enthesal change as an indicator of physical activity in the Post-Medieval Low Countries: potential and limitations*. Doctoral dissertation.
- Patriquin, M. L., Steyn, M., & Loth, S. R. (2005). Metric analysis of sex differences in South African black and white pelves. *Forensic Science International*, 147, 119-127. doi:10.1016/j.forsciint.2004.09.074
- Peckmann, T. R., & Fisher, B. (2018). Sex estimation from the patella in an African American population. *Journal of Forensic and Legal Medicine*, 54, 1-7. doi:10.1016/J.FLM.2017.12.002
- Peckmann, T. R., Meek, S., Dilkie, N., & Rozendaal, A. (2016). Determination of sex from the patella in a contemporary Spanish population. *Journal of Forensic and Legal Medicine*, 44, 84-91. doi:10.1016/j.jflm.2016.09.007
- Peckmann, T. R., Orr, K., Meek, S., & Manolis, S. K. (2015). Sex determination from the talus in a contemporary Greek population using discriminant function analysis. *Journal of Forensic and Legal Medicine*, 33, 14-19. doi:https://doi.org/10.1016/j.jflm.2015.03.011
- Phenice, T. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, 30(2), 297-301. doi:10.1002/AJPA.1330300214
- Reichs, K. J. (1998). *Forensic Osteology: Advances in the Identification of Human Remains*. Charles C Thomas.
- Robinson, M. S., & Bidmos, M. A. (2011). An assessment of the accuracy of discriminant function equations for sex determination of the femur and tibia from a South African population. *Forensic Science International*, 206, 212.e1-212.e5. doi:10.1016/j.forsciint.2010.12.009
- Scheuer, L. (2002). Application of osteology to forensic medicine. *Clinical Anatomy*, 15(4), 297-312. doi:10.1002/CA.10028/FORMAT/PDF
- Shook, B., Nelson, K., Aguilera, K., & Braff, L. (2019). *Explorations: An Open Invitation to Biological Anthropology*. American Anthropology Association .
- Singh, S., & Potturi, B. R. (1978). Greater sciatic notch in sex determination. *Journal of Anatomy*, 125, 619-624.
- Spradley, M. K., & Jantz, R. L. (2011). Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. *Journal of Forensic Sciences*, 56(2), 289-296. doi:10.1111/j.1556-4029.2010.01635.x

- Steele, D. G. (1976). The Estimation of Sex on the Basis of the Talus and Calcaneus. *American Journal of Physical Anthropology*, 45, 581-588.
- Ubelaker, D. H., & DeGaglia, C. M. (2017). Population variation in skeletal sexual dimorphism. *Forensic Science International*, 278, 407-414.
doi:<https://doi.org/10.1016/j.forsciint.2017.06.012>
- van Spelde, F. J., & Hoogland, M. L. (2018). Context and material culture of the 18th- and 19th- century cemetery of Middenbeemster, the Netherlands. En R. M. van Oosten, R. Schats, K. Fast, N. Arts, & H. M. Bouwmeester, *The urban graveyard: archaeological perspectives* (págs. 307-336). Sidestone Press.
- Veldhuis, J. D., Roemmich, J. N., Richmond, E. J., Rogol, A. D., Lovejoy, J. C., Sheffield-Moore, M., . . . Bowers, C. Y. (2005). Endocrine Control of Body Composition in Infancy, Childhood, and Pubertt. *Endocrine Reviews*, 26(1), 114-146.
doi:10.1210/er.2003-0038
- Walker, P. L. (2005). Greater sciatic notch morphology: Sex, age, and population. *American Journal of Physical Anthropology*, 127(4), 385-391.
doi:10.1002/AJPA.10422
- Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, 136, 39-50.
doi:10.1002/ajpa.20776
- Walker, P. L., & Cook, D. C. (1998). Brief Communication: Gender and Sex: Vive la Difference. *American Journal of Physical Anthropology*, 106, 255-259.
doi:10.1016/b978-0-12-385189-5.00004-2
- WEA. (1980). Recommendations for Age and Sex Diagnoses of Skeletons. *Journal of Human Evolution*, 9, 517-549.
- White, T., & Folkens, P. (2005). *The human bone manual*. Elsevier Academic Press Publications.

Other

American Academy of Orthopaedic Surgeons. Talus Fractures.
<https://orthoinfo.aaos.org/en/diseases--conditions/talus-fractures>, accessed on 21-04-2022.