Boy or Girl

Determining the sex of subadults from the Middenbeemster collection with twelve non-metric methods

T. Vermaas, 0911089
MSc Thesis Archaeology, 1040X3053Y
Prof. Dr. M.L.P. Hoogland
Human Osteology and Funerary Archaeology
University of Leiden, Faculty of Archaeology
Eindhoven, 03-06-2014, Final version

Content

Acknowledgments	7
1. Introduction	9
1.1 Metric methods	10
1.2 Non-metric methods	14
1.3 Metric versus non-metric methods	15
1.4 Importance for archaeology	19
1.5 Middenbeemster	20
1.6 Research questions	20
2. Accuracies of methods to determine the sex of subadults	23
2.1 Accuracies of non-metric methods of the pelvis	23
2.1.1 Weaver	24
2.1.2 Schutkowski	27
2.1.2.1 Angle of the greater sciatic notch	28
2.1.2.2 Depth of the greater sciatic notch	29
2.1.2.3 Arch criterion	31
2.1.2.4 Iliac crest	33
2.1.2.5 Summary	34
2.2 Accuracies of non-metric methods of the skull	34
2.2.1 Schutkowski	35
2.2.1.1 Protrusion of the chin	35
2.2.1.2 Shape of the anterior dental arcade	36
2.2.1.3 Eversion of the gonion region	38
2.2.1.4 Summary	39
2.2.2 Loth and Henneberg	39
2.2.3 Molleson and colleagues	42
2.2.3.1The mandibular angle	43
2.2.3.2 The mentum	44
2.2.3.3 Orbital morphology	45
2.2.3.4 CFS scores	46
2.2.3.5 Summary	47

2.3 Overview and discussion 47
3. Materials and methods 49
3.1 Materials 49
3.1.1 Historical background 49
3.1.2 Excavation 50
3.1.3 Lab procedures 51
3.2 Methods 52
3.2.1 The pelvis 52
3.2.1.1 Elevation of the auricular surface 53
3.2.1.2 Angle of the greater sciatic notch 53
3.2.1.3 Depth of the greater sciatic notch 55
3.2.1.4 The arch criterion 56
3.2.1.5 The iliac crest 57
3.2.2 The skull 58
3.2.2.1 Protrusion of the chin 58
3.2.2.2 The shape of the anterior dental arcade 59
3.2.2.3 The eversion of the gonion region 59
3.2.2.4 The complete mandible 60
3.2.2.5 The mandibular angle 61
3.2.2.6 The mentum 62
3.2.2.7 The orbital morphology 63
3.3 Lab procedures 64
3.4 Statistical analyses 65
3.4.1 The χ^2 test 65
3.4.2 The ϕ test 66
3.4.3 Logistic regression 67
4. Results 69
4.1 General results 69
4.1.1 Statistical analyses 73
4.2 Results for the pelvis methods 73
4.2.1 Auricular surface 74

4.2.2 Greater sciatic notch angle 76
4.2.3 Greater sciatic notch depth 78
4.2.4 Arch criterion 80
4.2.5 Iliac crest 82
4.3 Results for the skull methods 85
4.3.1 Chin prominence 85
4.3.2 Anterior dental arcade 87
4.3.3 Eversion of the gonion region 89
4.3.4 Complete mandible 91
4.3.5 Mentum 93
4.3.6 Mandibular angle 95
4.3.7 Orbital morphology 97
4.4 Summary 98
5. Discussion 101
5.1 Low accuracy rates 101
5.1.1 Population difference 101
5.1.2 Sexual dimorphism in Dutch populations 102
5.1.3 Inter- and intraobserver error 104
5.1.4 Endocrine abnormalities, disease and malnutrition 105
5.2 Female bias 107
5.3 Age groups 107
5.3.1 Age related methods 108
5.4 Confounding factors 109
6. Conclusion 111
6.1 Further research 113
Summary 115
Samenvatting 116
Bibliography 117
List of figures 128
List of tables 132
Appendix A Overview of all the individuals and their scores 137

Acknowledgments

I would like to thank the staff of the Laboratory for Human Osteoarchaeology, Leiden University. Especially Dr. Sarah Inskip who has given me a lot of advice and comments. Menno Hoogland made it possible for me to use the subadults from the Middenbeemster collection and he was a helpful supervisor. Without the help of Sebastiaan Zeeff the statistics would still have been incomprehensible for me. In addition, I would like to thank Mariska Versantvoort for supporting me even though she is miles away.

This thesis would not have been possible without the support from my parents Hugo Vermaas and Anne van Berkel. Thank you for helping me through these last two years. Lastly, I would like to thank my sister, Emmelot. Without her support, her advice and our discussions I would not have been where I am now.

1. Introduction

It has only been since the 1960s that osteologists and bioarchaeologists have been interested in subadults. The topics of subadult mortality, child growth and child development have been widely studied since then (Saunders 2008, 117). Most information gathered about subadults consists of age, stature and diseases. This information is only meager compared to the information that is gathered about adults, namely age, stature, sex, ancestry and diseases (Sheuer and Black 2004, 1). Estimating the sex of subadults is still a problem (WEA 1980, 517-518) even while methods have been researched since the 1870s (Boucher 1957, 581-582). The basis of the idea that it is possible to estimate sex in subadults, is that fetal testosterone is present in male fetuses from the 10th week with a peak around the 15th week. After that the testosterone will decrease until puberty. It is thought that this testosterone peak creates a difference between males and females from six months in gestation onwards (Lewis 2007, 47). The differences between males and females comes best apparent during puberty after the age of ten years (Saunders 2008, 123). There are indications that at the age of eight the ilium might be sexual dimorphic (Humphrey 1998, 67). It is thought that these differences can be used to sex adults (Lewis 2007, 47). However, some researchers believe that these differences are already visible in the fetus and thus can be used to estimate sex for subadults (Weaver 1980, 191-192).

In this thesis various methods to estimate the sex of subadults will be gathered and examined. These methods will be tested on a known-sex sample from Middenbeemster to see whether the hypothesis is valid that sex can be determined on subadults.

Currently there are two ways that can be used to determine the sex of subadults. The first way is by using metric methods and the second way by using non-metric methods. These two ways and their history will be described first. After that it will be decided which one of these ways is most interesting to use in this research. Then the importance of estimating the sex for subadults will be discussed together with the sample choice and the research questions.

1.1 Metric methods

Metric methods are measurements taking from the skeletons and can be seen as the quantitative aspects of the human skeleton (Pietrusewsky 2008, 487). The first research was by Hunt and Gleiser (1955) and was about sexing children with their bones and teeth. Their argument is that females mature faster than males (Lewis 2007, 48) and that this is visible in the permanent teeth and bones of subadults. For each child the teeth and bones were examined. An example: when a tooth has two-thirds of the root completed, than the mean age for males will be 84.3 months (7 years) and for a female 80.7 months (6.7 years). If the male bone age would be 7 years and the female bone age 5.36 years, than the male ages for bones and teeth overlap the most (Hunt and Gleiser 1955, 481). In this fictional case the skeleton would be sexed male with an age of 7 years. Linear equations were made both of the age and sex of males and females. For the first permanent mandibular molar the equation is y = 0.95x whereby y is the age of females and x the age of males. For the skeletal maturation the equation is y = 0.80x. A total of 93 radiographs of children from the growth study at the Harvard School of Public Health were examined and three age groups were created namely 2 years (n = 33), 5 years (n = 33) and 8 years (n = 27). For each individual the dental and bone ages were calculated with the equations and the sex was determined as the one in which the bone and dental ages agreed the most. This gave an accuracy of 76.3% (Hunt and Gleiser 1955, 481-483). In 1964 the conclusions were confirmed by Bailit and Hunt (1964).

Garn and his colleagues (1977) wanted to know the maximum extent of sexual discrimination within crown sizes and what the minimum number of teeth were that should be used to accurately estimate sex. This was done by measuring the right permanent teeth of 204 subadults. The results were used in discrimant analyses and the average accuracy was 87%. This was achieved by using six teeth (the upper and lower canines, lower second molar, upper and lower lateral incisors and the upper second premolars) and by only using the mesiodistal diameter (Garn *et al.* 1977). Black (1978) did the same kind of study, but on the deciduous teeth. He used the same sample as Garn *et al.* (1977), but he only used 133 children. A

Discriminant functions for sexing the deciduous dentition

Functions	Percent correctly classified		
	Male	Female	Total
1. Function using the maxillary teeth alone: $-2.91 + 1.512(m^1\ b\text{-}1) - 1.585(i^1\ m\text{-}d)$	69.6	57.8	63.9
2. Functions using the mandibular teeth alone: $-8.66 + 1.792(m_2\ b\text{-}1) - 1.528(i_2\ m\text{-}d)$	69.6	62.5	66.2
3. Function using mesiodistal measurements alone: $-7.64 + 1.096(m_2\ m\text{-}d) - 1.976(i^1\ m\text{-}d) + 1.439\ (\text{max. } c\ m\text{-}d)$	68.1	68.8	68.4
4. Function using buccolingual measurements alone: $-11.10 + 2.192(m^1\ b\text{-}1) - 1.924(m_1\ b\text{-}1) + 1.311(m_2\ b\text{-}1) - 1.042(i^1\ b\text{-}1)$	66.7	62.5	64.7
5. Function using mesiodistal and buccolingual measurements of the maxillary and mandibular teeth: $-8.18 + 2.343(m^1\ b\text{-}1) - 2.192(i^1\ m\text{-}d) + 2.046(m_2\ b\text{-}1) - 2.187(m_1\ b\text{-}1)$	72.5	62.5	67.7

Fig. 1. Table with five discriminant functions that Black has created from his study (Black 1978, 80).

total of twenty measurements were taken from the right teeth and these were used to created discriminant functions. He obtained an average accuracy of 75% (Black 1978). His discriminant functions are visible in figure 1. De Vito and Saunders (1990) created discriminant functions based on the deciduous teeth as well and included dental measurement of the first permanent molar. A total of twelve functions proved to be interesting for determining sex of children. Of these twelve, five had an accuracy rate over 80% (table 1). It shows us that there is sexual dimorphism between males and females. These standards can be used for the determination of sex of subadults and especially for modern forensic sciences (De Vito and Saunders 1990).

Table 1. Five functions from De Vito and Saunders with their accuracy rates (1990; 850-853).

Functions	Percent correctly classified		
	Male	Female	Total
1. 1.5 (FL R max li) + 1.091 (FL R max ci) + 0.654 (FL L max dm^2) - 1.489 (FL L max c) + 1.64 (MD R mand c) - 20.342	100%	83.3%	90.5%
2. 1.38 (FL R max li) + 0.896 (FL R max ci) + 0.357 (FL L max dm^2) - 1.474 (FL L max c) + 2.266 (MD R mand c) - 19.736	80%	80%	80%
3. 0.542 (FL R max li) + 0.279 (FL L max dm^2) - 0.723 (FL L max c) - 1.058 (MD R mand c) + 1.837 (FL L max M1) + 0.628 (MD L mand M1) - 1.692 (FL L mand M1) - 17.423	80%	90.9%	85.7%
4. 0.574 (FL R max li) + 0.393 (FL L max dm^2) - 0.371 (FL L max c) + 1.521 (FL L max M1) - 21.314	80%	90.9%	85.7%
5. 2.049 (MD R mand c) + 0.887 (MD L mand M1) - 0.516 (FL L mand M1) - 16.872	80%	90.9%	85.7%

Research on older and archaeological samples were conducted by Rösing (1983), Okazaki (2005), Cardoso (2008), Żądzińska et al. (2008) and Viciano et al. (2011). Rösing (1983) created discriminant functions by using the measurements of 55 adults which were sexed with morphognostic criteria. Four measurements were taken from the mandibular and maxillary teeth from the first incisors to the first molars. A total of 66 discriminant functions were created whereby four functions had an accuracy of 97% and 12 an accuracy between 90 and 95%. The same functions were then used to determine sex for subadults, however the accuracy could not be determined, because the sex was unknown (Rösing 1983). Okazaki (2005) used discriminant functions based on the permanent tooth crown measurements taken from two modern groups and two archaeological samples from the same regions. The sex was known from the two modern groups and this resulted in an accuracy of 97.3% for both groups. The functions were used for the archaeological samples to see if there are any differences in male:female ratio (Okazaki 2005). Cardoso (2008) used logistic regressions on adults and subadults from Lisbon. All the teeth, except the third molar, were measured and it showed that in adults the sexual dimorphism is larger than in subadults. The logistic regression showed that there was an average accuracy rate of 50 to 60% for all the teeth except for maxillary and mandibular canine which had an accuracy of 80% (Cardoso 2008). Żądzińska and colleagues conducted research on a Polish archaeological sample on the deciduous dentition. The sex of the sample was obtained by DNA analysis. Afterwards measurements were taken and a linear equation was created that could estimate the sex of subadults. Their accuracy rates lay between 69 and 88% (Żądzińska *et al.* 2008). The last research done on an archaeological sample was by Viciano and colleagues on the archaeological sample of Herculaneum. The sex of the subadults was first determined by using critized non-metric methods proposed by Schutkowski (1993). These non-metric methods are the greater sciatic notch angle, the greater sciatic notch depth, the "arch" criterion, the iliac crest, the chin prominence, the anterior dental arcade and the gonion region (Schutkowski 1993, 200-201). These non-metric methods will be described in more detail in the next section. Viciano and colleagues created discriminant functions based on the adult teeth and these were tested on the subadults. The accuracy rate is 76.2% (Viciano *et al.* 2011).

Other metric studies concerned the pelvis, mandible and basicranium. The first studies on the pelvis were only concentrated around acknowledging that there was sexual dimorphism in the subadult pelvis. These studies were done by Thomson (1899), Reynolds (1945), Boucher (1957), Weaver (1980) and Vlak et al. (2008). The earliest studies done by Thomson (1899), Reynolds (1945) and Boucher (1957) all indicate that there are metric differences between male and female pelvis from the 3rd fetal month and thereafter. Discriminant functions created by Schutkowski (1987) indicate that these functions can be used to correctly estimate sex with an accuracy over 70%. However, studies done by Weaver (1980) and Vlak *et al.* (2008) came to the conclusion that the pelvis is not sexually dimorphic within subadults.

Within the metric studies, digitization is getting more important. Within digitization the outline of the bone is scanned and landmarks are created. By putting this in a computer, differences within the outline and the landmarks can be viewed to calculate whether there is sexual dimorphism or not (Wilson *et al.* 2008, 270). Holcomb and Konigsberg (1995) digitized the iliac outline and created landmarks on the greater sciatic notch. Their conclusion was that the sciatic notch is not sexually dimorphic (Holcomb and Konigsberg 1995). Wilson et al. (2008) used the same kind of technique and created five landmarks. According to them, there is sexual dimorphism within the ilium (Wilson *et al.* 2008). The same technique and discriminant functions were also tested on a more extended sample and the results were the same (Wilson *et al.* 2011).

Two other areas have been studied in search for a way to correctly estimate the sex of subadults. These areas are the mandible and the skull. Franklin and colleagues used digitization on the mandible in 2007. Their conclusion is that the mandible is not sexually dimorphic (Franklin *et al.* 2007). The other metric study was done on the basicranium of subadults. The foramen magnum and the occipital condyles were measured and tests were done to determine whether the foramen magnum and the occipital condyles are sexually dimorphic. The results indicate that they are positively dimorphic for subadults over eight years (Veroni *et al.* 2010).

1.2 Non-metric methods

The first observation concerning sexual dimoprhism in subadults was made by Verneau in 1875 when he noted that females had larger subpubic angles than males (Verneau 1875, 71-74). After that it took more than one hundred years before an actual criterion was developed. This happened in 1980 with the research of Weaver. His research was about metric criteria and one non-metric criterion namely the elevation of the auricular surface. This method proved to be more reliable than the metric approaches with an accuracy of 73% (Weaver 1980). Mittler and Sheridan (1992) and Sutter (2003) replicated this method and got accuracies ranging between 71% and 74%. When Hunt (1990) retested the method he came to the conclusion that the elevation of the auricular surface was age-related and not sex-related. He did not use known-sex subadults so an accuracy rate could not be obtained. His conclusion is based on the male:female ratio within the sample when he used the method of Weaver (Hunt 1990).

After Weaver another method was proposed by Schutkowski (1993). His method was not only directed towards the pelvis, but also to the mandible. He proposed four reliable methods for the pelvis and three for the mandible. These methods are the greater sciatic notch angle, the greater sciatic notch depth, the "arch" criterion and the iliac crest for the pelvis; and the chin prominence, the anterior dental arcade and the gonion region for the mandible (Schutkowski 1993, 200-201). The accuracies ranged between 65 to 81% (Schutkowski 1993, 202-203). The methods of Schutkowski were replicated by Sutter (2003) with accuracies ranging between 59 to 81%. When the greater sciatic notch and angle were retested by Vlak and colleagues in 2008 the results were disappointing. Their accuracies are 53% for the greater sciatic notch angle and 58% for the depth which is barely better than chance (Vlak *et al.* 2008). In addition, Cardoso and Saunders (2008) tested the "arch" criterion and included a study into intra- and interobserver error. Their results were an average accuracy of 42% and high intra- and interobserver errors (Cardoso and Saunders 2008).

The mandibular methods of Schutkowski (1993) were also retested by Loth and Henneberg (2001) who could not replicate the results of Schutkowski (1993). Their accuracies were 33% (Loth and Henneberg 2001). They decided to create their own method which they believed could be used to estimate sex in subadults. This method is the complete mandible method. Their accuracy rates are 81% and did seem promising (Loth and Henneberg 2001). However, when it was retested by Scheuer (2002) and Coqueugniot and colleagues (2002), the results were poor. Both authors could not replicate the results of Loth and Henneberg (2002) and they questioned whether the method was reliable or not (Coqueugniot *et al.* 2002; Scheuer 2002).

The last research regarding the estimation of sex of subadults was done by Molleson et al. (1998). They created two methods for the mandible and one method for the skull. The methods for the mandible had an accuracy of 70 and 57.9% and the skull method had an accuracy of 66.7%. Their conclusion was that these three methods could be used to estimate the sex of subadults (Molleson *et al.* 1998). However, these methods have never been retested.

1.3 Metric versus non-metric methods

Both metric and non-metric methods have been investigated for the ability to estimate the sex of subadults. However, with both methods there are some problems.

A couple of the metric methods for the teeth have been based on radiographs of living children and they require realistic skeletal ages (Bailit and Hunt 1964; Hunt and Gleiser 1955) which is often difficult to get in archaeological samples. Some other researchers created discriminant functions for the teeth on living children (Black 1978; De Vito and Saunders 1990; Garn *et al.* 1977). The problems of this research is that the research was not done on an archaeological sample and that the discriminant functions could not be used on any other sample than the sample it was created on (Black 1978; De Vito and

Saunders 1990; Garn *et al.* 1977). Other researchers used the teeth of adults to create discriminat functions or logistic regressions that were used to sex subadults. In addition, all these functions are sample-specific. This means that for another sample the adults need to be measured first, then the functions need to be calculated and then the subadults might be correctly sexed. This can be time-consuming (Dawson *et al.* 2011, 81). Another confounding factor is that within the archaeological sample there need to be enough adults to create a discriminant function. Within archaeology, the sample are often small and have a poor preservation. The statistical procedures that are used for metric functions work best with many complete skeletons with known-sex. This means that functions created on small archaeological sample should be used with caution (Walker 2008, 39) Also, the accuracy rate of the function will be unknown for the subadult sample. Another problem is in the forensic sciences. For a single case it is often not known to which population the person belonged. Therefore, a discriminant function analysis cannot be used to determine the sex, because each function is population-specific and cannot be used on a different population than the original population that it was created on (Bruzek and Murail 2006, 234). Only the method of Viciano and colleagues might be applicable since they have created a discriminant function that might be applicable to other samples (Viciano *et al.* 2011). The last research on teeth was done by Żądzińska and colleagues (2008) using DNA to confirm their metric results. However, they did not create a real test to determine sex (Żądzińska *et al.* 2008).

Using teeth for estimating the sex of subadults can be limited by the state of preservation. The teeth need to be healthy which means that teeth with caries, calculus deposits, hypoplastic defects, wear, dental anomalies, trauma, fractures and taphonomic/diagenetic effects cannot be used (Viciano *et al.* 2011, 98). These teeth can give wrong measurements and can give false functions. This means that within an archaeological sample it is possible that many skeletons cannot be sexed with methods for the teeth.

The first metric research of the pelvis was concentrated around confirming that the pelvis was sexually dimorphic (Boucher 1957; Reynolds 1945; Thomson

1899; Vlak *et al.* 2008; Weaver 1980). After that, most of the metric methods of the pelvis are based on digitization of the ilium which seems promising (Holcomb and Konigsberg 1995; Wilson *et al.* 2011; Wilson *et al.* 2008), but is not feasible in all research, because special equipment need to be used for this. In addition, it is often difficult to capture subtle sexually dimorphic variations with measurements (Walker 2008, 39). The mandible was also digitized by Franklin et al. (2007),but it seems it cannot be used to determine the sex of subadults. The last metric research was concentrated around the basicranium (Veroni *et al.* 2010) and this study can be applied on other samples.

The non-metric methods seems to be most successful with the authors that created them. There some exceptions. The method of Weaver (1980) was successfully retested by Mittler and Sheridan (1992) and Sutter (2003). The methods of Schutkowski (1993) were retested by Sutter (2003) who had high accuracy rates even though the sample was quiet different from the sample used by Schutkowski (1993). This means that the methods of Schutkowski can be used on other samples. However, both the method of Weaver and the methods of Schutkowski were also retested by other authors who failed to reach the accuracy rates of the original authors. Hunt (1990) could not replicate the results of Weaver (1980). Vlak et al. (2008), Cardoso and Saunders (2008) and Loth and Henneberg (2001) could not replicate the results of Schutkowski (1993) when tested on other samples. The method created by Loth and Henneberg (2001) seems useful, but it was discarded by Scheuer (2002) and Coqueugnoit et al. (2002). The last methods that have been considered in the previous section are from Molleson and colleagues (1998). They have never been retested and it would be interesting to test them on another sample.

Although non-metric methods seem interesting, there are some problems with the use of them. There is always the chance of inter- and intraobserver error (Pretorius *et al.* 2006, 64). With metric analyses, the inter- and intraobserver errors are much smaller, since the methods are more objective (Dawson *et al.* 2011, 81). Another problem is that some non-metric methods can work very good on one population, but cannot work correctly on another population. The methods of Schutkowski (1993) have been created on an American population and were successfully used on an Amerindian population from Chile (Sutter 2003). It seems that at least the methods of Schutkowski (1993) are applicable to other populations. Another problem is that a local population can show sexually dimorphic changes that can happen within a few decades (Walker 2008, 39). This is problematic when the dates of the burials are more than hundred years apart. Another confounding factor of non-metric methods is the classification of morphological characteristics. Especially words as wide, narrow or intermediate can be interpreted in different ways and this can cause differences in determining sex (Pretorius *et al.* 2006, 64).

It seems that both metric and non-metric methods have their pitfalls. Creating a metric method is difficult for an archaeological sample. In addition, the metric methods created cannot be used on other populations, so there is no reference material. In addition, some researchers even say that non-metric methods are more reliable than metric methods (Baker *et al.* 2005). Therefore, non-metric methods will be further researched within this thesis. There are several reasons for this. First, the non-metric methods are easily applied on another population. Several of the non-metric methods have been tested on other populations and did perform well. This could indicate that the methods can be used on other populations. Until now, nobody has done research on all twelve non-metric methods together. It would be interesting to see whether all the methods are getting the same accuracy rates as the researchers who have presented these methods. In addition, it is possible that combining these methods will give higher accuracy rates. The results of the non-metric methods can also be compared with the results of the other researchers which gives a better understanding of their ability to work on other populations. Therefore, this thesis will be concentrated around researching non-metric methods to estimate the sex of subadults.

1.4 The importance for archaeology

The capacity to determine the sex of subadults is by many osteologists seen as an important factor for social studies and it can increase our ability to estimate age (Baker *et al.* 2005, 48). It is known that females grow and mature faster than males. When osteologists are able to estimate the sex of subadults, it would also be possible to refine our methods to assign age to subadults as well (Perry 2006, 90).

Not only osteologists would profit from the ability to determine the sex of subadults, but archaeologists as well. If we could sex fetuses, we would be able to learn more about changing sex ratios in a cemetery. These changing ratios could tell us more about their predisposing factors like parental investment, economic status and the fragility of male fetuses. Not only the changing sex ratios of fetuses is interesting, but also the changing sex ratios of older subadults. The ratios of degrees of trauma and pathology can tell more about which sex was preferred, the life course of males and females, susceptibility to diseases, specific burial rites for each gender, the division of labour and at what age males and females became accepted as adults in society (Baker *et al.* 2005, 48).

Since the late 1990s it was thought that DNA could resolve the problem of not knowing the sex of subadults. Several studies have focused on the question of infanticide in the Roman period (Fearman *et al.* 1998; Mays and Fearman 2001; Waldron *et al.* 1999) and two studies have incorporated older children (Cappellini *et al.* 2004; Cunha *et al.* 2000). There are however some problems with using ancient DNA to sex subadults. First, there is a problem with the method used. The differential amplification of the X-specific and Y-specific sequences method is often used. However, males can be wrongly identified as female when the X-specific sequence is preferentially amplified over the shorter and more easily degraded Y-chromosome. Caution in the use of this method is advised. Another factor is the choice of bone from where the sample is taken from. Often, DNA is better preserved in the teeth, cranial bones and long bones. These bones should be used when using DNA techniques (Lewis 2011).

Being able to identify the sex of subadults is interesting and important for archaeology. It could help us with better age estimations and give us more information about how children were cared for and treated in the past. Ancient DNA is an option, but it is expensive, the methods should be carefully chosen and the bones need to be well preserved.

1.5 Middenbeemster

For this research a skeletal collection is needed for which the sex of the subadults is known. Not many skeletal collections has this, but for most of the subadults of Middenbeemster the sex is known. During burial the sex was most often registered in the archives that have been preserved. The volunteers of the Historisch Genootschap Beemster and Prof. Dr. M.L.P. Hoogland have been able to match the names within the archives with the skeletons that have been excavated. Because the sex of the subadults within the Middenbeemster collection is known, this research will use the Middenbeemster collection. This collection is housed at the University of Leiden. More information about the collection will be giving in chapter 3.

1.6 Research questions

As seen before knowing the sex of subadults is interesting for archaeology. It could improve our methods to estimate age for subadults and give us more information about how children were cared for and treated in the past. Ancient DNA is an option, but there are various reasons why it is not often employed on subadults. Since the 1870s various methods have been proposed and the most promising methods are non-metric methods. These methods can be further subdivided in non-metric methods for the pelvis, mandible and skull. As noted before, the accuracies of these methods vary much between the original authors and authors who retested these methods.

The main goal of this research is to assess whether the non-metric methods for sexing subadults indeed have low accuracies when retested on another population or that the methods are capable of estimating the sex of subadults. Therefore the main question of this thesis is:

Is it possible to accurately determine the sex of subadults below 21 years with non-metric methods on the subadults from the cemetery of Middenbeemster?

The sub-questions are:

- Which non-metric methods are available from the literature to estimate the sex of subadults?
- What are the accuracy rates of all the methods in the literature?
- What are the accuracy rates of these methods when tested on the known-sex sample from Middenbeemster?
- Is there a higher accuracy rate for a specific age group?
- Is it possible to correctly estimate sex of subadults by using one or a combination of these methods?

Before this research can be done, the non-metric methods need to be identified within the literature. The accuracy rates that have been obtained before in previous research will be given to provide a framework for a better understanding of the accuracy rates for each individual method. These rates will be used as a comparison with the rates obtained in this research. The research itself will be a blind test on the subadults of Middenbeemster. Before the start of the research, the researcher was unaware of the male:female ratio nor any other things that could help to estimate the sex.

Before going further with the research itself, it is necessary to understand more of the research that has been done on this subject before. Within the literature several non-metric methods and their accuracy rates are mentioned. These non-metric methods and their accuracies are described in chapter two. In chapter three the materials that have been used will be named. Also, the history behind the Middenbeemster collection and the research on it are depicted. In addition, the non-metric methods and their descriptions will be given and the statistical analyses that will be used. In chapter four the results with the accuracy rates and the results of the statistical analyses will be given. Chapter five will give an overview of the discussion that stems from the results. Chapter six will be the last chapter and it will give a conclusion and an answer to the main question and the sub-questions.

2. Accuracies of methods to determine the sex of subadults

Various methods have been researched to determine the sex of subadults. These methods are discussed in chapter one. A distinction can be made between metric and non-metric methods. Within this thesis I will be focusing on the non-metric methods. There are several reasons to study the non-metric methods instead of the metric methods. These reasons have been discussed in chapter one.

The non-metric methods can be divided in non-metric methods of the pelvis and the skull. In the following sections the accuracies of these methods will be described. The total accuracy of a method will be described, but the accuracy will also be divided in the accuracy for males and females. This is to see whether the method is only highly accurate for one sex or for both sexes. The method is not interesting when only one of the sexes has a high accuracy and the other sex not. The accuracy will also be further divided in age classes which can tell us if a method is performing the best in a few age groups and whether a method is age-related and not sex-related. Within the tables, the accuracies that are 50% or less are made red to make the accuracies below chance more clear.

Within his research Sutter stated that a method needs to be at least 75% accurate before it can be used within forensic anthropology (2003, 928). Although there is no general consensus among forensic anthropologists or osteologists, the minimum standard within this thesis will be set at 75%.

2.1 Accuracies of non-metric methods of the pelvis

Two researchers can be seen as the inventors of non-metric methods of the pelvis. In 1980 Weaver used a non-metric method of the pelvis to sex fetal and infant skeletons and Schutkowski created in 1993 seven methods to sex subadults skeletons including four methods for the pelvis. All methods were retested by various authors. The accuracies for Weaver's method will be discussed first together with the accuracies of the authors who retested it. After that the same will be done for Schutkowski. Images of these methods will be presented in chapter three.

2.1.1 Weaver

Weaver (1980) was the first who came up with a non-metric method to estimate the sex of subadults. He looked at the sex differences in the ilia of fetal and infants skeletons. A sample of 153 infants between the ages of six fetal months and six months were taken from the collection of the Smithsonian Institution. Of this sample 71 were female and 82 were male (Weaver 1980, 192).

Weaver wrote in his article about determining the sex of subadults through the creation of three indices and one non-metric method, the elevation of the auricular surface. The indices did not perform well, but the non-metric method did. He recommended this method to be used to determine the sex of subadults. When the auricular surface is elevated, the skeleton is scored as female (Weaver 1980, 192-195). According to Weaver, the method has an accuracy of 72.5%. However, the males were estimated correctly more often than the females (Weaver 1980, 195).

This method was retested by Hunt (1990), Mittler and Sheridan (1992) and Sutter (2003). Hunt (1990) retested this method on a sample of 275 ilia from skeletons with ages between six fetal months and 6.5 years old. The skeletons were taken from an American Indian population. The ages are unknown, so Hunt (1990) looked at the prevalence rate of the raised method. Within the newborns the raised method was 5.6 times more prevalent than the non-raised method. From the age of 2.5 years the non-raised method was 3.5 to 4 times more prevalent than the raised method. These results gave Hunt the conclusion that the method is an indicator of age and not of sex (Hunt 1990).

Mittler and Sheridan (1992) also retested the method on 54 ilia of known-age and sex from Nubia. The ages of these skeletons ranged between newborn to eighteen years. The overall accuracy was 74.1%. The method appeared to perform better on males, with an accuracy of 85.3%. The performance on females was much lower with an accuracy rate of 58.3% (table 2; next page). When the age groups were studied more closely it became apparent that in the age group of birth to nine years the method performed better for males (70.6% accuracy) than for females (33.3% accuracy). In the age group of ten to eighteen, the method worked better overall and the accuracy rates for both females and males increased. Females had an accuracy of 66.7% and males even an accuracy of 100%. It seems that the older the children, the better this method works. The authors think that the method is not useful for archaeologists, but that it can be used in forensic applications. When the child is over nine and has an elevated auricular surface, there is a 99% probability that that child is a female. A non-elevated pattern cannot be used to sex children (Mittler and Sheridan 1992).

Table 2. Table with the various accuracies for the elevation of the auricular surface (based on Weaver 1980, 195; Mittler and Sheridan 1992, 1073 and Sutter 2003, 930).

Auricular surface	Female	Male	Combined
Weaver (1980)	57.7%	58.3%	60.9%
Mittler and Sheridan (1992)	85.4%	85.3%	77.8%
Sutter (2003)	73.5%	74.1%	71.7%

The most recent research with this method was conducted by Sutter (2003). His research was conducted on 85 known-sex mummies from Chile between the ages of birth till fifteen years. The combined accuracy was 71.7% (Sutter 2003, 931) (table 2; above).

Sutter also had the problem that the method scored males much better than the females. When dividing the age in various groups it is visible that again the method scores the younger groups much lower than the older age groups (table 3; next page). The age group newborn to one year has an accuracy of 53.8% which is slightly better than chance. Within this age group, (42.8%). The females score much better for the second age group (2 - 5 years) with an accuracy of 80%. Males score an accuracy of 100%. The method also scores this for the third age group for males. The overall accuracy rates is high with 91.7%. After this age group, the method performs much less with each age group. The age group six to ten years has a combined accuracy rate of 81.8%. The method works well on males (100%), but performs worse on females (50%). Strangely, the method works good on females in the age group eleven to fifteen years (71.4%), but males

Table 3. The accuracies obtained by Sutter for the elevation of the auricular surface divided over four age groups (based on Sutter 2003, 930).

Auricular surface	Female	Male	Combined
NB-1 year	42.8%	67%	53.8%
2-5 years	80%	100%	91.7%
6-10 years	50%	100%	81.8%
11-15 years	71.4%	58.8%	62.5%

perform worse (58.8%). There is not a real trend visible within these accuracy rates and the age groups. The age group six to ten years has a combined accuracy of 81.8%, but females had a low accuracy of 50%. The last age group is eleven to fifteen years and this group has an combined accuracy of 62.5%, but the method performed very good on females who had an accuracy of 71.4%. The method performed worse on the males who had an accuracy of 58.8% (Sutter 2003, 930).

The accuracies differ much from the accuracies Mittler and Sheridan (1992) obtained in their sample. Mittler and Sheridan had an accuracy of 33.3% for the method when used on females of the age group newborn to nine years (Mittler and Sheridan 1992, 1073) while Sutter had an accuracy of 44.4% for the same group (Sutter 2003, 930). Sutter had a little higher accuracy than Mittler and Sheridan, but this is not much and indicates that the method is not really reliable for females till the age of ten. Both Mittler and Sheridan and Sutter obtained higher accuracies for the method when tested on the males. They are respectively 70.6% and 90% (Mittler and Sheridan 1992, 1073; Sutter 2003, 930). For males the method is much more reliable.

When looking at the older age groups, the accuracies are much better. The method performed better on the females when Mittler and Sheridan tested it on females over ten years (Mittler and Sheridan 1992, 1073) while Sutter had an accuracy of 71.4% (Sutter 2003, 930). Both accuracies are much higher and could indicate that the method is more reliable with older individuals. For males the accuracies were different namely 100% for Mittler and Sheridan (1992, 1073) and 58.8% for Sutter (2003, 930). The discrepancies within the male accuracies between Mittler and Sheridan and Sutter might be explained by the fact that they used samples from different populations. It is known that there is a difference in sexual dimorphism between populations (Maclaughlin and Bruce 1986).

The non-metric method of Weaver does seem interesting. Three researchers retested the method. Hunt (1990) could not replicate the results of Weaver. Mittler and Sheridan (1992) and Sutter (2003) did have combined accuracies that were matching with the results of Weaver. The results indicate that the method is more accurate on older individuals and that the accuracy is higher for males (Mittler and Sheridan 1992; Sutter 2003). Mittler and Sheridan state in their article that the high accuracy among males stems from a nonelevated pattern among young-aged and that this pattern is persistent within one-third of the older females (Mittler and Sheridan 1992, 1074). Sutter (2003) also says that this method should not be used to determine the sex of subadults (Sutter 2003, 934). Both the accuracy rates of Weaver (1980) and Sutter (2003) are below the standard of 75%. The accuracy rate of Mittler and Sheridan (1992) was above 75%. There seems to be no real consensus whether the elevation of the auricular surface can be used to determine the sex of subadults or not.

2.1.2 Schutkowski

In 1993 Schutkowski wrote an article about estimating the sex of subadults by using seven different non-metric methods. Four of these methods are used on the pelvis. These four methods are the angle and depth of the greater sciatic notch, the arch criterion and the curvature of the iliac crest. These four methods were tested on 61 known-sex skeletons from the Spitalfields collection, London. Five of these skeletons were between six and eleven years and the others were between newborn and five. These five skeletons were omitted from the results, because there are too few of them (Schutkowski 1993, 200)

First, the accuracy rates for the angle of the greater sciatic notch will be discussed together with the accuracy rates from other authors. After this, the depth of the greater sciatic notch, the arch criterion and the iliac crest will be discussed with the accuracy rates from other authors.

2.1.2.1 Angle of the greater sciatic notch

This method got a combined accuracy rate of 81.6% with Schutkowski (1993) (table 4, below). Authors who retested this method got accuracy rates that were less. Sutter retested the method on 85 known-sex mummies from Chile. This method had a combined accuracy rate of 74.1% (2003, 931). The accuracy rate is high enough to say that this method can be used to determine sex of subadults. In addition, Vlak and colleagues (2008) tested this method as well. They used a sample of 56 known-sex and age subadults from the Bocage Museum in Lisbon. Their accuracy rates are low with a combined rate of 53.6%. The angle of the greater sciatic notch does not perform well on either males or females. Both Schutkowski and Sutter had better rates for both males and females (table 4). Only Schutkowski (1993) had a higher accuracy rate than the minimum of 75%

Table 4. Table with the accuracy rates for the greater sciatic notch angle based on (Schutkowski 1993, 203; Sutter 2003, 930 and Vlak et al. 2008, 311).

Greater sciatic notch angle	Female	Male	Combined
Schutkowski (1993)	95.2%	71.4%	81.6%
Sutter (2003)	78.6%	69.2%	74.1%
Vlak *et al.* (2008)	52.5%	54.5%	53.6%

The accuracy rates can also be divided in age groups. This can be seen in table 5 (next page) where the rates of Schutkowski (1993), Sutter (2003) and Vlak *et al.*(2008) are presented. All three authors do not have corresponding accuracy rates for all age groups. Only the fourth age group (11 to 15 years) seems to be corresponding while the other three groups are varied among the authors. The first age group differs among all three authors. Schutkowski (1993) and Vlak *et al.* (2008) have the same accuracy rates for the females, but not for the males. At the same time, Schutkowski (1993) and Sutter (2003) have the same rates for males, but not for the females or the combined rate. This difference might be explained by a difference in sexual dimorphism between the populations. This is known for other adults populations as well (Maclaughlin and Bruce 1986). While Schutkowski (1993) and Sutter (2003) had corresponding accuracy rates within

Table 5. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311) for the greater sciatic notch angle divided over four age groups.

Greater sciatic notch angle		**Female**	**Male**	**Combined**
NB - 1 year	Schutkowski (1993)	100%	55%	77.3%
	Sutter (2003)	71.4%	50%	61.5%
	Vlak *et al.* (2008)	100%	78%	87.5%
2-5 years	Schutkowski (1993)	90%	82%	85.2%
	Sutter (2003)	85.7%	86%	85.7%
	Vlak *et al.* (2008)	37.5%	40%	38.9%
6-10 years	Sutter (2003)	25%	100%	78.6%
	Vlak *et al.* (2008)	20%	71%	50%
11-15 years	Sutter (2003)	71.4%	94%	86.9%
	Vlak *et al.* (2008)	33.3%	100%	80%

the second age group, the rates of Vlak and colleagues failed. Their rates for all three factors are below chance. This can again be caused by a difference in sexual dimorphism between the populations. Within the third age group, the method has low accuracy rates for the females, but high rates for the males. The last age group is again strangely divided. Sutter has high accuracy rates for both sexes, but the method does not perform well with Vlak and colleagues for the females. Again, a difference in sexual dimorphism between the populations or anomalous individuals might have caused this difference between the authors. Vlak and colleagues state that the difference in accuracy rates is caused by the fact that the method is age-related and not sex-related (2008, 311). This is indeed a possibility.

2.1.2.2 Depth of the greater sciatic notch

This method was created by Schutkowski in 1993 and retested later by Sutter (2003) and Vlak and colleagues (2008). Sutter retested the method on 85 known-sex mummies from Chile. His accuracy rates were different from Schutkowski (1993)(table 6; next page). In addition, Vlak and colleagues (2008) retested the methods as well and their accuracy rates differ much from both authors. Both Sutter and Schutkowski had high combined accuracy rates for this method. The method performed less on the females with Schutkowski, but for

Table 6. The various accuracy rates for the depth of the greater sciatic notch based on Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311).

Greater sciatic notch depth	**Female**	**Male**	**Combined**
Schutkowski (1993)	68.4%	86.7%	79.6%
Sutter (2003)	92.9%	69.2%	81.5%
Vlak *et al.* (2008)	43.5%	69.7%	58.9%

Sutter it was the other way around. The method performed less on the males. Still, the accuracies are pretty high around 68-69%. Vlak and colleagues were less able to reproduce the results of the previous authors. Their accuracy rates are clearly below the other two authors. This can be caused by a difference in sexual dimorphism between the populations which is known for adult populations (Maclaughlin and Bruce 1986).

In table 7 (below) the accuracy rates are divided in four age groups for all three authors. Several patterns can be seen within this table. The first pattern is that the method underperformed for males within the first age group. All three authors have this same pattern. This makes clear that the depth of the greater

Table 7. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311) for the greater sciatic notch depth divided over four age groups.

Greater sciatic notch depth		**Female**	**Male**	**Combined**
NB - 1 year	Schutkowski (1993)	80%	40%	60%
	Sutter (2003)	85.7%	0%	46.1%
	Vlak *et al.* (2008)	100%	44%	68.7%
2-5 years	Schutkowski (1993)	83.8%	64%	73.1%
	Sutter (2003)	100%	71%	85.7%
	Vlak *et al.* (2008)	25%	50%	38.9%
6-10 years	Sutter (2003)	33.3%	80%	69.2%
	Vlak *et al.* (2008)	20%	100%	66.7%
11-15 years	Sutter (2003)	28.6%	88%	69.6%
	Vlak *et al.* (2008)	0%	100%	70%

sciatic notch cannot be used on children aged below one year. Another pattern is that the method did not perform well on males in the first two age groups when Vlak and colleagues (2008) tested. The accuracy rates for the males improve in the next two age groups. The method performs different for females. They have a high accuracy rate for the first age group and their accuracy rate drops after that. Vlak and colleagues state that this is caused by the fact that the method is age-related and not sex-related (2008, 311). Their research also indicate that the method cannot be used in any of the age groups. The results of Sutter (2003) indicate that the method can be used on the second age group, but not on any of the other age groups.

2.1.2.3 The arch criterion

The arch criterion was created by Schutkowski (1993) on 56 known-sex skeletons from birth to five years old. It was retested by Sutter (2003) and by Cardoso and Saunders (2008). Sutter (2003) retested the method on 85 known-sex skeletons from Chile between the ages of newborn and fifteen years. Cardoso and Saunders (2008) retested the iliac crest on 97 known-sex skeletons from the Bocage Museum in Lisbon. Their ages ranged between newborn and fifteen years. Cardoso and Saunders (2008) did not only retest the method itself, but also the inter- and intraobserver error. Within table 8 (below) the accuracy rates for all the authors are visible. In addition, the different rates that the different observers got during the research of Cardoso and Saunders (2008) is also displayed.

Table 8. The various accuracy rates for the arch criterion based on Schutkowski (1993, 203), Sutter (2003, 930) and Cardoso and Saunders (2008,27).

Arch criterion	Female	Male	Combined
Schutkowski (1993)	60%	81.5%	72.3%
Sutter (2003)	85.7%	76.9%	81.5%
Cardoso and Saunders (2008) 1	40%	50.9%	46.4%
Cardoso and Saunders (2008) 2	50%	31.6%	39.2%
Cardoso and Saunders (2008) 3	60%	26.3%	40.2%

Both Schutkowski (1993) and Sutter (2003) (table 8, previous page) have a high performance for the arch criterion. Sutter (2003) even improved the accuracy rates for this method. However, when Cardoso and Saunders (2008) tested this method, the accuracy rates dropped. All the observers have low accuracy rates and there is little correspondence between the observers. Observers one and three both had only one time that the method performed better than chance. All the other accuracy rates are below chance. The authors state that the arch criterion should not be used to sex subadults (Cardoso and Saunders 2008, 27).

The accuracy rates for the arch criterion are divided in four age groups in table 9, below. There is not one age group that is interesting for archaeology or

Table 9. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Cardoso and Saunders (2008, 27) for the arch criterion divided over four age groups.

Arch criterion		**Female**	**Male**	**Combined**
NB - 1 year	Schutkowski (1993)	70%	82%	76.2%
	Sutter (2003)	71.4%	50%	61.5%
	Observer 1	61.5%	34.8%	44.4%
	Observer 2	61.5%	21.7%	36.1%
	Observer 3	61.5%	21.7%	36.1%
2 - 5 years	Schutkowski (1993)	50%	81%	69.2%
	Sutter (2003)	100%	100%	100%
	Observer 1	40%	43.8%	42.3%
	Observer 2	70%	25%	42.3%
	Observer 3	60%	37.5%	46.2%
6 - 10 years	Sutter (2003)	25%	100%	78.6%
	Observer 1	36.4%	77.8%	55%
	Observer 2	45.5%	33.3%	40%
	Observer 3	63.6%	22.2%	45%
11 - 15 years	Sutter (2003)	25%	100%	84.2%
	Observer 1	0%	77.8%	46.7%
	Observer 2	0%	66.7%	40%
	Observer 3	50%	22.2%	33.3%

forensics. In all cases the method performs well on either females or males except for the first age group (newborn to one year) tested by Schutkowski (1993) and the second age group (two to five years) tested by Sutter (2003). All three observers of Cardoso and Saunders (2008) have differing rates in all the age groups and this only emphasizes that there is an inter- and intraobserver error. Only within the first age group, for the females, do the observers correspond with each other, but it is not clear whether they corresponded on the skeletons or only in rates (table 9; previous page).

2.1.2.4 The iliac crest

The method iliac crest was created by Schutkowski in 1993 on the Spitalfields collection. He used 56 known-sex skeletons from the ages of newborn to five years. This method was retested by Sutter in 2003 on 85 known-sex skeletons from the ages of newborn to fifteen years. Their accuracy rates are visible in table 10 (below). The combined accuracy rate is low for both authors which indicates that this method is not interesting at all. The accuracy rates for females are higher than for males. If only females are determined correctly, it is not useful for archaeologists. It seems that the arch criterion cannot be used to determine the sex of subadults.

Table 10. The various accuracy rates for the iliac crest based on Schutkowski (1993, 203) and Sutter (2003, 930).

Iliac crest	Female	Male	Combined
Schutkowski (1993)	85.7%	54.2%	68.9%
Sutter (2003)	92.9%	38.5%	66.6%

The accuracy rates for the iliac crest method can be further divided in four age groups (table 11; next page). The method performs well in the second age group. Both Schutkowski (1993) and Sutter (2003) had good rates. This method did not work well on the other age groups. Both Schutkowski (1993) and Sutter (2003) had low accuracy rates when the iliac crest was used on males within the first age group. Females had much higher accuracy rates. The last two age groups have low accuracy rates for the females, and high accuracy rates for the males. Still, the iliac crest cannot be used within archaeology or forensics.

Table 11. The accuracies obtained by Schutkowski (1993, 203) and Sutter (2003, 930) for the iliac crest divided over four age groups.

Iliac crest		Female	Male	Combined
NB - 1 year	Schutkowski (1993)	80%	40%	60%
	Sutter (2003)	85.7%	0%	46.1%
2 - 5 years	Schutkowski (1993)	83.8%	64%	73.1%
	Sutter (2003)	100%	71%	85.7%
6 - 10 years	Sutter (2003)	33.3%	80%	69.2%
11 - 15 years	Sutter (2003)	28.6%	88%	69.6%

2.1.2.5 Summary

The four pelvic methods of Schutkowski (1993) are debated. Sutter (2003) has shown that three of the four methods could be interesting for estimating the sex of subadults. However, when the sample is divided in four different age groups and also split by sex, a different picture comes to light. Both the angle and depth of the greater sciatic notch are more likely age-related than sex-related. This is made clear in the research of Vlak and colleagues (2008, 311). The research of Schutkowski and Sutter does seem to indicate more that the methods are sex-related, otherwise they would not have gotten such high accuracy rates. It is interesting to see whether this research indicates that the angle and depth of the greater sciatic notch are age-related or sex-related. The age group two to five years performs the best in the angle of the greater sciatic notch, the depth of the greater sciatic notch and the iliac crest. For all the other age groups, the method either determines females or males incorrectly. Additional research by Cardoso and Saunders (2008) has also revealed that there is a high intra- and interobserver error for the arch criterion.

2.2 Accuracies of non-metric methods of the skull

After the pelvis, the skull is seen as a good indicator of sex of subadults. Within this thesis the mandible is grouped with the skull. The first methods for the mandible were created by Schutkowski (1993) while he also created the methods for the pelvis. Other researchers who created methods for the skull are Loth and Henneberg (2001) and Molleson and colleagues (1998).

First, the methods and accuracies of Schutkowski will be discussed and also the authors who retested this method. Second, the method and accuracy of Loth and Henneberg will be discussed and the authors who retested this method and lastly the methods of Molleson and colleagues will be discussed.

2.2.1 Schutkowski

The first research carried out on sexual dimorphism of the mandible in subadults was by Schutkowski (1993). He found three different methods. These methods are the protrusion of the chin, the shape of the anterior dental arcade and the eversion of the gonion region. As mentioned earlier, Schutkowski used a sample of 61 known-sex skeletons between newborn and eleven years (1993). The five skeletons between the ages of six and eleven were excluded from the sample, because there were too few skeletons (Schutkowski 1993, 200)

First, the protrusion of the chin will be discussed which will be followed by the shape of the anterior dental arcade and the eversion of the gonion region. For all methods, the accuracy rates will be given and the accuracy rates will be further divided in four age groups. If the method has been retested by another author(s), than these accuracy rates will be presented as well.

2.2.1.1 Protrusion of the chin

This method was created by Schutkowski in 1993 on 56 known-sex skeletons between the ages of newborn and five years. It was retested in 2003 by Sutter. He did his research on 85 known-sex skeletons from Chile between the ages of newborn and fifteen years. Their accuracy rates are visible in table 12 on the next page. All the rates are above chance (50%) which is interesting. However, Schutkowski (1993) gets such a high accuracy, because the method works really well on the females. Only Sutter (2003) meets the criterion of at least 75% accuracy for this method. For Sutter (2003) it is the other way around. This could indicate that there is a difference in the sexual dimorphism between the two populations which is known for other adult populations (Maclaughlin and Bruce 1986).

Table 12. The various accuracy rates for the protrusion of the chin based on Schutkowski (1993, 202) and Sutter (2003, 931).

Protrusion of the chin	Female	Male	Combined
Schutkowski (1993)	92.3%	59.3%	70%
Sutter (2003)	64.3%	93.9%	83.1%

The accuracy rates for the protrusion of the chin are further divided in four age groups within table 13 (below). The combined accuracy rates and the individual accuracy rates do seem interesting. Both the first and second age group can be used for estimating the sex of subadults. Especially when looking at the results from Sutter (2003). The protrusion of the chin performed very good on both males and females for Sutter (2003), but only for the first two age groups. The other two age groups have low accuracy rates for females and good accuracy rates for males. It is possible that this is caused by the fact that the method was developed on children between newborn and five years and that it cannot be used on children older than five.

Table 13. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the protrusion of the chin divided over four age groups.

Protrusion of the chin		Female	Male	Combined
NB - 1 year	Schutkowski (1993)	100%	61.5%	73.7%
	Sutter (2003)	77.8%	100%	92%
2 - 5 years	Schutkowski (1993)	85.7%	57.1%	67.7%
	Sutter (2003)	71.4%	100%	89.5%
6 - 10 years	Sutter (2003)	40%	100%	72.7%
11 - 15 years	Sutter (2003)	57.1%	80%	72.7%

2.2.1.2 Shape of the anterior dental arcade

This method was created by Schutkowski in 1993 on 56 known-sex skeletons from the ages of newborn to five years. This method was retested by Sutter (2003) on 85 known-sex skeletons from Chile between the ages of newborn and fifteen years. This resulted in the accuracy rates presented in table 14 (below). The combined accuracy rates of both authors is interesting. It is not really high, but it can be used within archaeology. The method performed on females and male pretty good within the research of Schutkowski (1993). Sutter had more problems with correctly identifying the females within his sample. This can be caused by a difference in sexual dimorphism between these populations as is known for adult populations (Maclaughlin and Bruce 1986). Only Sutter (2003) obtained an accuracy rate that is above the criterion of 75%.

Table 14. The various accuracy rates for the shape of the anterior dental arcade based on Schutkowski (1993, 202) and Sutter (2003, 931).

Anterior dental arcade	Female	Male	Combined
Schutkowski (1993)	69.2%	73.1%	71.8%
Sutter (2003)	53.6%	89.9%	76.6%

Within table 15 (below) the accuracy rates are further divided in four age groups and according to sex. The anterior dental arcade performs bad on females in the first age category. Schutkowski (1993) has an accuracy of 50% and Sutter (2003) performs worse with an accuracy of 11.1%. This indicate that the anterior dental arcade cannot be used on subadults below one year. Both Sutter and Schutkowski state that the method should not be used on females below the age of one year (1993, 202; 2003, 934). However, the sex of the subadults is unknown, which means that when the method performs well on males, the method still

Table 15. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the anterior dental arcade divided over four age groups.

Anterior dental arcade		Female	Male	Combined
NB - 1 year	Schutkowski (1993)	50%	67.7%	61.1%
	Sutter (2003)	11.1%	100%	68%
2 - 5 years	Schutkowski (1993)	85.7%	78.6%	81%
	Sutter (2003)	57.1%	91.7%	78.9%
6 - 10 years	Sutter (2003)	80%	83.3%	81.8%
11 - 15 years	Sutter (2003)	85.7%	80%	81.8%

cannot be used to determine the sex of subadults. The other three age groups perform much better. Both Schutkowski (1993) and Sutter (2003) have high accuracy rates within these groups and this indicates that this method can be used for older individuals. Especially the two older age groups perform well within the research of Sutter (2003). The accuracy rates are around 80% which indicates that the method can be used for archaeology and forensics. Sutter even recommends the use of the anterior dental arcade for subadults over the age of six (2003, 934).

2.2.1.3 Eversion of the gonion region

The eversion of the gonion region was created by Schutkowski (1993) on 56 known-sex skeletons between the ages of newborn and five years. Sutter retested this method in 2003 on 85 known-sex skeletons between newborn and fifteen years. They accuracy rates that both authors obtained are visible in table 16. Within this table it is visible that the method improved when Sutter (2003) used it on his skeletons. The method still did not score really great on the females, but the accuracy rate for males was much higher. The combined rate was also much higher. This rate is above the criterion of 75% which indicates that the eversion of the gonion region can be used by archaeologists and forensic anthropologists.

Table 16. The various accuracy rates for the eversion of the gonion region based on Schutkowski (1993, 202) and Sutter (2003, 931).

Eversion of the gonion region	Female	Male	Combined
Schutkowski (1993)	60%	68%	65%
Sutter (2003)	63.4%	92.1%	81.7%

The method can be further divided in four age groups who are visible in table 17 on the next page. This table shows that the method can be used really well on the age group two to five years. Both Schutkowski (1993) and Sutter (2003) have good accuracy rates within this group. The first age group can be correctly used according to Sutter (2003), but it performed less with Schutkowski. The method did not sex males very good. It was barely above chance (54.6%). Within the last two age groups, the method performs not well on females. The accuracy rates are only 25%. It is possible that the older age groups perform worse, because the method was created on skeletons between the age of newborn and five years.

Table 17. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the eversion of the gonion region divided over four age groups.

Eversion of the gonion region		Female	Male	Combined
NB - 1 year	Schutkowski (1993)	100%	54.6%	77.3%
	Sutter (2003)	71.4%	80%	61.5%
2 - 5 years	Schutkowski (1993)	90%	82.3%	85.2%
	Sutter (2003)	100%	100%	100%
6 - 10 years	Sutter (2003)	25%	100%	78.6%
11 - 15 years	Sutter (2003)	25%	100%	84.2%

2.2.1.4 Summary

There is not one age group that performs the best within the three methods. For the protrusion of the chin the first two age groups are interesting. For the anterior dental arcade and the eversion of the gonion region it seems that these methods are interesting for the second age group. The last two age groups perform the best with the anterior dental arcade. Both the protrusion of the chin and the eversion of the gonion region indicate that the last two age groups are not usable for archaeology or forensics. In each instance, the method performs worse on either males or females. This indicates that the method can be used for one sex, but not for the other. This is meaningless for archaeology, since we do not know the sex before we start the determination of the sex. It is possible that this is caused by the fact that the methods were created on skeletons between the ages of newborn and five years and this can create a problem with older individuals.

2.2.2 Loth and Henneberg

Because the methods of Schutkowski did not prove to be accurate, Loth and Henneberg (2001) created their own method. Their research was conducted on the mandibles of 62 known-sex skeletons between the ages of newborn and nineteen years from the Dart Collection at the University of Witwatersrand. From these 62 mandibles, 19 were chosen to form the basis of this method. The other 43 mandibles were not seen as healthy or complete enough to be used to create a new method. The mandibles used were from skeletons between the ages of seven months and 3.5 years. The tests were conducted by both authors and two other experienced osteologists. The overall accuracy was 81% (table 18, below). Although, males had a higher accuracy (95.2%) than females (75%). Additionally, this method was tested on two South African forensic cases and nine CT scans from living French children. These cases had an accuracy of 82% (Loth and Henneberg 2001, 181-185). These accuracy rates are above the 75% criterion set for the present study.

The authors recommend that this method should not be used on subadults older than six years. The reasons for that are that their sample was not older than six years and the transition to adult shapes begins around the sixth year. This transition could cause errors in sexing mandibles that are over six years (Loth and Henneberg 2001, 185).

The complete mandible was retested by Coqueugniot and colleagues (2002) and by Scheuer (2002) (table 18; below). Coqueugniot and colleagues (2002) tested the method on 76 known-sex skeletons between the ages of newborn and eighteen years from the Torino collection in Turin, Italy. They also used mandibles of children older than six years even though Loth and Henneberg had stated that this could cause more errors (2001, 185). The overall accuracy is only 52.6%. Females and males had different accuracy rates for this method. Females have a high accuracy rate of 62.5% and males only 41.7%. The accuracy rates are visible in table 18, below.

Table 18. The various accuracy rates for the complete mandible based on Loth and Henneberg (2001, 183), Scheuer (2002, 190) and Coqueugniot et al. (2002, 136).

Complete mandible	Female	Male	Combined
Loth and Henneberg (2001)	75%	95.2%	81%
Scheuer (2002) attempt 1	66.7%	77.8%	75%
Scheuer (2002) attempt 2	44.4%	85.2%	75%
Coqueugniot *et al.* (2002)	62.5%	41.7%	52.6%

Scheuer (2002) tested the method on 36 known-sex mandibles from the Spitalfields collection in London. The ages ranged between five months and five years. The mandibles were tested twice by the author with two days apart. In both cases the accuracy was 75%, but there are large differences when the method is tested on males and females separate (table 18; previous page). In both instances the method performs less on females than on males.

When the results of both sessions are viewed the accuracy falls down to 63.9% when taking only the mandibles that were correctly estimated in both attempts (table 19; below). For females the accuracy is extremely low namely 44.4% and for males it is higher with 70.4%. This is visible in table 19 below. These differences show that there is a case of intraobserver error. The accuracy rates are low for both attempts and the author does not succeed in replicating the results of the first attempt.

Table 19. The accuracy rates for the complete mandible and the allocation in both sessions (Scheuer 2002, 190).

Complete mandible	Female	Male	Combined
Correct both sessions	44.4%	70.4%	63.9%
Incorrect both sessions	33.3%	7.4%	13.9%
Inconsistent both sessions	22.2%	22.2%	22.2%

The accuracy rates for the complete mandible can be further divided in two age groups. These groups are visible in table 20 on the next page. In both cases, the method did not perform well on females. Within the first age group, Loth and Henneberg (2001) have a 60% rate for females and Scheuer (2002) has a rate of 0%. It seems that the method is not interesting for females below one years and therefore this method cannot be used on subadults below one year. In addition, the method does not perform much better on the females in the second age group. The accuracy is there around 60% which is below the criterion of 75%. This method seems interesting, because of the combined accuracy rates, but when looking at females and males separately, it becomes clear that the complete mandible method cannot be used. Even the combined accuracy rates are not near 75% which means that the complete mandible method cannot be used for archaeology or forensics.

Table 20. The accuracy rates for the complete mandible divided in two age groups (Loth and Henneberg 2001, 183; Scheuer 2002, 190).

Complete mandible		**Female**	**Male**	**Combined**
NB - 1 year	Loth and Henneberg (2001)	60%	100%	71.4%
	Scheuer (2002)	0%	78%	70%
2 - 5 years	Loth and Henneberg (2001)	57%	100%	72.7%
	Scheuer (2002)	60%	69%	62.5%

Coqueugniot and colleagues (2002) used subadults above six years although Loth and Henneberg stated that those should not be used, because the method would become useless after that age (2001, 185). They have divided their accuracy rates in subadults younger than six and subadults over six. Their accuracy rates are visible in table 21, below. Strangely, their accuracy rate rises after the age of six. Before the age of six the accuracy rate is 43.5%, but over the age of six the accuracy rate is 56.6%. Another trend is also visible. Before the age of six the method works better on males (58.3%) than on females (27.3%). Over the age of six the method works better on females (75.9%) than on males (33.3%). This could indicate that the method is more age-related than sex-related.

Table 21. The accuracy rates of Coqueugniot et al. (2002) for the complete mandible divided in two age groups.

Complete mandible	**Female**	**Male**	**Combined**
< 6 years	27.3%	58.3%	43.5%
> 6 years	75.9%	33.3%	56.6%

2.2.3 Molleson and colleagues

Molleson and colleagues used three adult methods for the skull on subadults. These methods are the mandibular angle, the mentum and the orbital morphology. All three methods were tested on a sample of 53 known-sex adults and 20 known-sex subadults from the Spitalfield collection in London. In addition, the methods were tested on a unknown-sex sample of adults and subadults from Wharram Percy. All methods were scored with four scores. A -2 and a -1 was giving to a female appearance and a +1 and +2 were given to male appearances. These scores were added up to created a Combined Facial Score (CFS). First, the mandibular angle, the mentum and the orbital morphoology will be discussed. After this the outcomes of the CFS scores will be discussed.

2.2.3.1 The mandibular angle

The mandibular angle was tested on both known-sex adults and subadults from the Spitalfields collection in London. The ages of the subadults ranged between newborn and fourteen years. The accuracy rates are visible in table 22. The accuracy rates for the adults are high and very good. The accuracy rates for the subadults are less interesting. The combined accuracy rate does not meet the criterion of 75%. In addition, the mandibular angle does not work well on males and therefore, this method cannot be used within archaeology or forensics.

Table 22. The accuracy rates for the mandibular angle for the Spitalfields sample(Molleson et al. 1998, 724).

Mandibular angle	Female	Male	Combined
Adults	88.6%	100%	92.4%
Subadults	78.6%	50%	70%

This method was also tested on the Wharram Percy adults and subadults. The sex of this sample is unknown, so checking the reliability of this method is more difficult. A total of 28 adults and 57 subadults were chosen. For the adults the sex was estimated with pelvic criteria. The accuracy rates are visible in table 23. The combined accuracy rate is high and the method performs good on both males and females. The picture is quite the same as for the Spitalfields sample.

Table 23. The accuracy rates for the mandibular angle for the Wharram Percy sample (Molleson et al. 1998, 724).

Mandibular angle	Female	Male	Combined
Adults	84.6%	92.8%	88.9%

A total of 57 subadults between five and seventeen years were taken from the Wharram Percy collection to be used within this research. For the subadults it was impossible to use pelvic criteria to estimate the sex. Therefore the lower permanent canine was measured. This tooth is known to be sexually dimorphic. It was hypothesized that the skulls with a male mandibular angle would have large canines. The skeletons with a male mandibular angle did indeed have larger canines and so giving evidence that this method does indicate sexual dimorphism (Molleson *et al.* 1998, 724-727). However, no accuracy rate for the subadults could be created.

2.2.3.2 Mentum

The mentum was tested on 53 known-sex adults and 20 known-sex subadults from the Spitalfields collection. For the subadults the ages ranged between newborn and fourteen years. In table 24 are the accuracy rates for the mentum visible for the adults and subadults. The adults score good with this method. The method works the best for males (100%) and although the females have a lower accuracy rate, the combined accuracy rate still indicates that this method can be used within archaeology and forensics. The accuracy rate of 85.3% for females is aboven the criterion of 75%. It is not interesting to use the mentum on subadults. The method performs less on males (16.7%) and not really great on females (76.9%). The combined rate is barely above chance (57.9%).

Table 24. The accuracy rates for the mentum for the Spitalfields sample (Molleson et al. 1998, 724).

Mentum	Female	Male	Combined
Adults	85.3%	100%	90.6%
Subadults	76.9%	16.7%	57.9%

The method was also tested on the unknown-sex adults and subadults from Wharram Percy. A total of 28 adults and 57 subadults were used. The subadults ranged in age between five to seventeen years. For the adults pelvic criteria were used to estimate sex. This resulted in the accuracy rates visible in table 25. The combined accuracy rate is a bit lower than for the Spitalfields sample, but still high enough for archaeology and forensics. The method is not really good for females, but it can still be used.

Table 25. The accuracy rates for the mentum for the Wharram Percy sample (Molleson et al. 1998, 724).

Mentum	Female	Male	Combined
Adults	69.2%	92.8%	81.5%

For the subadults it was not possible to use pelvic criteria to estimate the sex. Instead, the sexual dimorphism of the lower permanent canine was used. The lower permanent canine was measured within each skeletons and it was hypothesized that the skull with a male mentum would have larger canines. This turned out to be true and this indicates that the mentum is sexually dimorphed (Molleson *et al.* 1998, 724-727). However, no accuracy rate for the subadults could be created.

2.2.3.3 Orbital morphology

The orbital morphology was tested on the same sample from Spitalfields as the mandibular angle and the mentum. The accuracy rates are visible in table 26. The accuracy rates for the adults are very good. The orbital morphology performed very well for both females and males. For the subadults, only three skulls were available. All the other skulls were too damaged to be used. This resulted in an accuracy rate of 0% for the females and an accuracy rate of 100% for the males. These results should be treated with caution.

Table 26. The accuracy rates for the orbital morphology for the Spitalfields sample (Molleson et al. 1998, 724).

Orbital morphology	Female	Male	Combined
Adults	91.2%	94.7%	92.4%
Subadults	0%	100%	66.7%

The orbital morphology was also tested on the unknown-sex adults and subadults from the Wharram Percy collection. The accuracy rates for the adults are visible in table 27. The sex was determined through pelvic criteria. The accuracy rates are all above 75% which indicates that this method can be used to determine the sex of adults.

Table 27. The accuracy rates for the orbital morphology for the Wharram Percy sample (Molleson et al. 1998, 724).

Orbital morphology	Female	Male	Combined
Adults	100%	85.7%	91.7%

There were 57 subadults ranging in age between five to seventeen years. The researchers could not use pelvic criteria to determine the sex of subadults. Therefore, they used the measurements of the lower permanent canine. It is known that this tooth is sexually dimorphic and it was hypothesized that male orbital morphology would have larger canines. This turned out to be true and this might be an indication that this method indicates sexual dimorphism (Molleson *et al.* 1998, 724-727). However, no accuracy rates could be obtained.

2.2.3.4 CFS scores

All three methods have been scored with a -2 and -1 for females and a +1 and +2 for males. This was done to create a Combined Facial Score (CFS). This was done for both the subadults and adults from the Spitalfields collection and the Wharram Percy collection. The CFS accuracy rates are visible in table 28. The CFS of the adults from Spitalfields and Wharram Percy are very good. It seems that by combining the three methods, the accuracy rates are still high. For the subadults, combining the methods is really interesting. The combined accuracy rate is much higher than the combined accuracy rate of each method separately. This indicates that the combination of methods can give high accuracy rates.

Table 28. The accuracy rates for the CFS for both the Spitalfields sample as the Wharram Percy sample (Molleson et al. 1998, 724).

CFS		Female	Male	Combined
Spitalfields	Adults	85.3%	94.7%	88.7%
	Subadults	60%	84.6%	77.8%
Wharram Percy	Adults	92.3%	86.7%	89.3%

For the subadults from the Wharram Percy sample accuracy rates could not be determined. As mentioned before, the lower permanent canine was measured. For the CFS it was hypothesized that a male CFS would have a larger permanent canine than the female CFS. This turned out to be true and this indicates that the CFS can indicate sexually dimorphism.

2.2.3.5 Summary

For each method separately the accuracy rates were not impressive when the method was used on the Spitalfields sample. For the Wharram Percy sample no accuracy rates could be determined, because the sex is unknown for the subadults. A comparison between the methods and measurements of the lower permanent canine do indicate that the methods can indicate sexual dimorphism. The CFS does improve the accuracy rate for the subadults from the Spitalfields sample which indicates that the combination of all three methods can be interesting.

2.3 Overview and discussion

As seen in the previous sections a total of twelve non-metric methods can be found in the literature to determine the sex of subadults. These methods are the elevation of the auricular surface, the greater sciatic notch depth and angle, the arch criterion, iliac crest, chin prominence, anterior dental arcade, eversion of the gonion region, the complete mandible, mandibular angle, mentum and the orbital morphology. Of these methods the anterior dental arcade and the complete mandible are not derived from adult methods. The other ten methods are derived from adult non-metric methods with varying accuracies. Within adults, the eversion of the gonial angle is seen as not usable (Loth and Henneberg 2000) and the elevation of the auricular surface is only used with additional methods, because the method itself is not sufficient enough to correctly estimate sex (White and Folkens 2005, 394). The remaining eight methods are seen as good indicators and are widely used (Bruzek 2002; Rösing *et al.* 2007; WEA 1980; White and Folkens 2005).

As seen in de previous paragraphs the accuracies of the various methods are variable. The original authors have high frequencies as seen for Schutkowski (between 70 and 80% accurate (Schutkowski 1993)) and Weaver (72.5% accurate (Weaver 1980)). However, when these methods were retested, the accuracies were often lower. Sutter had approximately the same accuracies as Schutkowski and Weaver for the pelvic methods, but often there are indications that the methods are age-related and not sex-related. When Sutter retested the mandibular methods he had much lower accuracies than Schutkowski and again it seems that the methods are associated with age and not with sex (Sutter 2003). Mittler and Sheridan also had the same accuracies as Sutter and Weaver (Mittler and Sheridan 1992). However, all the other authors that retested the methods of Schutkowski and Weaver had much lower accuracies (Cardoso and Saunders 2008; Loth and Henneberg 2001; Vlak *et al.* 2008). The accuracies are sometimes below chance which indicates that randomly assigning sex has a higher accuracy. The method proposed by Loth and Henneberg (2001) was also retested and both studies indicate that the method is not usable, because of low accuracies (Coqueugniot *et al.* 2002; Scheuer 2002). The methods of Molleson *et al.* (1998) have never been retested, but the adult accuracy rates do seem promising.

3. Materials and Methods

Within this chapter the materials used for the present research will be discussed. In addition, information about the background of the place of excavation, the excavation itself and the procedures in the lab will be presented. After this the non-metric methods that have been chosen for the present study will be outlined. A short description will be given together with pictures and sometimes with drawings as well. Following this, the lab procedures for the present research are described. The chapter will be concluded with a description of the statistical analyses that will be used within this research.

3.1 Materials

The research sample for this study consists of 59 skeletons from the Middenbeemster skeletal collection housed at the University of Leiden. The individuals have been identified through the archival records that were kept by the local church. The Historisch Genootschap Beemster and Prof. Dr. M.L.P. Hoogland have used these records to identify the skeletons from the excavation. The 59 skeletons are between newborn and twenty-one years old. Information about sex and age is visible in appendix A.

The Middenbeemster collection was excavated in the summer of 2011 by Hollandia archeologen in conjunction with the University of Leiden. The excavation took place near the Keyser church in Middenbeemster (Hakvoort *et al.* 2013, 11).

3.1.1 Historical background

The village Middenbeemster lies in the Beemster polder in North Holland in the Netherlands. The Beemster polder was created by drainage of the Beemster lake between 1607 and 1612 (Aten 2012, 11). The land that was created in this way was divided in squares. The entire plan was based on ideas from the Roman architect Vitruvius whereby virtue, usability and beauty were the main principles. The Beemster polder has been placed on the UNESCO World Heritage List in 1999, because of its unique design (Aten 2012, 30).

Part of the design was a church in the middle of the polder (Aten 2012, 28-29), the Reformed Church De Keyser. The building of the church commenced in 1618 and it was inaugurated in 1623 although the tower was not completed until 1661 (Kurpershoek 2012, 207-209).

Before the building of the church, the land was already in use as a graveyard. For the build of the church it was decided that the graveyard would be raised and that in the middle the church would be built. After the build, people were buried in the graveyard and from 1638 also inside the church when they had enough money. Because there were no other graveyards in the Beemster municipality, everyone was buried within this graveyard. In 1829 people were no longer buried inside the church and from 1867 the graveyard was no longer in use (Hakvoort *et al.* 2013, 14).

In 1829 the graveyard was bought by the municipality and the old graveyard was cleared out. The clay ground was dug away and new sand was laid in place. Most of the burials found in the excavation probably date from the period 1829 to 1867 (Hakvoort *et al.* 2013, 29).

Although the graveyard was no longer in use, the church continues to be used till the present day. In 1954 till 1959 the church was restored to its former glory. However, from the end of the 19th century less people came to the church. Nowadays, the church is still in use for various religious and non-religious occasions (Kurpershoek 2012, 229-232)

3.1.2 Excavation

In June to August 2011 the excavation at the graveyard of Middenbeemster was carried out (fig. 2). This excavation was carried out by Hollandia

Fig. 2. Excavation of Middenbeemster seen from the church top (image taken by T. Vermaas).

archeologen in conjunction with the University of Leiden (Hakvoort *et al.* 2013, 11). A total of approximately 450 individuals have been excavated and are currently housed at the University of Leiden. The graveyard was laid out in a grid design and burial layers were put on top of each other. Everyone was buried in a wooden rectangular chest which was still visible during the excavations (Hakvoort *et al.* 2013, 31). From 1829 the grid design was recorded on a map which was used to keep track of who was buried where. However, only people that paid for the place were put on the map. This means that it is possible that some people from one place were not recorded on the map. However, it is still possible to distinguish people from the map with the skeletons that have been excavated. This knowledge is used to identify people within the collection.

3.1.3 Lab procedures

After excavation, all the skeletons were put in separate boxes to be dried in a ventilated warehouse. Afterwards they were taken to the University of Leiden where first DNA samples were taken from the teeth and later on cleaned (Hakvoort *et al.* 2013, 20). The lab analysis consists of laying out the skeletons in anatomical order after which sex and age were estimated. Sex was estimated with the methods recommended by the WEA, the os pubis method of Phenice (1969), the morphology of the sacrum and osteometric tecnhiques. Subadults are not sexed. The age is estimated for adults with various techniques including the sympheseal method of Suchey and Brooks (1990) and the auricular surface of Lovejoy and colleagues (1985). For subadults other methods are used including tooth eruption and ossification progress. Additionally, the length and several indices are calculated as well for adults. The condition of the teeth, the non-metric methods and the pathologies of both adults and subadults are registered as well. The information is put in a database that can be used by everyone who is doing research with the Middenbeemster collection (Hakvoort *et al.* 2013, 36-40).

3.2 Methods

Within chapter two all the known non-metric methods that can be used to determine the sex of subadults are discussed. Their origin, their accuracy rates and researchers who retested the methods have been discussed in that chapter. These methods have all been selected for various reasons. The methods of Weaver (1980), Schutkowski (1993) and Loth and Henneberg (2001) have been tested by various researchers who all got different kind of accuracy rates (Cardoso and Saunders 2008; Coqueugniot *et al.* 2002; Hunt 1990; Mittler and Sheridan 1992; Scheuer 2002; Sutter 2003; Vlak *et al.* 2008). Some researchers got the same accuracy rates as the original researchers but others got accuracy rates who were lower than chance. The original researchers also had good statistics (Loth and Henneberg 2001; Schutkowski 1993; Weaver 1980) indicating that there is a relationship between the method and the true sex. these methods can be used to determine sex. Therefore, these methods are used within this research to see whether the high accuracy rates obtained by several authors, will also be obtained within this research. The methods of Molleson and colleagues have never been retested by anyone. All three methods are used within adult sex estimation with success (WEA 1980) and the research of Molleson and colleagues indicate that the methods can be used to determine the sex of subadults (1998).Within this section more details will be given about the non-metric methods and how they are scored by the original authors and within this research. Both pictures and drawings will be provided to give more detail about the various methods and how they are scored.

3.2.1 The pelvis

The methods for the pelvis consists of the elevation of the auricular surface by Weaver (1980) and the four methods of Schutkowski (1993). For each method a description and figures are provided. Most methods have two pictures of the male method and two pictures of the female method of different ages.

3.2.1.1 Elevation of the auricular surface

Weaver (1980) describes his non-metric method as follows: *"... if the sacro-iliac surface was elevated from the ilium along its entire length and along both the anterior and posterior edges of the sacro-iliac surface, the auricular surface was considered elevated and was so scored ..." (Weaver 1980, 192).*

When the auricular surface is elevated, the skeleton is scored as a female. When it is not elevated, it is scored as a male (Weaver 1980, 192). Both the auricular surface of a male and of a female are visible on figures 3 and 4 below.

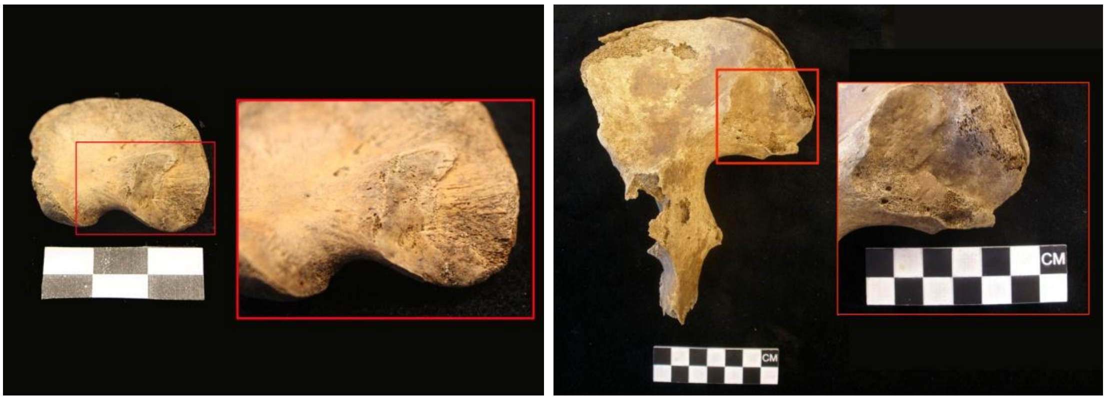

Fig. 3. Two female pelves with an elevated auricular surface. On the left S082V0084 (17 days old) and on the right S344V0730 (21 years old) (images taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; figures created by H. Vermaas).

Fig. 4. Two male pelves with a non-elevated auricular surface. On the left S035V0031 (2 years old) and on the right S196V0437 (13 years old) (images taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; figures created by H. Vermaas).

3.2.1.2 Angle of the greater sciatic notch

Schutkowski (1993) was the first person who described this method for subadults. The angle of the greater sciatic notch can best be scored when the bone is *"viewed from the ventral aspect and positioned in such a way that the anterior side of the greater sciatic notch is aligned vertically (Schutkowski 1993, 200-201)"*. When the greater sciatic notch has an angle over 90°, the skeleton is scored as female (fig. 5; below). When the greater sciatic notch has an angle that is around or smaller than 90°, the skeletons is scored as a male (fig. 6; below) (Schutkowski 1993, 201).

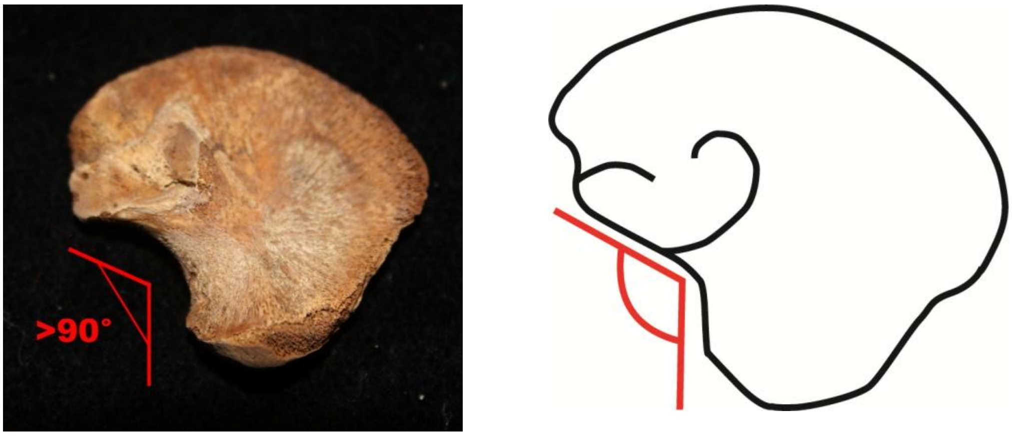

Fig. 5. A picture and a drawing of a female greater sciatic notch angle. On the left a picture of S373 V0798 (7 days old) and on the right a drawing (from Schutkowski 1993, 201) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H.

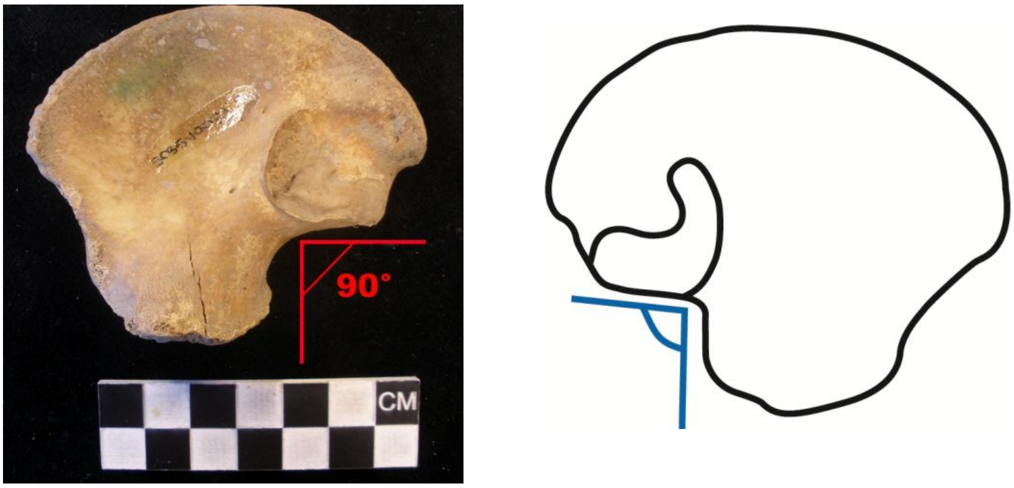

Fig. 6. A picture and a drawing of a male greater sciatic notch angle. On the left a picture of S035V0031 (2 years old) and on the right a drawing (from Schutkowski 1993, 201) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

3.2.1.3 Depth of the greater sciatic notch

This method was created by Schutkowski in 1993. The depth of the greater sciatic notch can best be scored *"from the dorsal aspect, the spina iliaca posterior inferior and the dorsal rim of the acetabular region point downwards in a line (Schutkowski 1993, 201)"*. Within females the notch is shallow while in males the notch is pretty deep (fig. 7 and 8) (Schutkowski 1993, 201).

Fig. 7. A picture and a drawing of a female greater sciatic notch depth. On the left a picture of S082 V0084 (17 days old) and on the right a drawing (from Schutkowski 1993, 201)(picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

Fig. 8. A picture and a drawing of a male greater sciatic notch depth. On the left a picture of S035V0031 (2 years old) and on the right a drawing (from Schutkowski 1993, 201)(picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

3.2.1.4 Arch criterion

This method was created by Schutkowski (1993) as well. The view that is used for the angle of the greater sciatic notch can also be applied to this method. The ilium is *"viewed from the ventral aspect and positioned in such a way that the anterior side of the greater sciatic notch is aligned vertically (Schutkowski 1993, 200-201)"*. Females have an arch when from the vertical side of the greater sciatic notch a line is drawn that crosses the auricular surface (fig. 9). Within males the line will not cross the auricular surface, but go to the lateral rim of the auricular surface (fig 10) (Schutkowski 1993, 201).

Fig. 9. Two pictures and a drawing of a female arch criterion. On the left a picture of S082V0084 (17 days old) and in the middle a picture of S286V0469 (10 years old). On the right a drawing (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

Fig. 10. Two pictures and a drawing of a male arch criterion. On the left a picture of S035V0031 (2 years old) and in the middle a picture of S350V0844 (19 years old). On the right a drawing (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

3.2.1.5 Iliac crest

The last method of the pelvis and the last method that Schutkowski (1993) created is the curvature of the iliac crst. The best way to score the curvature of the iliac crest is by viewing the ilium from the top and aligning the dorsal surface horizontally. Females will have an iliac crest with a faint S-shape, while males will have a clear S-shaped iliac crest (fig. 11 and 12) (Schutkowski 1993, 201).

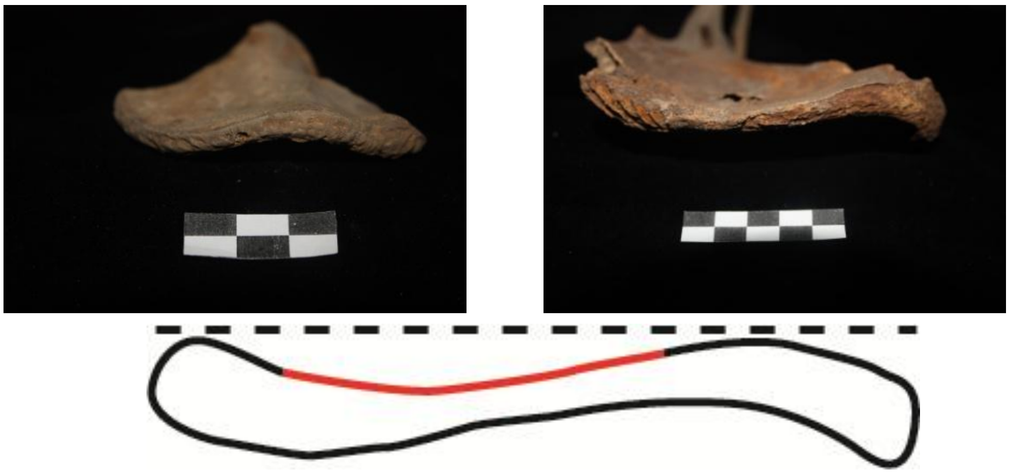

Fig. 11. Two pictures and a drawing of a female iliac crest. On the left a picture of S165V0242 (2 years old) and on the right a picture of S047V0045 (21 years old). Below a drawing of the iliac crest (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

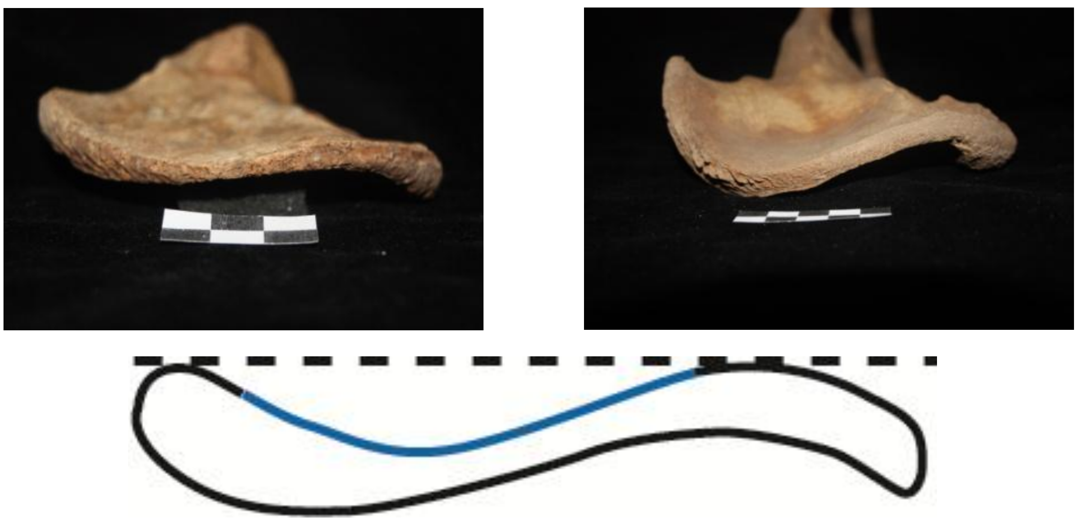

Fig. 12. Two pictures and a drawing of a male iliac crest. On the left a picture of S364V1517 (6 years old) and on the right a picture of S350V0844 (19 years old). Below a drawing of the iliac crest (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas).

3.2.2 The skull

For the skull seven methods to determine the sex of subadults are known. These methods are the protrusion of the chin, the shape of the anterior dental arcade, the eversion of the gonion region, the complete mandible, the mentum, the mandibular angle and the orbital morphology. The methods will be described in the following sections including images of both female and male expressions of these methods.

3.2.2.1 Protrusion of the chin

The protrusion of the chin was created by Schutkowski in 1993. He describes it for the females as follows: *"The chin region is not prominent and the mandible does not show distinct elevated rough structures at the side of the median sagittal line. The surface of the bone is smooth. Seen from the top, the mental protrusion appears faint and narrow and sometime tapers (Schutkowski 1993, 200) (fig.13)."* For males the chin region is more prominent and rough. The protrusion of the chin is best seen from the top and anterior (Schutkowski 1993, 200) (fig. 14).

Fig. 13. A female protrusion of the chin. On the left a picture of S167V0270 (12 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

Fig. 14. A male protrusion of the chin. On the left a picture of S384V0839 (8 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

3.2.2.2 The shape of the anterior dental arcade

This method was created by Schutkowski in 1993. The mandible is viewed from the top and from the side. For females, the shape of the anterior dental arcade is: *"... a roughly parabolic shape (Schutkowski 1993, 200) (fig. 15)."* This is caused by the fact that the alveoli of the front teeth form a rounded contour and the dental arch is not protruded by the canines. For males the anterior dental arcade is more U-shaped (Schutkowski 1993, 200) (fig. 16).

Fig. 15. A female anterior dental arcade. On the left a picture of S046V0023 (2 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

Fig. 16. A male anterior dental arcade. On the left a picture of S037V0013 (3 months old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

3.2.2.3 The eversion of the gonion region

The last method of Schutkowski for the mandible is the eversion of the gonion region. The mandible is viewed from the frontal aspect and for females "...

the outer surfaces of the mandible are aligned with the gonion region (Schutkowski 1993, 200)(fig. 17)." For males the gonion is *"... everted, protruding slightly beyond the general surface of the mandible's horizontal rami (Schutkowski 1993, 200) (fig. 18)."*

Fig. 17. A female eversion of the gonion region. On the left a picture of S082V0084 (17 days old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

Fig. 18. A male eversion of the gonion region. On the left a picture of S400V0859 (24 days old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University).

3.2.2.4 The complete mandible

Loth and Henneberg (2001) also created a non-metric method for the mandible to estimate the sex of subadults. The mandible is divided in two parts namely the symphyseal base and the body. The symphyseal base is rounded in females and squared within males. The female shape is gradually curved with a gradual transition between the symphyseal base and the body. For males the transition between the symphyseal base and the body is clearly angled (Loth and

Henneberg 2001, 182) (fig. 19 and 20). The method of Loth and Henneberg is named the complete mandible method within this research.

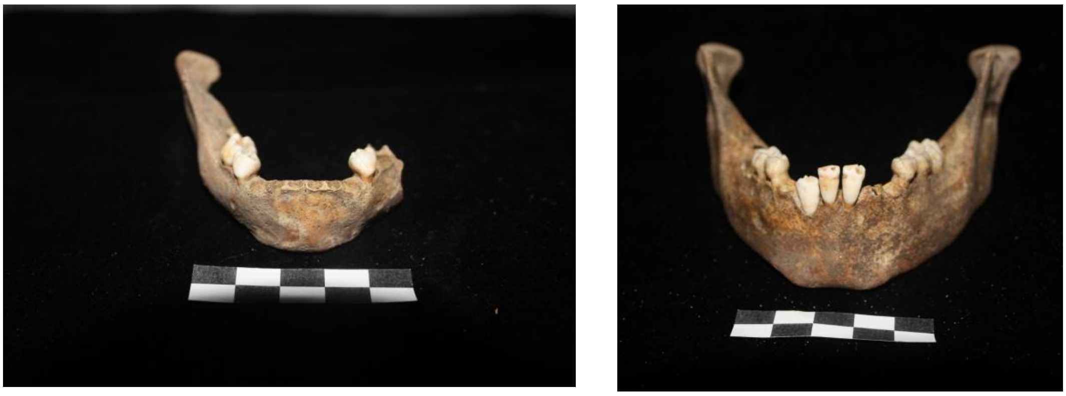

Fig. 19. A female and a male complete mandible. On the left a female (S328V0701; 3 years old) and on the right a male (S384V0839; 8 years old) (pictures taken by T. Vermaas; from the Laboratory for Human Osteology, Leiden University).

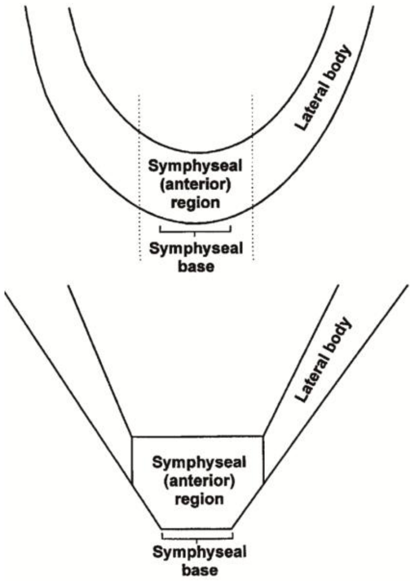

Fig. 20. A schematic design of the complete mandible method from Loth and Henneberg (2001, 181).

3.2.2.5 The mandibular angle

Molleson and colleagues (1998) adopted this method from the adults methods to estimate sex. In the original article a five-point scale is used, but for this research the scale will be changed to a male and female division for better comparison with the other methods. Within the five-point scale a distinction is made between male, probable male, indeterminate, probable female and female.

For this research this is changed to male and female by taking the probable male and male together and the probably female and female. The indeterminate is discarded from this research since indeterminate indicates that the method cannot see whether the skeleton is male or female. This means that the description of this method a combination is from the probably male and male for the male description and probably female and female for the female description.

For the mandibular angle, the mandible is viewed from the side, perpendicular to the ramus. A female mandibular angle is underdeveloped. A less female mandibular angle has an ascending ramus that is evenly developed whereby the outline smooth is and the angle moderate (fig. 21). A male mandibular angle has an ascending ramus that is more hinged on the body. It is possible that rugosity and thickening of the bone is visible near the angle (fig. 22) (Molleson *et al.* 1998, 720).

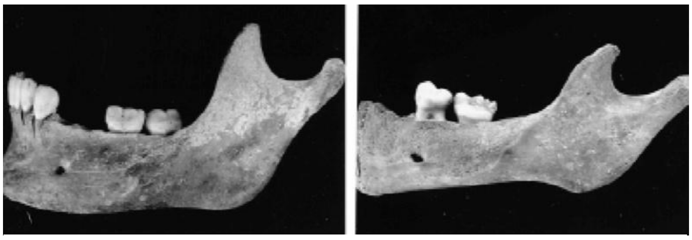

Fig. 21. Two female mandibular angles with on the left a less female mandibular angle and on the right a clear female mandibular angle (Molleson et al. 1998, 721).

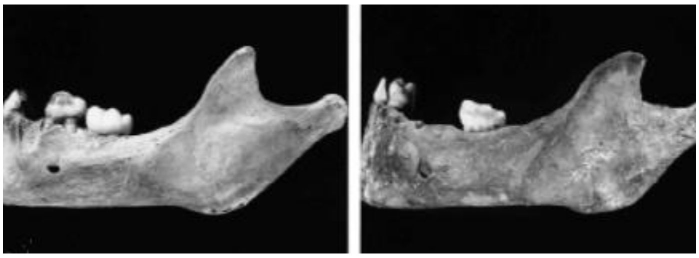

Fig. 22. Two male mandibular angles with on the left a less male mandibular angle and on the right a clear male mandibular angle (Molleson et al. 1998, 722).

3.2.2.6 The mentum

The second method from Molleson and colleagues (1998) is the mentum. As mentioned before, within the original article a five-point scale is used, but this is changed to a distinction in female and male within this research. The description of the mentum is a combination from the probably male and male for the male description and probably female and female for the female description.

The mentum is viewed from above. For females is the outline of the mandibular body smooth and evenly curved. The tubercles, found on the lower part of the mandibular body, are only weakly developed, sometimes undeveloped, and wide apart. In addition, the mental triangle is narrow and does not protrude the line of the symphysis when looking in a lateral view (fig. 23). For males, the tubercles on either of the mentum, square off the outline of the mandible. Sometimes, the tubercles are clearly pronounced. The mental triangle is broad and prominent and clearly delimited (fig. 24)(Molleson *et al.* 1998, 723).

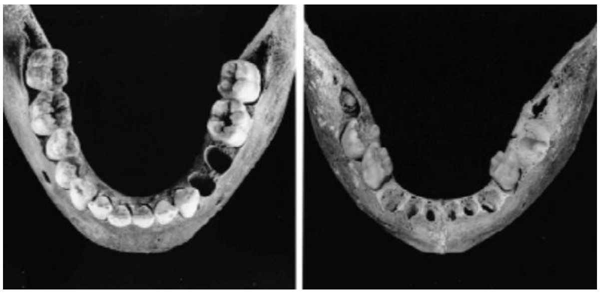

Fig. 23. Two female mentums with on the left a less female mentum and on the right a clear female mentum (Molleson et al. 1998, 721).

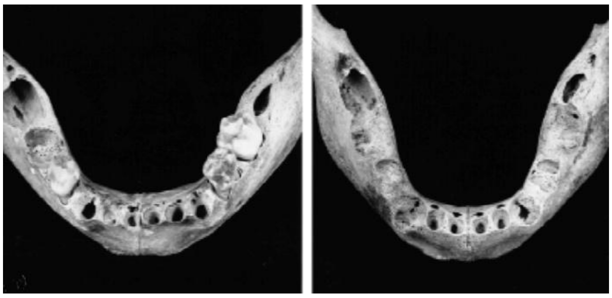

Fig. 24. Two male mentums with on the left a less male mentum and on the right a clear male mentum (Molleson et al. 1998, 722).

3.2.2.7 The orbital morphology

The last method for this research is the orbital morphology. This method was used by Molleson and colleagues in 1998. The original authors used a five-point scale as mentioned before. Within this research a binary division in female and male is used. This means that the classes probable female and female are taken together in the description for females and that probably male and male are also taken together in the description for males.

Fig. 25. A female orbital morphology (Molleson et al. 1998, 721).

To assess the orbital morphology, the skull should be viewed from the front, full-face, with the face vertical. For females the shape of the orbit is symmetrical, with a thin or a sharp rim and the lateral part of the roof is deep-set. When looking at the orbit from the side, the profile does not interrupt the outline of the face (fig. 25; previous page). For males, the orbit is more asymmetrical in outline, with a thick rim and the lateral part of the roof is shallow. In addition, the surface of the brow may bulge slightly and the lower border of the cheek region can be thickened. When looking at the orbit from the side, the orbit interrupts of the line of the face (fig. 26)(Molleson *et al.*1998, 720).

Fig. 26. A male orbital morphology (Molleson et al. 1998, 722).

3.3 Lab procedures

For this research a blind test was conducted on the skeletons. This means that the author was unaware of the sex division before starting the research. The aforementioned division of male and female was only known after the research was conducted. In this way it was ensured that the author had no bias when doing the research and that the test was done completely blind.

All the before mentioned methods are scored in the lab on the form that can be found in appendix B. The right side will be scored unless the right side is missing. In that case the left side was scored. All the methods have been described in the previous paragraphs. To ensure that there was no influence from the outcomes of the other methods for one skeleton, it has been decided that each method has its own form.

Only bones that were not damaged on the parts of the methods that were scored were used. This means that for some methods more skeletons were available than for other methods. It also means that for some methods, namely the chin prominence, complete mandible and mentum, the mandible needs to be fully closed. This happens during the first year (Black and Scheuer 2000, 147) and therefore some skeletons will not be scored on these four methods. For the orbital morphology methods there is the same problem. The bones that make up the orbit are not fused before the second or third year (Black and Scheuer 2000, 124). This could limit the amount of skeletons that can be used for the orbital morphology method.

After scoring, the results were put into an SPSS database. The results for each skeleton is compared with the true sex of the skeleton. This comparison will result in accuracy rates for each method. These accuracy rates can then be compared to the accuracy rates obtained in previous research as discussed in chapter two. Additionally, accuracy rates for the pelvis, skull and overall will be created.

The accuracy of the pelvis was determined by using the five pelvic methods, one of Weaver (1980) and four of Schutkowski (1993). The amount of female and male determinations were calculated and the highest of the two was seen as the indicator for the pelvis accuracy. The other methods were all counted for the skull accuracy. When female and male were equal the skeleton was rated as indeterminate. The indeterminate are not counted with the accuracy rates. It is unknown whether the sex is estimate correctly, so they have been omitted from the sample. The determinations were compared with the archival age.

3.4 Statistical analyses

For this research several statistical tests will be done. These tests are the chi-square (χ^2) test, the phi (ϕ) test and the binary logistic regression. All the tests were conducted by using SPSS 22.0 (Statistical Package for the Social Sciences).

3.4.1 The χ^2 test

The χ^2 test can tell whether there is a relationship between categorical variables. The test compares the observed frequencies within the categories with the frequencies that can be expected within those categories by chance. For the results of the present study only two categorical variables are used namely male and female (Field 2009, 687-689). This test is important for understanding whether the accuracy rates obtained within this research are caused by the method or by chance. For the χ^2 test two hypotheses are created. The first hypothesis (H_0) is that the used method cannot determine sex better than chance. The second hypothesis (H_1) is that the used method can determine sex better than chance. The H_1 hypothesis is only accepted when there is a 95% chance that the method can determine the sex of subadults. Therefore, the p-value is set at ≤ 0.05 (Howell 2007, 144-148). This means that there is a 5% chance that we accept the hypothesis that the method can determine sex better than chance while this is not the case. This is called the Type I error. The χ^2 test is done on all twelve methods of this study. This means that the chance of a Type I error increases to 60% ($12*0.05 = 0.6 = 60\%$). To counteract this problem the Bonferroni correction can be used. This is done by dividing the p-value of 0.05 by the amount of tests (Field 2009, 372-373). This means that the p-value drops to 0.0042. The H_1 hypothesis is only accepted when the p-value is ≤ 0.0042.

For the χ^2 test the expected frequencies in each cell must be at least 5. When the expected frequency of at least one cell is below 5, the sample size is too small. This means that the sampling distribution might be too deviant from a χ^2 distribution to be of any use. To counteract this problem the Fisher's exact test has been created. In this way, a reliable χ^2 test can be done on small samples (Field 2009, 690). It is possible that for some methods within the present study the amount of skeletons is too small for a normal χ^2 test. In those cases a Fisher's exact test will be done. Within the results section it will be mentioned whether a χ^2 test was done or a Fisher's exact test.

The χ^2 test will be used on all twelve method and on the accuracy rates for the pelvis, skull and overall.

3.4.2 The ϕ test

The ϕ test gives an indication whether there is a relation or association between a certain method and a certain variable. It seems that this test is the same as the χ^2 test, but that test tells us whether the variables are correlated. Both the χ^2 test and the ϕ test give the same answer, but the ϕ test indicates the degree or the magnitude between the variables. Within the present study the χ^2 test is used to see whether there is a correlation between the method and the true sex while the ϕ test is used to see whether the relationship between the method and the true sex is strong enough to say, with certainty, that the method can be used to determine the sex of subadults (Howell 2007, 309-313).

The ϕ test will give a value between 0 and 1. The given value can be interpreted based on these rules:

- 1 = complete association
- 0.75 = strong association
- 0.50 = moderate association
- 0.25 = weak association
- 0 = no association

The ϕ test will be done on all twelve method and on the accuracy rates for the pelvis, skull and overall.

3.4.3 Logistic regression

In addition a logistic regression analysis will be done on the results. Within this analysis a regression function is fitted to the data to predict the outcome of a variable (Meyers *et al.* 2006, 221-222). In the case of the present data it would be interesting to know that if a skeleton is classified as female with the method, that the predicted outcome will be female as well. To see whether that is the case, a logistic regression analysis will be used. Because the data is binary, a skeleton is either classified as male or female, a linear function is not feasible. Therefore, a S-shaped function is created (fig. 27). The S-shaped function gives this analysis the predictive power it

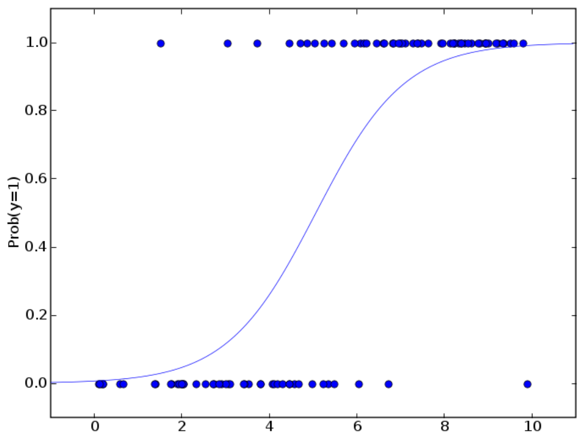

Fig. 27. An example of a logistic regression analysis with a clear S-shaped function (http://abel.ee.ucla.edu/cvxopt/examples/book/figures/fig-7-1).

needs (Meyers *et al.* 2006, 226). First, a χ^2 test will be conducted to see whether the model predicts the sex at levels that are greater than chance. After this several log-tests will be conducted and a model is created which can be used to determine the sex of subadults (Howell 2007, 561-563). When interpreting the logistic regression analysis, it is first important to see what the outcomes are of the χ^2 test. If the p-value of this test is above 0.0042, the same p-value as was used in the regular χ^2 test, the model cannot be used.

Within this research the logistic regression analysis will be used on all the methods and the accuracy rates of the pelvis, skull and overall.

4. Results

Within this chapter the results of the research will be analyzed. First, the general results will be shown which will follow with the results for each method separately. The results will be compared to the results in the literature as discussed in chapter two. The results are divided in the same age groups that have been used in chapter two. With each method the accuracy rates, the accuracy rates for each age groups and the results of the statistical analyses will be given. These results will be summarized in the final section.

4.1 General results

A total of 59 skeletons were used in this research (appendix A). From 48 skeletons there were enough bones to estimate the sex. For four skeletons the sex could not be determined by the archive. One skeleton was older than 21 years and was thus omitted from the sample. This skeleton was omitted, because the research is about subadults and the age range is set to newborn to 21 years.

A total of 44 skeletons were used with 26 female and 17 male (table 29). There are more females (61.4%) than males (38.6%) (table 29). This indicates that there is a skewed ratio of males and females. This could influence the results of the statistical analyses. However, statistics indicate that there is a 17% chance that this spread within a sample happens again. This indicates that the skewed ratio of females and males, should not be a problem for the statistical analyses.

All twelve non-metric methods were tested on the skeletons. Not all methods could be measured on all skeletons (appendix

Table 29. Table with the amount of females and males for each age group that were used within this research.

Age group	Female	Male	Total
NB - 1 year	3	4	7
2 - 5 years	8	2	10
6 - 10 years	8	2	10
11 - 15 years	3	3	6
16 - 21 years	5	5	10
Unknown	0	1	1
Total	27	17	44

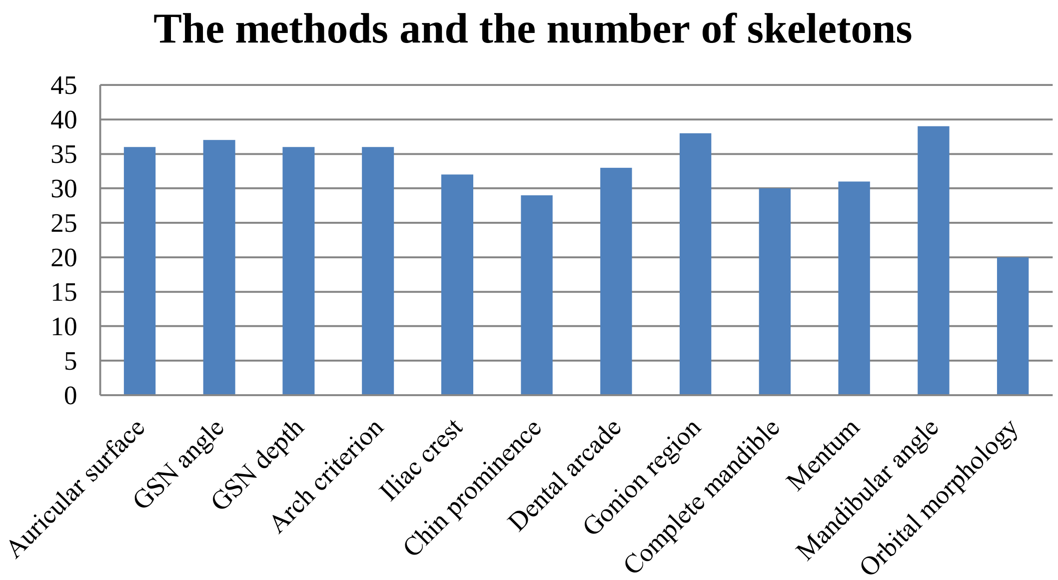

Fig. 28. All twelve methods with the amount of skeletons used for each method.

A). The number of skeletons for each method is visible in figure 28. General accuracy rates were determined, namely an accuracy for the pelvis, for the skull and an overall accuracy. The results of all three accuracies are visible in table 30.

Table 30. The accuracy rates of the pelvis, skull and overall. The number of skeletons is indicated with N.

	Female		Male		Combined	
	N	%	N	%	N	%
Pelvis	23	39.1%	13	84.6%	36	55.6%
Skull	24	62.5%	11	18.2%	35	48.6%
Overall	25	64%	14	50%	39	59%

These accuracy rates cannot be compared to any accuracy rates in the literature, because this kind of research has not been done before. The accuracy rates are low. The pelvis and the overall accuracy both show an accuracy rate barely above chance and the accuracy rate of the skull is even below chance. Within the pelvis the females are often not correctly estimated while males have a high accuracy of 84.6%. This could indicate that these methods indicate namely males and that therefore the males have a high accuracy rate. This has been noted before by Scheuer (2002) when she retested the method of Loth and Henneberg (2001). For the skull, and the overall accuracy rate, the males are below chance while females are barely above chance. This all indicates that a combination of these methods is not useful for estimating the sex of subadults.

Table 31. The accuracy rates for the pelvis, skull and overall divided in five age groups and by female and male.

		Female		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Pelvis	3	66.7%	1	100%	4	75%
	Skull	4	25%	1	0%	5	20%
	Overall	3	66.7%	2	50%	5	60%
2 - 5 years	Pelvis	5	40%	2	100%	7	57.1%
	Skull	6	66.7%	1	0%	7	57.1%
	Overall	7	57.1%	2	50%	9	55.6%
6 - 10 years	Pelvis	6	33.3%	2	50%	8	37.5%
	Skull	7	57.1%	2	0%	9	44.4%
	Overall	8	50%	2	0%	10	40%
11 - 15 years	Pelvis	3	33.3%	3	66.7%	6	50%
	Skull	2	50%	2	50%	4	50%
	Overall	3	66.7%	2	50%	5	60%
16 - 21 years	Pelvis	5	20%	4	100%	9	56%
	Skull	5	100%	4	0%	9	55.6%
	Overall	3	100%	5	60%	8	75%

One of the sub-questions of this thesis is whether a certain age group has higher accuracies than other age groups. The various age groups with their accuracy rates are visible in table 31, above. From this table it becomes clear that there is not really one age group which performs much better than the others. The age group 6 to 10 years performs the worst of all. Within this age group females perform the best of all with accuracy rates of 33.3% for the pelvis, 57.1% for the skull and an overall accuracy rate of 50%. For the males the accuracy rates are much worse namely 50% for the pelvis and 0% for both the skull and overall. It is possible that this is caused by the fact that Schutkowski has developed his method on subadults between the ages of newborn and five years (1993, 200). Therefore, it is possible that the methods do not work correctly on older individuals. However, research by Sutter (2003) has indicated that the methods should work on older individuals as well. In addition. The accuracy rates of the pelvis, skull and overall do not only consist of methods from Schutkowski (1993). The pelvis has one additional method, namely the auricular surface elevation from Weaver (1980). The skull accuracy rate has one method from Loth and Henneberg (2001) and three methods of Molleson and colleagues (1998). The methods from Molleson and colleagues have been tested on older age groups and this worked very well (1998). It is possible that the age groups over the age of five have lower accuracy rates, because the methods were mostly developed on subadults below five years. But there are indications that this should not be the case.

What is striking is that within all the age groups females perform better than males. In general, males perform better than females. (Baker *et al.* 2005, 10). For males, the pelvis performs best in the age groups newborn to one year, two to five years and sixteen to twenty-one years. In all these age groups the accuracy rate is 100%. The skull and overall rates are continuously low for the males. Within the age group 16 to 21 years is the overall accuracy rate 60%. In all the other age groups, the accuracy rate is either 0% or 50%. For females, the accuracy rates are more divided. For the newborn to one year group the accuracy rates for the pelvis and overall are 66.7%. For the skull the accuracy rate is only 25%. For the age group two to five years the accuracy rates are 40% for the pelvis, 66.7% for the skull and 57.1% for the overall rate. For the six to ten year olds, the accuracy rates are 33.3% for the pelvis, 57.1% for the skull and 50% for overall. The accuracy rates for the group eleven to fifteen years is 33.3% for the pelvis, 50% for the skull and 66.7% overall. The last age group, sixteen to twenty-one years, has an accuracy rate of 20% for the pelvis, but females score a 100% for both the skull and the overall accuracy rate.

It seems that there is not one specific age group in which the accuracy rates are much better than the other groups.

4.1.1 Statistical analyses

Three statistical analyses were executed on these results. On all three the χ^2 test, ϕ test and logistic regression analysis were used. For these statistics the Statistical Package for the Social Sciences (SPSS 22) was used. For all three tests the results are visible in table 32.

Table 32. The three rates and their statistical results.

		Chi-square		Phi	Logistic regression
	N	Value	P	Value	Value
Pelvis	36	2.207	0.137	0.248	0.293
Skull	35	1.306	0.253	0.193	0.217
Overall	39	0.727	0.394	0.137	0.448

Within the table it is visible that none of the statistical analyses shows any statistical significance between the estimation methods and the true sex. The χ^2 test shows that the pelvis methods cannot be used to determine sex better than chance ($\chi^2(1) = 2.207$, n = 36, p = 0.137). The same is true for the skull ($\chi^2(1) = 1.306$, n = 35, p = 0.253) and the overall rate ($\chi^2(1) = 0.727$, n = 39, p=0.394). The ϕ test indicates that there is no association between the pelvis and the true sex ($\phi = 0.248$). Which is also the case for the skull ($\phi = 0.193$) and overall ($\phi = 0.137$) accuracy rate. The logistic regression model for the pelvis rate was non-significant (χ^2 (1) = 2.364, p = 0.124) indicating that this rate cannot be used to determine the sex of subadults. For the skull rate the logistic regression model was also non-significant (χ^2 (1) = 1.388, p = 0.239). The logistic regression model for the overall rate was non-significant (χ^2 (1) = 0.723, p = 0.395) as well. All the statistical tests and accuracy rates indicate that combining the methods does not improve the accuracy rate or the ability to determine the sex of subadults.

4.2 Results for the pelvis methods

Within this paragraph the five pelvis methods will be discussed together with their accuracy rates and the results of the statistical analysis. The accuracy rates will be compared to the rates of the authors described in chapter two. The rates will be further divided by sex and in age groups.

4.2.1 Auricular surface

A total of 37 skeletons could be used for the sex estimation using the auricular surface treat of Weaver (1980). The accuracy rate found in this study is 45.9%. Three previous studies have reported accuracy rates for this method, namely the original article by Weaver (1980) and two other articles by Mittler and Sheridan (1992) and Sutter (2003). Their accuracy rates compared with the accuracy rate of this study are visible in table 33, below.

The accuracy rate obtained in this study is below chance (45.9%) and much lower than the accuracies of the other studies. Weaver (1980) had an accuracy rate of 73.5%, Mittler and Sheridan (1992) had an accuracy rate of 74.1% and Sutter (2003) had a rate of 71.7%. These accuracy rates are all much higher than the accuracy rate of 45.9% found in this study (table 33, below).

Within this study males have a high accuracy of 91.7% and the females had an accuracy rate of 24% (table 33, below). This indicates that this method can be used on males, but is useless for females. It still indicates that this method cannot be used within archaeology or forensics. That males have such a high accuracy rate and females such a low rate, could indicate that this method estimate almost everything as male. Therefore, the method scores good on males, but worse on females.

Table 33. The accuracy rates of Weaver (1980), Mittler and Sheridan (1992), Sutter (2003) and the present study divided in female, male and combined for the auricular surface. The number of skeletons is indicated with N.

Auricular surface	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	25	24%	12	91.7%	37	45.9%
Weaver (1980)	71	57.7%	80	85.4%	151	73.5%
Mittler and Sheridan (1992)	24	58.3%	34	85.3%	58	74.1%
Sutter (2003)	23	60.9%	37	77%	60	71.7%

The accuracy rates can also be divided in age groups. This can be seen in table 34, below. The age groups are compared with research done by Sutter (2003). For the males it is clear that all age groups are interesting for this method. The methods scores almost 100% in every case. Only for the older age group of sixteen to twenty-one years is the accuracy rate 75%. The method cannot be used on the females. The first age group, newborn to one year has a rate of 66.7%, but the following two age groups, two to five years and six to eleven years, have an accuracy rate of 0%. The method does not perform well on females in these two age groups. The age group eleven to fifteen has a low accuracy rate of 33.3% and for the last age group, sixteen to twenty-one, the accuracy rate climbs to 60%. Sutter (2003) had much higher accuracy rates for the age groups two to five years, six to ten years and eleven to fifteen years. Within this study the method scores better on the males than within Sutter (2003) his study. It seems that for females the method might be used for the age group of newborn to one. For the males, the method can be perfectly used for all the age groups. Unfortunately, there are too little skeletons to be able to confirm this with statistics for each age group. In addition, if the method works perfectly for only one sex, the method can still not be used within archaeology or forensics, since we do not know the sex beforehand.

Table 34. The accuracy rates divided in five age groups and divided by male and female for the auricular surface. For each age group the total number of skeletons (N) is given.

Auricular surface		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB-1 years	Present study	3	66.7%	1	100%	4	75%
	Sutter (2003)	7	42.8%	6	67%	13	53.8%
2-5 years	Present study	6	0%	1	100%	7	14.3%
	Sutter (2003)	5	80%	7	100%	12	91.7%
6-10 years	Present study	7	0%	2	100%	9	22.2%
	Sutter (2003)	4	50%	7	100%	11	81.8%
11-15 years	Present study	3	33.3%	3	100%	6	66.7%
	Sutter (2003)	7	71.4%	17	58.8%	24	62.5%
16-21 years	Present study	5	60%	4	75%	9	66.7%

For the statistical analysis a χ^2 test, a ϕ test and a logistic regression analysis were conducted. Two cells had an expected frequency below 5 which means that a Fisher's exact test had to be done. The test indicates that the auricular surface cannot be used to accurately determine the sex of subadults ($\chi^2(1) = 1.297$; $n = 37$; $p = 0.389$) since the p-value is above 0.0042. In addition, the ϕ test indicates that there is almost no association between the method and the true sex ($\phi = 0.187$). The logistic regression model turned out to be non-significant (χ^2 (1) = 1.455, $p = 0.228$) and this indicates that the auricular surface cannot be used to determine the sex of subadults. Both the statistical analyses and the accuracy rates indicate that the elevation of the auricular surface cannot be used to determine the sex of subadults.

4.2.2 Greater sciatic notch angle

For the greater sciatic notch angle 38 skeletons could be used. The accuracy rate is 62.1% (table 35, below). This is mostly caused by the fact that the males have a high accuracy rate of 84.6%. Females have an accuracy rate of 50%. These rates can be compared with the original author (Schutkowski 1993) and two authors who retested this method (Sutter 2003; Vlak *et al.* 2008). The accuracy rate of the present study is not the lowest accuracy ever achieved. Vlak and colleagues had a lower accuracy (2008). Still, the accuracy rate is not high enough to give reasons to use this method within archaeology or forensics.

Table 35. The accuracy rates of Schutkowski (1993), Sutter (2003), Vlak et al. (2008) and the present study divided in female, male and combined for the greater sciatic notch angle. The number of skeletons is indicated with N.

Greater sciatic notch angle	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	25	48%	13	84.6%	38	60.5%
Schutkowski (1993)	17	95.2%	27	71.4%	44	81.6%
Sutter (2003)	25	78.6%	37	69.2%	62	74.1%
Vlak *et al.* (2008)	23	52.5%	33	54.5%	56	53.6%

The accuracy rates can be further divided in five age groups, which can be compared to the rates that Schutkowski (1993), Sutter (2003) and Vlak *et al.* (2008) obtained (table 36, below). The age group analysis demonstrates that this method can be applied to the newborn to one year olds. The present study and all the other researchers obtained good accuracy rates for this age group. Within the other age groups, the method performed well for the males, but the accuracy rates for the females continuously decreased. The only age group for which the method does not perform well for the males is the third age group from six to ten years where the accuracy rate was 50%. For females the accuracy rates for the two to five year olds is 50%, for the six to ten year olds it is 42.9%, for the eleven to fifteen year olds the accuracy rate was 33.3% and for the last age group the

Table 36 The accuracy rates divided in five age groups and divided by male and female for the greater sciatic notch angle. For each age group the total number of skeletons (N) is given.

Greater sciatic notch angle		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	3	100%	1	100%	4	100%
	Schutkowski (1993)	11	100%	11	55%	22	77.3%
	Sutter (2003)	7	71.4%	6	50%	13	85.7%
	Vlak *et al.* (2008)	7	100%	9	78%	14	87.5%
2-5 years	Present study	6	50%	2	100%	8	62.5%
	Schutkowski (1993)	10	90%	17	82%	27	85.2%
	Sutter (2003)	7	85.7%	7	86%	14	85.7%
	Vlak *et al.* (2008)	8	37.5%	10	40%	18	38.9%
6-10 years	Present study	7	42.9%	2	50%	9	44.4%
	Sutter (2003)	4	25%	10	100%	14	78.6%
	Vlak *et al.* (2008)	5	20%	7	71%	12	50%
11-15 years	Present study	3	33.3%	3	66.7%	6	50%
	Sutter (2003)	7	71.4%	16	94%	23	86.9%
	Vlak *et al.* (2008)	3	33.3%	7	100%	10	80%
16-21 years	Present study	5	20%	4	100%	9	44.4%

accuracy rate decreased to 20%. It seems that this method can be used for the age group newborn to one year olds. There are unfortunately not enough skeletons in each age group to do any statistical analysis. For all the other age groups, the method works for males, but not for females. However, the method cannot be used to determine the sex of subadults within archaeological or forensic context.

A χ^2 test, a ϕ test and a logistic regression analysis were applied on this method. The expected frequency of one cell was below five which means that a Fisher's exact test had to be done instead of a χ^2 test. This test indicate that the method cannot be used to determine the sex of subadults ($\chi^2(1) = 3.910$; $n = 38$; $p = 0.077$). The p-value is above 0.0042. The ϕ test indicates that there is a small association between the greater sciatic notch angle and the true sex ($\phi = 0.321$). In addition, the logistic regression analysis was non-significant (χ^2 $(1) = 4.237$, $p = 0.040$). Both the accuracy rates as the statistical analyses indicate that the angle of the greater sciatic notch cannot be used to determine the sex of subadults.

4.2.3 Greater sciatic notch depth

A total of 37 skeletons could be used for the greater sciatic notch depth method. The accuracy for the greater sciatic notch depth in the present study is 45.9% (table 37). When the accuracy rates are divided in females and males, the accuracy rates are not much lower. These rates can be compared to the rates that the original author had and the two succeeding authors who retested the method. Schutkowski (1993), the original author, had a much higher rate of 76.9%. Sutter (2003) even had a higher rate of 81.5%. However, when Vlak and colleagues

Table 37. The accuracy rates of Schutkowski (1993), Sutter (2003), Vlak et al. (2008) and this study divided in female, male and combined for the greater sciatic notch depth. The number of skeletons is indicated with N.

Greater sciatic notch depth	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	25	44%	12	50%	37	45.9%
Schutkowski (1993)	17	68.4%	27	86.7%	54	79.6%
Sutter (2003)	25	92.9%	37	69.2%	62	81.5%
Vlak *et al.* (2008)	23	43.5%	33	69.7%	56	58.9%

retested the method, the accuracy fell to 58.9%. The accuracy rate of this study falls below chance and even below the accuracy rate of Vlak and colleagues. The accuracy rates for both males and females indicate that the method cannot be used to determine the sex of subadults.

The results can be further divided into five age groups (table 38, below). Schutkowski (1993), Sutter (2003) and Vlak *et al.* (2008) have used the same age groups making the data comparable. A general trend from the other authors is that the females perform the best in the younger age groups and the males in the older age group. Within the present study, this trend is visible as well. The method performs well on the females. An accuracy rate of 100% was achieved in the first age group of newborn to one year. In the ages two to five years the results

Table 38. The accuracy rates divided in five age groups and divided by male and female for the greater sciatic notch depth. For each age group the total number of skeletons (N) is given.

Greater sciatic notch depth		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	3	100%	1	0%	4	75%
	Schutkowski (1993)	11	80%	11	40%	22	60%
	Sutter (2003)	7	85.7%	6	0%	13	46.1%
	Vlak *et al.* (2008)	7	100%	9	44%	16	68.7%
2-5 years	Present study	6	66.7%	1	0%	7	57.1%
	Schutkowski (1993)	8	83.8%	19	64%	27	73.1%
	Sutter (2003)	7	100%	7	71%	14	85.7%
	Vlak *et al.* (2008)	8	25%	10	50%	18	38.9%
6-10 years	Present study	7	40%	2	0%	9	28.6%
	Sutter (2003)	4	33.3%	8	80%	12	69.2%
	Vlak *et al.* (2008)	5	20%	7	100%	12	66.7%
11-15 years	Present study	3	33.3%	3	66.7%	6	50%
	Sutter (2003)	7	28.6%	17	88%	24	69.6%
	Vlak *et al.* (2008)	3	0%	7	100%	10	70%
16-21 years	Present study	5	0%	4	100%	9	50%

decrease to 66.7%. Within the following three age groups the accuracy rate decreases further. For the six to ten year olds the accuracy rate is 40%, for the eleven to fifteen year olds the rate is 33.3% and for the oldest group (sixteen to twenty-one years) the accuracy rate is 0%. On the males, the method does not perform well in the first age groups. The accuracy rates for the males are 0% for the first three age groups from newborn to ten years. There were only four males in total for these age groups and therefore the results should be viewed with caution. For the later age groups the method performs much better on the males. For the age group eleven to fifteen years the accuracy rate is 66.7% and for the age group sixteen to twenty-one years the accuracy rate is 100%. It is possible that the depth of the greater sciatic notch is age-related and not sex-related. For the ages newborn to ten, the greater sciatic notch is shallow. From the age of ten the greater sciatic notch is deep. Unfortunately, there are too few skeletons in each age group, to do any statistics on each age group.

On the depth of the greater sciatic notch sample, the χ^2 test, the ϕ test and the logistic regression analysis were conducted. The χ^2 test indicates that this method cannot be used to correctly determine the sex of subadults ($\chi^2(1) = 0.118$; $n = 37$; $p = 0.732$) since the p-value is above the value of 0.0042. The ϕ test agrees with this of which can be concluded that there is no association between the method and the sex ($\phi = 0.056$). The logistic regression analysis indicate that the depth of the greater sciatic notch cannot be used to determine the sex of subadults (χ^2 (1) = 0.117, $p = 0.732$). Both the statistics and the accuracy rates indicate that the depth of the greater sciatic notch cannot be used to determine the sex of subadults.

4.2.4 Arch criterion

For the arch criterion 37 skeletons could be used. The combined accuracy rate is 67.5% (table 39, next page). This is the highest accuracy rate of all the accuracy rates for the pelvis. The arch criterion performs well on both males and females. The accuracy rate is lower than the accuracy rates obtained by

Schutkowski (1993) and Sutter (2003), but the present study scores better than the three observers from Cardoso and Saunders (2008).

Table 39. The accuracy rates of Schutkowski (1993), Sutter (2003), Cardoso and Saunders (2008) and the present study divided in female, male and combined for the arch criterion. The number of skeletons is indicated with N.

Arch criterion	Female		Male		Combined	
	N	%	N	%	N	%
Present study	25	72%	12	58.3%	37	67.5%
Schutkowski (1993)	17	60%	27	81.5%	44	72.3%
Sutter (2003)	25	85.7%	37	76.9%	62	81.5%
Cardoso and Saunders (2008) 1	40	40%	57	50.9%	97	46.4%
Cardoso and Saunders (2008) 2	40	50%	57	31.6%	97	39.2%
Cardoso and Saunders (2008) 3	40	60%	57	26.3%	97	40.2%

The accuracy rates can be further divided in age groups which are visible in table 40 on the next page. The accuracy rates are compared to the rates of Schutkowski (1993) and Sutter (2003). The present study shows that the arch criterion can be used for the age groups of newborn to one year and two to five years. The method performs well on both males and females within these two age groups. The method continue to perform well on the females. They have accuracy rates of 57.1% for the age group six to ten years, 66.7% for the age group eleven to fifteen years and 100% for the age group. The method performs worse on males. The age group six to ten years has an accuracy of 50% and the age group eleven to fifteen even has an accuracy of 0%. The last age group, sixteen to twenty-one years has a high accuracy of 75%. It is difficult to compare the results of the arch criterion with Schutkowski (1993) and Sutter (2003), since their accuracy rates of this method differ much from the present study. It seems that the arch criterion is effective within the first two age groups and the last age group. Unfortunately, there are too few skeletons in each age group to be able to conduct statistical analyses.

On the arch criterion a χ^2 test, a ϕ test and a logistic regression analysis were conducted. For one cell the expected frequency was below five which means

Table 40. The accuracy rates divided in five age groups and divided by male and female for the arch criterion. The number of skeletons for each age group is indicated with N.

Arch criterion		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	3	66.7%	1	100%	4	75%
	Schutkowski (1993)	10	70%	11	82%	21	76.2%
	Sutter (2003)	7	71.4%	6	50%	13	61.5%
2 - 5 years	Present study	6	66.7%	1	100%	7	71.4%
	Schutkowski (1993)	10	50%	16	81%	26	69.2%
	Sutter (2003)	7	100%	7	100%	14	100%
6 - 10 years	Present study	7	57.1%	2	50%	9	55.6%
	Sutter (2003)	4	25%	10	100%	14	78.6%
11 - 15 years	Present study	3	66.7%	3	0%	6	33.3%
	Sutter (2003)	4	25%	15	100%	19	84.2%
16 - 21 years	Present study	5	100%	4	75%	9	88.9%

that a Fisher's exact test had to be done. This test indicate that this method cannot be used to determine the sex of subadutls ($\chi^2(1) = 3.172$; $n = 37$; $p = 0.146$) since the p-value is above 0.0042. The ϕ test agrees with this, since this test indicates that there is only a weak association ($\phi = 0.293$) between the sex and the method. The logistic regression analysis turned out to be non-significant (χ^2 (1) = 3.133, p = 0.077) which indicates that this method cannot be used to determine the sex of subadults. The accuracy rates indicates that the arch criterion can be used for the ages newborn two to five years and from sixteen to twenty-one years old. The statistics indicate that overall the arch criterion cannot be used to accurately determine the sex of subadults.

4.2.5 Iliac crest

A total of 34 skeletons could be used for the iliac crest method. The accuracy rate is 44.1% which is below chance (table 41, next page). The method performed low on females who had an accuracy rate of only 34.8%. The iliac crest performed better on males who had a rate of 63.6% .Schutkowski (1993), the original author, did not have high accuracy rates for this method himself and neither did Sutter (2003) when he retested the method. Sutter (2003) had a high accuracy rate for the females (92.9%), but a low accuracy rate for the males (38.5%). It is striking that both Schutkowski and Sutter had high rates for females and low rates for males. In this study it is the opposite.

Table 41. The accuracy rates of Schutkowski (1993, Sutter (2003) and this study divided in female, male and combined for the iliac crest. The number of skeletons is indicated with N.

Iliac crest	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	23	34.8%	11	63.6%	34	44.1%
Schutkowski (1993)	17	85.7%	27	54.2%	44	68.9%
Sutter (2003)	25	92.9%	37	38.5%	62	66.6%

The accuracy rates for the iliac crest can be further divided in five age groups and compared to the accuracy rates of Schutkowski (1993) and Sutter (2003). These rates are visible in table 42, next page. The accuracy rates for all age groups are consistently lower than the rates obtained by the other authors. The combined accuracy rate for the newborn to one year olds (50%) is very close to what Schutkowski (1993) (60%) and Sutter (2003) (46.1%) obtained. The age group two to five years has an accuracy rate of 33.3%, which is much lower than the rates of Schutkowski (1993) (73.1%) and Sutter (2003) (85.7%). The accuracy rates of the other age groups are very low. The group six to ten years has an accuracy of 25%, the group of eleven to fifteen years has an accuracy of 50% and the last age group has an accuracy of 62.5%. When looking at how the iliac crest performed on both males and females, it becomes clear that there is not one age group that is interesting for both sexes. The age group sixteen to twenty-one years has a high accuracy rate for males (100%), but the method did not perform well on females who had an accuracy rate of 40%. It seems that this method cannot be used to determine the sex of subadults.

Three statistical tests were conducted on this method namely the χ^2 test, the ϕ test and the logistic regression analysis. For a χ^2 test every cell needs to have an expected frequency of five. For the iliac crest, one cell has an expected

Table 42. The accuracy rates divided in five age groups and divided by male and female for the iliac crest. For each age group the total number of skeletons (N) is given.

Iliac crest		Female		Male		Combined	
		N	%	N	%	N	%
NB - 1 year	Present study	3	66.7%	1	0%	4	50%
	Schutkowski (1993)	10	80%	10	40%	20	60%
	Sutter (2003)	7	85.7%	6	0%	13	46.1%
2 - 5 years	Present study	5	40%	1	0%	6	33.3%
	Schutkowski (1993)	12	83.8%	14	64%	26	73.1%
	Sutter (2003)	7	100%	7	71%	14	85.7%
6 - 10 years	Present study	6	16.7%	2	50%	8	25%
	Sutter (2003)	3	33.3%	10	80%	13	69.2%
11 - 15 years	Present study	3	33.3%	3	66.7%	6	50%
	Sutter (2003)	7	28.6%	16	88%	23	69.6%
16 - 21 years	Present study	5	40%	3	100%	8	62.5%

frequency below five. Therefore, a Fisher's exact test was conducted instead of a χ^2. This test indicate that the iliac crest cannot be used to accurately determine the sex of subadults ($\chi^2(1) = 0.008$; n = 34; p = 1.000) since the p-value is above 0.0042. The ϕ test also indicates that there is no association between the method and the sex ($\phi = 0.015$). The logistic regression analysis turned out to be non-significant (χ^2 (1) = 0.008, p = 0.928) which indicates that the iliac crest cannot determine the sex of subadults. Both the statistics and the accuracy rates indicate that the iliac crest cannot be used to determine the sex of subadults.

4.3 Results for the skull methods

In this paragraph the results of the seven methods of the skull will be presented. Both the accuracy rates of this study as of the other authors discussed in chapter 2 will be presented. In addition, statistical analyses will be used to determine whether the observed rates are not mere chance. The rates will be further divided by sex and five age groups.

4.3.1 Chin prominence

A total of 30 skeletons could be used for this method. The overall accuracy rate for this method is 46.7% (table 43). The chin prominence did not work well on females who only had an accuracy rate of 38.1%. Males had a higher rate of 66.7%. The combined accuracy rate of the present study is much lower than the accuracy rates for Schutkowski (1993)(70%) or Sutter (2003)(83.1%). The accuracy rate of the females does not come near the accuracy rates obtained by Schutkowski (1993) or Sutter (2003). The accuracy rate of the male comes close to Schutkowski (1993), but Sutter (2003) had a much higher rate namely of 93.9%. The chin prominence cannot be used to determine the sex of subadults.

Table 43. The accuracy rates of Schutkowski (1993), Sutter (2003) and this study divided in female, male and combined for the chin prominence. The number of skeletons is indicated with N.

Chin prominence	Female		Male		Combined	
	N	%	N	%	N	%
Present study	21	38.1%	9	66.7%	30	46.7%
Schutkowski (1993)	17	92.3%	27	59.3%	44	70%
Sutter (2003)	28	64.3%	48	93.9%	76	83.1%

The results for the chin prominence can be further divided in five age groups (table 44, next page). In this case, the age group newborn to one is not available. The method could only be scored when the mandible was fully closed and this happens during the first year of life (Black and Scheuer 2000, 147). Therefore, the method was not scored on any of the first age group mandibles. The table shows us that the chin prominence is not reliable in any of the age groups. The age group two to five years and the age group six to ten years has the worst accuracy rates from all the age groups. The age group two to five years has a combined rate of 28.6%, while the group six to ten years has a rate of 42.9%. The method does perform well on males in the six to ten year group, but the females only have an accuracy rate of 33.3%. This means that the method can be used on males, but not in females which does not help within archaeology or forensics. The other two age groups do seem more promising. The combined accuracy rates of both groups is 60% which is promising. However, the method does not perform well on males in the age group eleven to fifteen years (50%) and does not perform well on females within the sixteen to twenty-one years age group (40%). This indicates, that the chin prominence cannot be used to determine sex. Both Schutkowski (1993) and Sutter (2003) had much higher accuracy rates. Within the age group six to ten years both Sutter and the present study have around the same accuracy rates. In all the other age groups the method does not perform well when compared to both other authors.

Table 44. The accuracy rates divided in five age groups and divided by male and female for the chin prominence. For each age group the total number of skeletons (N) is given.

Chin prominence		Female		Male		Combined	
		N	%	N	%	N	%
2 - 5 years	Present study	6	33.3%	1	0%	7	28.6%
	Schutkowski (1993)	7	85.7%	14	57.1%	21	67.7%
	Sutter (2003)	7	100%	12	71%	19	85.7%
6 - 10 years	Present study	6	33.3%	1	100%	7	42.9%
	Sutter (2003)	5	40%	6	100%	11	72.7%
11 - 15 years	Present study	3	66.7%	2	50%	5	60%
	Sutter (2003)	7	57.1%	15	80%	22	72.7%
16 - 21 years	Present study	5	40%	5	80%	10	60%

On the sample of the chin prominence three statistical tests were conducted namely the χ^2 test, the ϕ test and the logistic regression analysis. Since the expected frequency of one cell was below 5, a Fisher's exact test was conducted instead of a χ^2 test. The Fisher's exact test indicates that the chin prominence cannot be used to determine the sex of subadults ($\chi^2(1) = 0.062$; n = 30; p = 1.000) since the p-value is above 0.0042. The ϕ test agrees with this, since it indicates that there is no association between the method and the sex ($\phi = 0.045$). The logistic regression analysis showed that the chin prominence cannot be used to determine the sex of subadults correctly (χ^2 (1) = 0.062, p = 0.803). Both the accuracy rates and the statistical analysis indicates that the chin prominence cannot be used to correctly determine the sex of subadults.

4.3.2 Anterior dental arcade

For the anterior dental arcade a total of 34 skeletons could be used. An accuracy rate of 53% was obtained (table 45, below). For females, the rate was much higher (69.6%) than for males (18.2%). With such a low accuracy rate for males, it could be questioned whether the method will ever really work on males. The results of Schutkowski (1993) and Sutter (2003) are quite different. Schutkowski obtained a rate of 71.8% and Sutter of 76.6%. Their rates for males are also much higher than the present study has obtained. The present study indicates that this method cannot be used within archaeology of forensics.

Table 45. The accuracy rates divided in five age groups and divided by male and female for the anterior dental arcade. The number of skeletons is indicated with N.

Dental arcade	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	23	69.6%	11	18.2%	34	53%
Schutkowski (1993)	28	69.2%	48	73.1%	76	71.8%
Sutter (2003)	17	53.6%	27	89.8%	44	76.6%

Within table 46 (next page) there is a division visible in the results by age groups. The anterior dental arcade works much better on females than males. Only within the first age group (newborn to one year) females have a low accuracy rate of 0%. Both Schutkowski (1993) and Sutter (2003) had the same problem. It seems that this method cannot be used on individuals below the age of one year.

For the other age groups, the method does work well on females. The age group two to five years has a rate of 66.7%, the group six to ten years has a rate of 60% and the last two age groups, eleven to twenty-one years, even have a rate of 100%. For the males, the anterior dental arcade only works well for the first age group from newborn to one year. Within the other age groups, the accuracy rate decreases. For the age group two to five years there were no male skeletons available. Both the age groups six to ten years and sixteen to twenty-one years have an accuracy rate of 0%. For the age groups eleven to fifteen years the method has a rate of 33.3.% for the males. The anterior dental arcade performed much better on the males within the research of both Schutkowski (1993) and Sutter (2003). The accuracy rates of the present study indicates that the anterior dental arcade cannot be used to determine the sex of subadults.

Table 46. The accuracy rates divided in five age groups and by male and female for the anterior dental arcade. For each age group the total number of skeletons (N) is given.

Anterior dental arcade		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	3	0%	1	100%	4	25%
	Schutkowski (1993)	6	50%	12	67.7%	18	61.1%
	Sutter (2003)	9	11.1%	16	100%	25	68%
2 - 5 years	Present study	6	66.7%	0	N/O	6	66.7%
	Schutkowski (1993)	7	85.7%	14	78.6%	21	81%
	Sutter (2003)	7	57.1%	12	91.7%	19	78.9%
6 - 10 years	Present study	5	60%	2	0%	7	42.9%
	Sutter (2003)	5	80%	6	83.3%	11	81.8%
11 - 15 years	Present study	3	100%	3	33.3%	6	66.7%
	Sutter (2003)	7	85.7%	15	80%	22	81.8%
16 - 21 years	Present study	5	100%	5	0%	10	44.4%

All three statistical tests were conducted on the anterior dental arcade. These tests are the χ^2 test, the ϕ test and the logistic regression analysis. The χ^2 test was replaced by the Fisher's exact test, because one cell had an expected frequency below five. This test indicates that this method cannot be used to determine the sex of subadults ($\chi^2(1) = 0.574$; n = 34; p = 0.682) since the p-value is above 0.0042. The ϕ test indicates that there is no association between the method and the sex ($\phi = 0.130$). The logistic regression analysis indicates that this method is non-significant (χ^2 (1) = 0.600, p = 0.438) and that the anterior dental arcade cannot be used to determine the sex of subadults. Both the accuracy rates and the statistical analyses indicates that the anterior dental arcade cannot be used to determine the sex of subadults.

4.3.3 Eversion of the gonion region

For the eversion of the gonion region a total of 39 skeletons could be used. This yielded an accuracy rate of 56.4% (table 47). This combined accuracy rate is lower than the rates that Schutkowski (1993) and Sutter (2003) obtained. Schutkowski had a rate of 65% and Sutter had an accuracy rate of 81.7% for the eversion of the gonion region. The obtained accuracy rate for females comes close to both authors. The accuracy of the present study is 64%. Schutkowski (1993)

Table 47. The accuracy rates of Schutkowski (1993), Sutter (2003) and the present study divided in female, male and combined for the eversion of the gonion region. The number of skeletons is indicated with N.

Gonion region	Female		Male		Combined	
	N	%	N	%	N	%
Present study	25	64%	14	42.9%	39	56.4%
Schutkowski (1993)	17	60%	27	68%	44	65%
Sutter (2003)	28	63.4%	48	92.1%	76	81.7%

obtained a rate of 60% and Sutter (2003) obtained a rate of 63.4%. The eversion of the gonion region performed less on the males. Within the present study only an accuracy rate of 56.4% was obtained. Schutkowski (1993) had a rate of 68% and Sutter (2003) even had a rate of 92.1%. The present study indicates that the eversion of the gonion region cannot be used to determine the sex of subadults.

When the results are divided in five age groups and compared with Schutkowski (1993) and Sutter (2003), it becomes clear that there is not a single age group for which this method works (table 48, next page). The combined accuracy rate for the age group newborn to one year is only 57.1% and the accuracy rate decreases further to 42.9% for the age group two to five years. The group six to ten years has the higher accuracy rate (66.7%). After this the accuracy rate decreases further to 50% for the group eleven to fifteen years and then rises again to 62.5% for the age group sixteen to twenty-one years. The eversion of the gonion region seems to work on the females much better than on the males. Only the age groups two to five years and eleven to fifteen years, have low accuracy rates for the females (respectively 50% and 33.3%). Sutter (2003) also had low accuracy rates for the age group eleven to fifteen years for the females. His accuracy rate was only 25%. Sutter (2003) has also low rates for the group six to ten years, but within the present research the accuracy rate was 85.7%. This is a large increase compared with Sutter (2003). The eversion of the gonion region when applied to males, does not seem to work. For the age group newborn

Table 48. The accuracy rates divided in five age groups and divided by male and female for the eversion of the gonion region. For each age group the total number of skeletons (N) is given.

Gonion region		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	3	66.7%	4	50%	7	57.1%
	Schutkowski (1993)	7	100%	11	54.6%	18	77.3%
	Sutter (2003)	7	71.4%	6	80%	13	61.5%
2 - 5 years	Present study	6	50%	1	0%	7	42.9%
	Schutkowski (1993)	8	90%	14	82.3%	22	85.2%
	Sutter (2003)	7	100%	7	100%	14	100%
6 - 10 years	Present study	7	85.7%	2	0%	9	66.7%
	Sutter (2003)	4	25%	10	100%	14	78.6%
11 - 15 years	Present study	3	33.3%	3	66.7%	6	50%
	Sutter (2003)	4	25%	15	100%	19	84.2%
16 - 21 years	Present study	5	80%	4	50%	9	62.5%

to one year, the accuracy rate is only 50%. For the age groups two to five years and six to ten years the accuracy rate decreases further to 0%. Only the age group eleven to fifteen years has a high accuracy rate of 66.7%. This decreases again to 50% for the last age group from sixteen to twenty-one years. It seems that the method eversion of the gonion region does not work on males. However, neither within archaeology nor forensics can we use a method to determine sex that can only be used on one of the sexes. This method cannot be used to determine the sex of subadults based on the accuracy rates.

Three statistical analyses were conducted on the sample of the eversion of the gonion region. These three analyses are the χ^2 test, the ϕ test and the logistic regression analysis. The χ^2 test indicates that this method cannot be used to determine the sex of subadults ($\chi^2(1) = 0.178$; $n = 38$; $p = 0.673$) since its p-value is above 0.0042. The ϕ test indicates that there is no association between the method and the sex ($\phi = 0.068$). The logistic regression analysis indicates that the method cannot be used to determine the sex of subadults (χ^2 (1) = 0.177, p = 0.674). Both the accuracy rates and the statistical analyses indicate that the eversion of the gonion region cannot be used to determine the sex of subadults.

4.3.4 Complete mandible

For the complete mandible method 31 skeletons could be used. An accuracy of 46.6% was obtained for this method (table 49). This accuracy rate indicates that this method cannot be used to determine the sex of subadults. The

Table 49. The accuracy rates of Loth and Henneberg (2001), Scheuer (2002), Coqueugniot et al. (2002) and this study divided in female, male and combined for the complete mandible method. The number of skeletons is indicated with N.

Complete mandible	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	21	47.6%	10	40%	31	45.2%
Loth and Henneberg (2001)	12	75%	7	95.2%	19	81%
Scheuer (2002) attempt 1	9	66.7%	27	77.8%	36	75%
Scheuer (2002) attempt 2	9	44.4%	27	85.2%	36	75%
Coqueugniot *et al.* (2002)	40	62.5%	36	41.7%	76	52.6%

The method had an accuracy rate of 47.6% for the females and an even lower accuracy rate of 40% for the males. The accuracy rate for the present study is the lowest that has been obtained for the complete mandible ever. Based on these accuracy rates, the complete mandible method cannot be used to determine the sex of subadults in either forensics nor within archaeology.

The results for the complete mandible can be further divided in age groups and compared with both Loth and Henneberg (2001) and Scheuer (2002) (table 50; below). The results of Coqueugniot *et al.* (2002) cannot be further divided, because the article does not provide enough information for this. In addition, only the first two age groups have information from Loth and Henneberg (2001) and Scheuer (2002). Table 50 shows, that the complete mandible method can be used for the age group newborn to one year and for six to ten years. However, the age group newborn to one year only had one skeleton that could be used. This indicates that the results for this age group are probably not representative for the entire age group. For the age group six to ten years the accuracy rate is 66.7%. When the complete mandible method is applied on females the accuracy rate is

Table 50. The accuracy rates divided in five age groups and divided by male and female for the complete mandible method. For each age group the total number of skeletons (N) is given.

Complete mandible		**Female**		**Male**		**Combined**	
		N	%	N	%	N	%
NB - 1 year	Present study	1	100%	0	N/O	1	100%
	Loth and Henneberg (2001)	5	60%	3	100%	8	75%
	Scheuer (2002)	1	0%	9	78%	10	70%
2 - 5 years	Present study	6	33.3%	1	0%	7	28.6%
	Loth and Henneberg (2001)	7	57%	4	100%	11	72.7%
	Scheuer (2002)	8	60%	16	69%	24	62.5%
6 - 10 years	Present study	5	60%	1	100%	6	66.7%
11 - 15 years	Present study	3	33.3%	3	66.7%	6	50%
16 - 21 years	Present study	5	40%	5	20%	10	30%

only 60%. For males the accuracy rate increases to 100%. It seems that the method can be used for this age group, even though Loth and Henneberg stated that this method cannot be used on subadults over the age of six (2001, 184). For the older age groups, the method is not reliable when viewing the accuracy rates. For the age group eleven to fifteen years the accuracy rate is 50% and for the age group sixteen to twenty-one years the accuracy rate decreases even further to 30%. It seems that the complete mandible method is only reliable for the age group six to ten years and in addition maybe for the age group newborn to one year as well. For all the other age group this method is not reliable based on these accuracy rates.

Three statistical analyses were conducted on the sample for the complete mandible. These three tests are the χ^2 test, the ϕ test and the logistic regression analysis. One of the cells had an expected frequency below five. This means that the χ^2 test cannot be used and that a Fisher's exact test needs to be conducted. This test indicates that this method cannot be used to determine the sex of subadults ($\chi^2(1) = 0.416$; $n = 31$; $p = 0.704$) since its p-value is above 0.0042. The ϕ test indicates that there is no association between the sex and the method ($\phi = 0.116$). The logistic regression analysis shows that the method is non-significant ($\chi^2 (1) = 0.418$, $p = 0.518$) and that the complete mandible method cannot be used to determine the sex of subadults. Both the accuracy rates and the statistical analyses indicate that the complete mandible method cannot be used to determine the sex of subadults.

4.3.5 Mentum

For the mentum method a total of 31 skeletons could be used. The accuracy rates of the present study are visible in table 51 (next page) together with the accuracy rates of Molleson and colleagues (1998). The present study has a higher combined accuracy rate than that Molleson *et al.* (1998) obtained. The mentum performed not too good on males (40%), but it was higher than what was obtained by Molleson and colleagues in 1998 (16.7%). It seems that this method has a higher accuracy rate than the original authors had. Still, the accuracy rates are too low to be used within archaeology or forensics, because an accuracy rate of 65% is considered to be too low to be useful in either of these disciplines (De Vito and Saunders 1990).

Table 51. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the mentum. The number of skeletons is indicated with N.

Mentum	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	22	77.2%	10	40%	32	65.6%
Molleson *et al.* (1998)	6	76.9%	13	16.7%	19	57.9%

The results can be further divided in five age groups (table 52). The article from Molleson *et al.* (1998), did not have enough information to be able to distinguish the five age groups. Dividing in age groups can give information about whether the method works better for a specific age group and whether the method is related to age or sex. The combined accuracy rates indicate that the metum can be used for the ages six to fifteen years. For the age group newborn to one year the accuracy rate was 50%. However, there were no male skeletons that were available for this group and accuracy rate can be different when the males are included as well. For the age group two to five years the accuracy rate is 57.1%. Unfortunately, only one male was available and this might have skewed the accuracy rate. It seems that for the mentum only the age group eleven to fifteen years interesting is.

Table 52. The accuracy rates divided in five age groups and divided by male and female for the mentum. For each age group the total number of skeletons (N) is given.

Mentum	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
NB - 1 year	2	50%	0	N/O	2	50%
2 - 5 years	6	50%	1	100%	7	57.1%
6 - 10 years	5	100%	1	0%	6	83.3%
11 - 15 years	3	100%	3	66.7%	6	83.3%
16 - 21 years	5	100%	5	20%	10	60%

For the statistical analysis three tests were used namely the χ^2 test, the ϕ test and the logistic regression analysis. The χ^2 test could not be used, because one of the cells has an expected frequency below five. This is too low, but a Fisher's exact test can be used instead. This test indicates that the mentum cannot be used to determine the sex of subadults ($\chi^2(1) = 1.015$; $n = 32$; $p = 0.407$), since its p-value is above 0.0042. In addition, the ϕ test indicates that there is no association between the mentum and the sex ($\phi = 0.178$). The logistic regression analysis indicates that the mentum cannot be used to determine the sex of subadults (χ^2 (1) = 0.982, p = 0.322). The accuracy rates indicate that the mentum can be used for the age group eleven to fifteen years. The statistical analyses indicate that overall the mentum cannot be used to determine the sex of subadults.

4.3.6 Mandibular angle

A total of 39 skeletons were available for this method. The accuracy rates are compared with Molleson *et al.* (1998) in table 53. The results of the present study are lower than what Molleson and colleagues obtained. The combined accuracy rate is only 53.8% which is too low to be considered useful in either archaeology or forensics. When the results of the mandibular angle are divided in males and females it becomes clear that these results are lower than the accuracy rates that Molleson and colleagues obtained. The accuracy rates indicate that the mandibular angle is not reliable for determining the sex of subadults.

Table 53. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the mandibular angle. The number of skeletons is indicated with N.

Mandibular angle	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
Present study	24	62.5%	15	40%	39	53.8%
Molleson *et al.* (1998)	6	78.6%	14	50%	20	70%

In addition, the results can be divided in five age groups to see if there is an age group that has more interesting results (table 54, next page). The combined accuracy rates of the mandibular angle are not high. For the age group newborn to one year the accuracy rate is only 42.9%. This accuracy rate increases to 50% for the age group two to five years. For the group six to ten years, the accuracy rate decreases to 33.3%. The only two interesting age groups are the age group of eleven to fifteen years (83.3%) and the group of sixteen to twenty-one years (70%). The mandibular angle does not perform well on males. The age group newborn to one year has a rate of 25%. For the age groups two to five years and six to ten years the accuracy rate decreases to 0%. For the group eleven to fifteen years the accuracy rate is high with 100%. The last age group, sixteen to twenty-one years is only 40%. It seems that the method works much better on females. An accuracy rate of 66.7% was obtained for the age groups newborn to one year, and eleven to fifteen years. The age group two to five years had an accuracy rate of only 60% . The age group sixteen to twenty-one years even obtained an accuracy rate of 100%. The only age group with a rate below chance is the group six to ten years who had a rate of 42.9%. It seems that the mandibular angle cannot be used to determine the sex of subadults.

Table 54. The accuracy rates divided in five age groups and divided by male and female for the mandibular angle. For each age group the total number of skeletons (N) is given.

Mandibular angle	**Female**		**Male**		**Combined**	
	N	%	N	%	N	%
NB - 1 year	3	66.7%	4	25%	7	42.9%
2 - 5 years	5	60%	1	0%	7	50%
6 - 10 years	7	42.9%	2	0%	9	33.3%
11 - 15 years	3	66.7%	3	100%	6	83.3%
16 - 21 years	5	100%	5	40%	10	70%

To see whether the mandibular angle can be used to determine the sex of subadults, three statistical tests were used namely the χ^2 test, the ϕ test and the logistic regression analysis. The χ^2 test indicates that the method cannot be used to determine the sex of subadults ($\chi^2(1) = 0.024$; $n = 39$; $p = 0.876$) since the p-value is above 0.0042. In addition, the ϕ test indicates that there is no association between the method and the sex ($\phi = 0.025$). The logistic regression analysis indicates that the mandibular angle cannot be used to determine the sex of subadults (χ^2 (1) = 0.024, p = 0.876). The accuracy rates indicate that the mandibular angle can be used for the age group eleven to fifteen years. The statistical analyses indicate that, overall, the mandibular angle cannot be used to determine the sex of subadults.

4.3.7 Orbital morphology

For the orbital morphology a total of 21 skeletons could be used. As can be seen in table 55 (next page), the method performed well on males (100%) for Molleson and colleagues (1998) while females had a low accuracy rate of 10%. Within the present study, the accuracy rate is less biased towards the males. The orbital morphology had a combined accuracy of 52.4%. This is barely above chance and not interesting for archaeology or forensics. The method did perform better on females within this study, than the study of Molleson and colleagues (1998). They had a rate of 0%, but the present study had a rate of 35.7%. The accuracy rate of the males within the present study decreased slightly compared with Molleson and colleagues (1998) to 85.7%. Although the accuracy rates indicate that the orbital morphology can be used on males, this does not mean that the method can be used to determine the sex of subadults. Within archaeology nor in forensics, the sex of the skeletons is not known beforehand.

Table 55. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the orbital morphology. The number of skeletons is indicated with N.

Orbital morphology	Female		Male		Combined	
	N	%	N	%	N	%
Present study	14	35.7%	7	85.7%	21	52.4%
Molleson *et al.* (1998)	1	0%	2	100%	3	66.7%

The results for the orbital morphology can be further divided in age groups (table 56). For the age group newborn to one year, no skeletons were available for the orbital morphology. For the age group 2 to 5 years, only females were available. This resulted in an accuracy rate of 100%. Since no male skeletons were available for this age group, it is possible that the accuracy rates are not reliable. Therefore, nothing can be said about this age group. The age group six to ten years has a low accuracy rate of 14.3%. The age groups eleven to fifteen and sixteen to twenty-one years has much higher accuracy rates of 66.7%. Still, this is

Table 56. The accuracy rates divided in five age groups and divided by male and female for the orbital morphology.

Orbital morphology	Female		Male		Combined	
	N	%	N	%	N	%
2 - 5 years	2	100%	0	N/O	2	100%
6 - 10 years	5	0%	2	50%	7	14.3%
11 - 15 years	2	50%	1	100%	3	66.7%
16 - 21 years	5	40%	4	100%	9	66.7%

not high enough to meet the criterion of 75%. It seems that the orbital morphology can be used on males, but not of females. There is not one age group for which this method is interesting.

To see whether the accuracy rates are indicative of anything, three statistical tests were used. These tests are the χ^2 test, the ϕ test and the logistic regression analysis. Because two of the cells had an expected frequency below 5, the Fisher's exact test was used instead of the χ^2 test. This test indicates that the orbital morphology cannot be used to determine the sex of subadults ($\chi^2(1) = 1.050$; $n = 21$; $p = 0.613$).The p-value of this test is above 0.0042 which indicates that this method cannot be used. The ϕ test indicates that there is only a weak association between the method and the sex ($\phi = 0.224$). In addition, the logistic regression analysis indicates that the orbital morphology cannot be used to determine the sex of subadults (χ^2 (1) = 1.137, $p = 0.286$). Both the accuracy rates and the statistical analyses indicate that the orbital morphology cannot be used to determine the sex of subadults.

4.4 Summary

All the methods seem to have a low accuracy rate. The accuracy rates for the pelvis, skull and overall are around 50% which is not interesting for archaeology. The statistics also conclude that the methods cannot be used to correctly determine sex and that there is no association between the sex and the method. For each separate method, the rates do not improve. Three methods score above 60%, the greater sciatic notch angle, the arch criterion and the mentum, but the other methods are between 40 and 55% which is too low to use in archaeology or forensics. All the statistics conclude that the methods cannot be used to correctly determine sex and there is no association between the method and the sex in any of the methods. This is also true for the three rates that were above 60%.

5. Discussion

In the previous chapter the results of the present study have been presented. Within this research four discussion points came to light. The first point is that the accuracy rates and the statistics indicate that the methods cannot be used to determine the sex of subadults. This might be explained by four different causes. First of all, the low accuracy rates can be caused by a difference between the Middenbeemster population and the populations on which the methods were created. The second cause might be that Dutch skeletons are less sexually dimorphic, making it harder to determine the sex of Dutch skeletons. Thirdly, intra- and interobserver errors can cause low accuracy rates. Lastly, diseases, endocrine abnormalities and malnutrition can affect the ability to determine the sex of subadults.

The second discussion point is the influence of the female bias on the results and the accuracy rates. Thirdly, the age groups will be discussed and the possibility that some of the methods are age-related and not sex-related. Lastly, the confounding factors of this research will be discussed.

5.1 Low accuracy rates

Within this study it appeared that all the methods cannot be used to determine the sex of subadults. Both the accuracy rates and the statistical analyses proved this. There might be several reasons to explain these low accuracy rates. These reasons are a population difference, a difference in sexual dimorphism within Dutch populations, inter- and intraobserver error and disease, endocrine abnormalities and malnutrition that might cause problems with normal development.

5.1.1 Population difference

The Middenbeemster collection that has been used for this research might be different from the collections that other researchers have used for their research. For adult samples, it is known that methods can perform perfect on one sample, but that the methods cannot be used on a different sample (Mays and Cox 2000, 119). An example is the Phenice (1969) method that is used for estimating sex. This particular method was created on American adults, but the accuracy dropped significantly when tested on European samples (Maclaughlin and Bruce 1990). To better understand the populations on which the methods have been based, the various collections and their populations will be considered.

All the methods that have been used were created on populations that are similar of the Middenbeemster sample or were created on a diverse collection. The Middenbeemster collection is European which means that collections that are European from origin should work correctly on this sample. Methods created on a diverse collection can often be used on different populations, since the method is based on different populations.

The European collection that have been used by one author is the Spitalfields collection which came from the Spitalfields parish in the UK (Adams and Reeve 1987). This collection was used for the seven methods from Schutkowski (1993). The diverse collections that have been used are the Smithsonian Institution collection for the method of Weaver (1980, 192) and the Dart Collection for the method of Loth and Henneberg (2001). Both collections have skeletons originating from different sources and different races (Dayal *et al.* 2009; Weaver 1980, 192). Molleson and colleagues (1998) have used adult methods for which is known that they can be used on other historical European populations (Acsádi and Nemeskéri 1970, 91).

It seems that all the methods can be used on different populations and this might indicate that the low accuracy rates are not solely caused by problems relating to differences between the population that was used to create the method on and the population used within this research.

5.1.2 Sexual dimorphism in Dutch populations

Another problem can be a difference in the sexual dimorphism of Dutch populations and other populations. To get a better understanding of this, the term sexual dimorphism and its relationship to determining the sex of skeletons will be presented first.

Sexual dimorphism is the difference between males and females and it is the basis for determining sex. Within humans, the hormones are the cause of this sexual dimorphism (Mays and Cox 2000, 117). Males tend to be larger and more robust than females, because they have more muscles, a larger body and the pelvis is adapted to only walking. Females are more small and gracile and the pelvis is different, because of a compromise for females between walking and reproduction (Scheuer and Black 2006, 203). Within the pelvis, skull and long bones certain traits are considered exclusively female or male traits. These traits are used to determine the sex of a skeleton (Ali and Maclaughlin 1991, 57). A high degree of sexual dimorphism will give a more accurate sex determination. However, there are differences in the degree of sexual dimorphism between populations (Mays and Cox 2000, 122). A researcher will probably get accustomed to the manifestation of sexual dimorphism within the population he always works with. It is possible that when dealing with an unknown individual from another population, the past experiences may cause a bias. For example, the mandibles of female Eskimos are very pronounced and their mandibles can be mistakenly sexed as male when compared to another population. Certainly, when the sexing is done by a person who has no experience in sexing Eskimo females (Steele 1988, 53).

Several other studies have been done to show a difference of sexual dimorphism between populations (Walker 2005, 389). Two important studies for this research are the study by Maat and colleagues (1997) and by Maclaughlin and Bruce (1990). Both studies were conducted on Dutch adults and it can be assumed that if there are problems with the adult populations, those were probably already present in the subadult populations. Maat and colleagues (1997) did research into the reliability of sexing Dutch adults with methods for the pelvis and the mandible. They came to the conclusion that mandibles are less reliable as indicators of sex for the Dutch females. Of the total sample, 51.6% of the female mandibles were classified as male. They stated that mandibular methods are less reliable on Dutch populations (Maat *et al.* 1997). For the Dutch pelves, Maclaughlin and Bruce (1990) researched one Dutch collection and multiple English collections. It turned out that the Dutch pelves were less sexually dimorphic for the Phenice (1969) method than the English pelves. Maclaughlin and Bruce concluded that the Dutch pelves are less reliable for determining sex when using the Phenice (1969) method (Maclaughlin and Bruce 1990).

It seems that Dutch mandibles and pelves are less reliable for determining the sex of skeletons than other populations. It might be possible that this is one of the causes of the low accuracy rates within this research. More research is needed to learn more about the sexual dimorphism of the Middenbeemster collection.

5.1.3 Inter- and intraobserver error

Another possibility for the low accuracy rates is inter- and intraobserver error. An interobserver error is an error between two or more researchers which can cause one researcher to perform better than the other one. An intraobserver error is an error between two different sessions by the same observer. This is more likely to happen when the methods are difficult to use. For the twelve methods used within this research, inter- and intraobserver errors are known for the arch criterion (Cardoso and Saunders 2008) and the complete mandible method (Scheuer 2002). Cardoso and Saunders tested the arch criterion for inter- and intraobserver error. They found out that both errors can be easily made for the arch criterion (2008, 26). Scheuer only tested for intraobserver error for the complete mandible method. She found out that intraobserver error exists for this method (2002, 190). Unfortunately, no other methods were tested for inter- and intraobserver errors.

Errors in description of the methods or interpretation of these descriptions can also result in different accuracy rates. Cardoso and Saunders note that Sutter (2003) may have used a different description for the arch criterion and that this could have caused the increase in the accuracy for this method within his research (2008, 27). In addition, Loth and Henneberg mention that even with the description of the anterior dental arcade, they could not distinguish the method in any of the mandibles they used (2001, 180). It seems that inter- and intraobserver error or a fault in the interpretation of the description of the methods may cause a higher or a lower accuracy rate. Neither interobserver or intraobserver error were specifically investigated in this research and additional research is needed to understand if this is the cause of the lower accuracy rates that were encountered here.

5.1.4 Endocrine abnormalities, disease and malnutrition

Endocrine abnormalities, disease or malnutrition can affect normal development including areas that are used to determine the sex of subadults (Loth and Henneberg 2001, 183). Many endocrine diseases are known to affect bone. The most known conditions include vitamin D deficiency, hyperpituitarism, hypopituitarism, hyperthyroidism and hypothyroidism (Roberts and Manchester 2005, 234-249). Many more endocrine diseases are known that can affect the skeletons, but they are less common (Grumbach *et al.* 2003). Within the Middenbeemster collection approximately 20% of the children with were infected with rickets (Veselka 2012, 70). Within this research sample four skeletons have been diagnosed with rickets and one skeletons has possible rickets (Veselka *et al.* 2013). Any other endocrine diseases are not known for this research sample. Rickets created a 'softening' of the bones and this leads to deformation like bending of the bones and orbital porosity (Roberts and Manchester 2005, 237). The five skeletons with rickets did not show rickets in the pelvis, one showed angulation of the mandibular ramus and one showed orbital porosity (Veselka *et al.* 2013). The angulation of the mandibular ramus might cause problems for the method mandibular angle. Therefore, this skeleton was not scored for the mandibular angle. The skeleton with orbital porosity was not scored for the method orbital morphology. Although the other skeletons were not affected with any endocrine abnormalities that were visible, it is possible that there were abnormalities which were not clear yet. It is possible that endocrine abnormalities are one of the reasons of the lower accuracy rates.

Malnutrition can affect normal development as well. From the Spitalfields collection it is known that the children are shorter than their counterparts from the countryside which could indicate that they were less healthy (Lewis 2002). This could have affected their development and the expression of traits to determine sex. Since seven of the methods are based on the Spitalfields Collection (Schutkowski 1993), it is important to keep this in mind. The Chilean mummies that Sutter (2003) used in his research, were also not really healthy. Cribra orbitalia and Harris lines have been found on these mummies in large quantities. In addition, many adults had infectious lesions that gives evidence that they had probably already been present in their childhood (Arriaza 1995, 63-82). Their poor health could have affected their expression of traits.

The complete mandible method from Loth and Henneberg (2001) is based on the Dart collection whereby the health status of each individual is unknown. However, the collection is composed of unclaimed bodies and bequeathed bodies (Dayal *et al.* 2009, 327). Often, the unclaimed bodies are from poor families, are less educated and younger that the bequeathed bodies who often come from families with a high education and who are older with a better health status (Komar and Grivas 2008, 230). Unfortunately, it is unclear whether the subadult skeletons came from unclaimed bodies or were bequeathed. There is a possibility that the method of Loth and Henneberg has been created on a population with developmental problems.

It seems that malnutrition and other health problems might have affected the methods that have been created to determine the sex of subadults. It is known that both poor health and diseases affect the ability to sex skeletons and that it can cause differences in sexual dimorphism (Loth and Henneberg 1996, 481; Mays and Cox 2000, 125; Walker 2005, 390). Children found in cemeteries are the non-survivors. This means that they could not cope with their environment and were probably not healthy enough to reach adulthood (Lewis 2002, 218). The children within the Middenbeemster sample had many indications of stress or diseases and they were probably not really healthy (pers. com. Dr. S.A. Inskip and Prof. Dr. M.L.P. Hoogland). Since the health of a skeleton may affect the amount of sexual dimorphism within a person (Mays and Cox 2000, 125), it is possible that the low accuracy rates obtained within this research were caused by the fact that the children were not healthy. Another problem is that seven of the twelve methods have been created on an archaeological sample (Schutkowski 1993) and it would be interesting to know what the effect is of the health of the children on the expression of the traits used for sex determination.

It seems that poor health and diseases might have a lot of influence on the methods that have been developed to determine the sex of subadults and that both factors might have had influence on the accuracy rates that were encountered within this research. Therefore, more research is needed on the effect of health and expression of methods to determine the sex of skeletons.

5.2 Female bias

Within this sample there is a bias towards females. Of the entire sample, 61.4% is female and 38.6% is male (n = 43). This can cause a problem with the statistics. However, the statistics indicate that there is 17% chance that this spread of males and females in a sample could happen again. This indicates that the sample is not too divergent from a normal sample and that there should be no problem with the statistics.

However, there could be a problem with interpreting the accuracy rates. If a method performs well on one sex and not on the other, the combined accuracy rate may still be high. This is especially the case when there are more females or males than the opposite sex as is the case for this sample. Therefore, the accuracy rates of the age groups should be viewed with caution. In addition, some of the accuracy rates of the original authors and the authors who retested the methods, have these problems as well. Their results should be viewed with caution as well and certainly when there are more skeletons from one sex in the sample. A larger sample with a male:female division that is equal or near equal can fix this problem. Therefore, a larger sample with more males would be recommended for further research on determining the sex of subadults by using the known-sex subadults form Middenbeemster.

5.3 Age groups

Within this research all the accuracy rates have been divided in five age groups namely newborn to one year, two to five years, six to ten years, eleven to fifteen years and sixteen to twenty-one years. It seems that there is not one particular age group that works for all methods. For the elevation of the auricular surface, the greater sciatic notch angle, the arch criterion, the mentum and the mandibular angle, one or more age groups performed better than the others. For the elevation of the auricular surface, the greater sciatic notch angle and the arch criterion, the age group newborn to one year performed adequately. It is possible that this age group performed so well, because between the 16th and 24th week in gestation, a male fetus has testosterone levels comparable to adults. It is thought that this peak causes the sexual dimorphism that might be seen in infants (Cardoso and Saunders 2008, 28). The arch criterion can be used for the sixteen to twenty-one years olds as well, which can be explained by the fact that this age group has almost reached adulthood. For the adults the arc composé is used which is almost similar to the arch criteiron. This method performs well with an average accuracy of 80% within adults (Cardoso and Saunders 2008, 24) and this might explain why this age group has a high accuracy rate. The mentum and mandibular angle perform adequately for the age groups eleven to fifteen years. Since the age group sixteen to twenty-one years does not perform adequately, it is not possible to say that both methods perform better on older age groups. Why the methods performed adequately within this age groups is unknown.

More research and a larger sample is needed to get a better understanding for the relationships between methods and the different age groups.

5.3.1 Age related methods

Based on his research, Hunt (1990) concluded that the elevation of the auricular surface is related to age and not to sex (1990, 884). His conclusion was based on the fact that before the age of 2.5 years the raised auricular surface was 5.6 times more prevalent than the non-raised method. This would mean that before the age of 2.5 there were 5.6 times females than males within the sample. After the age of 2.5 years the non-raised elevation of the auricular surface was 3.5 to 4 times more prevalent indicating that there were 3.5 to 4 times more males than females within the sample (Hunt 1990, 884). If a method is related to sex, a more even spread of males and females is expected. Within this research, there were three methods that had a shift from a high accuracy rate of one sex in the younger age groups to a high accuracy rate of the other sex in older age groups. This was visible with the greater sciatic notch depth, the anterior dental arcade and the orbital morphology. It is possible that these three methods are related to age and not to sex. More research and a larger sample is needed to confirm that these methods are related to age and to see if any of the other methods are related to age and not to sex.

5.4. Confounding factors

This research had its own limitations. The beginning sample was 59 skeletons, but only 43 skeletons could be used in the end. Of these 43 skeletons 26 were female and 17 were female. This small amount of males created one problem. This problem is that there age groups had too few male skeletons and this may have caused problems when interpreting the data of the age groups. The entire sample was too small to be used for some statistics and the results of the other statistics should be interpreted with case, because more skeletons are needed to confirm that the statistics were not caused by chance.

6. Conclusion

The capacity to estimate the sex of subadults is seen as important for osteology. It gives us the ability to better estimate age and it is important for social studies (Baker *et al.* 2005, 48). Within osteology both metric methods and non-metric methods exist to estimate the sex of subadults. Within the present study non-metric methods have been studied on a known-sex sample from the Middenbeemster collection.

The main question of this research was "*Is it possible to accurately determine the sex of subadults below 21 years with non-metric methods on the subadults from the cemetery of Middenbeemster?".* To answer this question a number of sub-questions were posed. The first sub-question was to identify what non-metric methods were available from the literature to determine the sex of subadults. These were the following twelve non-metric methods: the elevation of the auricular surface by Weaver (1980), the greater sciatic notch angle, the greater sciatic notch depth, the arch criterion, the iliac crest, the chin prominence, the anterior dental arcade and the eversion of the gonion region by Schutkowski (1993), the complete mandible method by Loth and Henneberg (2001) and the mentum, mandibular angle and the orbital morphology by Molleson *et al.* (1998).

The second sub-question regarded the accuracy rates of these methods. High accuracy rates were obtained by the original authors and some authors who retested the methods (Loth and Henneberg 2001; Mittler and Sheridan 1992; Molleson *et al.* 1998; Schutkowski 1993; Sutter 2003; Weaver 1980). Often, the accuracy rates dropped when retested by other authors (Cardoso and Saunders 2008; Coqueugniot *et al.* 2002; Scheuer 2002; Vlak *et al.* 2008). These methods were tested and this was used to answer the third sub-question namely what were the accuracy rates of these methods when tested on the sample of Middenbeemster. The accuracy rates were low and the statistical analyses indicate that the methods cannot be used to determine sex.

In addition, it was also studied whether a combination of the methods would give better accuracy rates. However, only three male skeletons were available for the analysis and this is too low to conduct an analysis. Pelvic, skull and overall scores were obtained, but this did not result in higher accuracy rates.

The fifth question was whether a specific age group gave a better result. There were no specific age group for which the methods worked better. The age groups were too small to be useful for statistical analysis. Therefore, no conclusion can be made about this sub-question.

The main question for this research was whether it is possible to accurately determine the sex of subadults below 21 years with non-metric methods on the subadults from the cemetery of Middenbeemster. The results of this research indicate that this is not possible. The accuracy rates are low and the statistical analyses indicate that the methods cannot be used to correctly estimate sex. Combining the methods does not give any better results neither are there better results for certain age groups.

There are some possible explanations for the low accuracy rates within this research. One possible explanation is a difference in the origin of the population. However, the methods are either created on an European collection or based on a variance of populations. Another possible factor is the amount of sexual dimorphism within the Dutch population. A third possible factor is inter- and intraobserver error. The last possible factor is endocrine abnormalities, illnesses and malnutrition. Children within graveyards are non-survivors which means that they were probably not really healthy. This can have effects on the ability to determine sex. More research is needed to be more certain about these factors and their ability to influence the features that are used to estimate sex.

From this research it can be concluded that the twelve non-metric methods cannot be used to determine the sex of subadults.

6.1 Further research

This research has shown that further research is needed in several areas. The first area is sexual dimorphism within Dutch populations. This is known for Dutch adults and more research is needed to see whether this is also the case for the subadult pelves.

Only one non-metric method within the present study was research for inter- and intraobserver and more research is needed to see whether this is also the case for the other methods.

In addition, more research is needed on the influence of subadult health on the ability to sex subadults. It is possible that a poor health may affect the development of the areas that have been used within the present research for determining the sex of subadults. This could cause lower accuracy rates for some populations. More research within this field is needed.

The last area in which more research is required is the relationship between these methods and age. The present research has shown that at least three methods are related to age and probably not to sex. A larger sample could give a more detailed view of this. Some age groups are underrepresented and this could have caused a skewed view. A larger sample and more research would help us understand this better.

Summary

The determining the sex of subadults is seen as important for osteology and social studies. Within osteology several methods exist to determine the sex of subadults, but all these methods are under discussion. Often, the researchers who created the methods had high accuracies, but when it was later on retested by other researchers, the accuracy rates dropped. Within this research twelve non-metric methods were tested on known-sex subadults from Middenbeemster (n = 43). The twelve non-metric methods of this research are the elevation of the auricular surface, the greater sciatic notch angle, the greater sciatic notch depth, the arch criterion, the iliac crest, the protrusion of the chin, the anterior dental arcade, the eversion of the gonion region, the complete mandible, the mentum, the mandibular angle and the orbital morphology. The accuracy rates for each method were calculated and in addition separate accuracy rates were obtained for the pelvis, the skull and the overall skeleton. All the obtained accuracy rates were low. In addition, the statistical analyses indicate that the methods cannot be used to determine the sex of subadults. These low accuracy rates might be explained through with several factors. A population difference between the Middenbeemster sample and the sample of the other researchers might explain the low accuracy rate. In addition, some researchers have proved that there is less sexual dimorphism in Dutch populations than in other populations. Disease and malnutrition can also cause low accuracy rates. And inter- and intraobserver error can also cause low accuracy rates. In addition, this research was limited by a small sample and a overrepresentation of females within the sample. A division in five age groups made clear that there might be several methods that are age-related and not sex-related. It seems that the twelve non-metric method cannot be used to determine the sex of subadults, but more research is needed to get a better understanding of these methods.

Samenvatting

Het bepalen van het geslacht van kinderen wordt gezien als belangrijk voor de osteologie en de sociale wetenschappen. Binnen de osteologie bestaan verschillende methoden om het geslacht van kinderen te bepalen, maar deze methoden liggen allemaal onder vuur. Vaak is het zo dat de auteurs, die de methoden creëren, een goede nauwkeurigheid hebben, maar auteurs, die besluiten de methoden te testen, hebben vaak een veel lagere nauwkeurigheid. Dit onderzoek gaat over twaalf non-metrische methoden, die zijn getest op de kinderen van Middenbeemster (n = 43) waarvan het geslacht bekend is. De twaalf non-metrische methoden, die zijn gebruikt in dit onderzoek zijn de 'elevation of the auricular surface', de 'greater sciatic notch angle', de 'greater sciatic notch depth', 'arch criterion', 'iliac crest', 'protusion of the chin', de 'anterior dental arcade', de 'eversion of the gonion region', de complete kaak, de 'mentum', de 'mandibular angle' en de 'orbital morphology'. De nauwkeurigheid van elke methode is bepaald en ook de nauwkeurigheid voor de bekken, de schedel en het gehele skelet. De nauwkeurigheid van elke methode was laag en de statistiek wees uit dat de methoden niet kunnen worden gebruikt om het geslacht van kinderen te bepalen. Deze lage nauwkeurigheid kan worden verklaard door verschillende oorzaken. Ten eerste is er een bevolkingsverschil tussen de mensen van Middenbeemster en de mensen op wie de methoden zijn gemaakt. Daarnaast hebben een aantal onderzoekers bewezen dat de seksuele dimorfie in de Nederlandse bevolking kleiner is dan in andere bevolkingen. Ziekte en ondervoeding kunnen ook zorgen voor een lagere nauwkeurigheid. Als laatste kunnen inter- en intraobservatie fouten zorgen voor een lagere nauwkeurigheid. Dit onderzoek werd gelimiteerd door het kleine aantal skeletten en een oververtegenwoordiging van vrouwen in het onderzoek. Een verdeling van de resultaten in vijf leeftijdsgroepen laat zien dat een aantal van de methoden waarschijnlijk aan leeftijd zijn gerelateerd en niet aan geslacht. De twaalf non-metrische methoden, die hier zijn onderzocht, kunnen niet worden gebruikt om het geslacht van kinderen te bepalen. Meer onderzoek is nodig om een beter begrip van deze methoden te krijgen.

Bibliography

Acsádi, G. and J. Nemeskéri, 1970. *History of Human Life Span and Mortality*. Translated by K. Balás. Budapest: Akadémiai Kiadó.

Adams, M. and J. Reeve, 1987. Excavations at Christ Church, Spitalfields 1984-86. *Antiquity: A Quarterly Review of Archaeology* 61(2), 247-256.

Ali, R.S. and S.M. Maclaughlin, 1991. Sex identification from the auricular surface of the adult human ilium. *International Journal of Osteoarchaeology* 1(1), 57-61.

Arriaza, B.T., 1995. *Beyond Death: the Chinchorro Mummies of Ancient Chile*. Washington: Smithsonian Institution Press.

Aten, D., 2012. Van water tot werelderfgoed: de droogmaking en wat vooraf ging, in K. Bossaers and C. Misset (eds), *400 jaar Beemster 1612-2012*, Wormer: Stichting uitgeverij Noord-Holland, 10-33.

Bailit, H. and E.E. Hunt, 1964. The sexing of children's skeletons from teeth alone and its genetic implications. *American Journal of Physical Anthropology* 22(1), 171-174.

Baker, B.J., T.L. Dupras, M.W. Tocheri and S.M. Wheeler, 2005. *The Osteology of Infants and Children*. Texas: Texas A&M University Press.

Black, S. and L. Scheuer, 2000. *Developmental juvenile osteology*. London: Elsevier Academic Press.

Black, T.K., 1978. Sexual dimorphism in the tooth-crown diameters of the deciduous teeth. *American Journal of Physical Anthropology* 48, 77-82.

Boucher, B.J., 1957. Sex differences in the foetal pelvis. *American Journal of Physical Anthropology* 15(4), 581-600.

Brooks, S. and J. Suchey, 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution* 5, 227-238.

Bruzek, J., 2002. A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* 117, 157-168.

Bruzek, J. and P. Murail, 2006. Methodology and reliability of sex determination from the skeleton, in A. Schmitt, E. Cunha and J. Pinheiro (eds), *Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death*, Totowa: Humana Press, 225-242.

Cappellini, E., B. Chiarelli, L. Sineo, A. Casoli, A.D. Gioia, C. Vernesi, M.C. Biella and D. Caramelli, 2004. Biomolecular study of the human remains from tomb 5859 in the Etruscan necropolis of Monterozzi, Tarquinia (Viterbo, Italy). *Journal of Archaeological Science* 31, 603-612.

Cardoso, H.F.V., 2008. Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions. *Journal of Archaeological Science* 35, 158-168.

Cardoso, H.F.V. and S.R. Saunders, 2008. Two arch criteria of the ilium for sex determination of immature skeletal remains: a test of their accuracy and an assessment of intra- and inter-observer error. *Forensic Science International* 178, 24-29.

Coqueugniot, H., G. Giacobini and G. Malerba, 2002. L'Utilisation de caractères morphologiques dans la diagnose sexuelle des mandibules d'enfants: application à la collection ostéologique de Turin (Italie). *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 14(1-2), 131-139.

Cunha, E., M.-L. Fily, I. Clisson, A.L. Santos, A.M. Silva, C. Umbelino, P. César, A. Corte-Real, E. Crubézy and B. Ludes, 2000. Children at the convent: comparing historical data, morphology and DNA extracted from ancient tissues for sex diagnosis at Santa Clara-a-Velha (Coimbra, Portugal). *Journal of Archaeological Science* 27, 949-952.

Dawson, C., D. Ross and X. Mallett, 2011. Sex determination, in S. Black and E. Fergusan (eds), *Forensic Anthropology 2000 to 2010*, Boca Raton: Taylor and Francis, 61-94.

Dayal, M.R., A.D.T. Kegley, G. Štrkalj, M.A. Bidmos and K.L. Kuykendall, 2009. The history and composition of the Raymond A. Dart collection of human skeletons at the University of the Witwatersrand, Johannesburg, South Africa. *American Journal of Physical Anthropology* 140, 324-335.

De Vito, C. and S.R. Saunders, 1990. A discriminant function analysis of deciduous teeth to determine sex. *Journal of Forensic Sciences* 35(4), 845-858.

Fearman, M., G.K. Bar-Gal, D. Filon, C.L. Greenblatt, L. Stager, A. Oppenheim and P. Smith, 1998. Determining the sex of infanticide victims from the late Roman Era through ancient DNA analysis. *Journal of Archaeological Science* 25, 861-865.

Field, A., 2009. *Discovering Statistics using SPSS Third Edition (and Sex and Drugs and Rock 'n' Roll)*. Dubai: Oriental Press.

Franklin, D., C.E. Oxnard, P. O'Higgins and I. Dadour, 2007. Sexual dimorphism in the subadult mandible: quantification using geometric morphometrics. *Journal of Forensic Sciences* 52(1), 6-10.

Garn, S.M., P.E. Cole, R.L. Wainwright and K.E. Guire, 1977. Sex discriminatory effectiveness using combinations of permanent teeth. *Journal of Dental Research* 56, 697.

Grumbach, M.M., F.A. Conte and I.A. Hughes, 2003. Disorders of sex differentiation, in P. R. Larsen, H. M. Kronenberg and S. Melmed (eds), *Williams Textbook of Endocrinology*, Pennsylvania: Elsevier Health Sciences, 1303-1425.

Hakvoort, A., M. Sonders, S. de Bruin, S. Lemmers, R. Schats, M. Hoogland and A. Waters-Rist, 2013. *De Begravingen bij de Keyserkerk te Middenbeemster*. Zaandijk: Hollandia archeologen.

Holcomb, S.M.C. and L.W. Konigsberg, 1995. Statistical study of sexual dimorphism in the human fetal sciatic notch. *American Journal of Physical Anthropology* 97, 113-125.

Howell, D.C., 2007. *Statistical Methods for Psychology*. Wadsworth: Cengage Learning.

Humphrey, L.T., 1998. Growth patterns in the modern human skeleton. *American Journal of Physical Anthropology* 105(1), 57-72.

Hunt, D.R., 1990. Sex determination in the subadult ilia: an indirect test of Weaver's nonmetric sexing method. *Journal of Forensic Sciences* 35(4), 881-885.

Hunt, E.E. and I. Gleiser, 1955. The estimation of age and sex of preadolescent children from bones and teeth. *American Journal of Physical Anthropology* 13, 479-487.

Komar, D.A. and C. Grivas, 2008. Manufactured populations: what do contemporary reference skeletons collectinos represent? A comparative study using the Maxwell museum documented collection. *American Journal of Physical Anthropology* 137, 224-233.

Kurpershoek, E., 2012. Gemeenteleden en parochianen: kerken en geloof, in K. Bossaers and C. Misset (eds), *400 jaar Beemster 1612-2012*, Wormer: Stichting uitgeverij Noord-Holland, 206-233.

Lewis, M., 2011. The osteology of infancy and childhood: misconceptions and potential, in M. Lally and A. Moore (eds), *(Re)Thinking the Little Ancestor: New Perspectives on the Archaeology of Infancy and Childhood*, Oxford: Archaeopress, 1-13.

Lewis, M.E., 2002. Impact of industrialization: comparative study of child health in four sites from Medieval and Postmedieval England (A.D. 850-1859). *American Journal of Physical Anthropology* 119, 211-223.

Lewis, M.E., 2007. *The Bioarchaeology of Children: Perspectives from Biological and Forensic Anthropology*. Cambridge: Cambridge University Press.

Loth, S.R. and M. Henneberg, 1996. Mandibular ramus flexure: a new morphological indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology* 99, 473-485.

Loth, S.R. and M. Henneberg, 2000. Gonial eversion: facial architecture, not sex. *HOMO* 51(1), 81-89.

Loth, S.R. and M. Henneberg, 2001. Sexually dimorphic mandibular morphology in the first few years of life. *American Journal of Physical Anthropology* 115, 179-186.

Lovejoy, C.O., R.S. Meindle, T.R. Pryzbeck and R.P. Mensforth, 1985. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology* 68(1), 15-28.

Maat, G.J.R., R.W. Mastwijk and E.A. Van der Velde, 1997. On the reliability of non-metrical morphological sex determination of the skull compared with that of the pelvis in the Low Countries. *International Journal of Osteoarchaeology* 7(6), 575-580.

Maclaughlin, S.M. and M.F. Bruce, 1986. Population variation in sexual dimorphism in the human innominate, in M. Pickford and B. Chiarelli (eds), *Sexual Dimorphism in Living and Fossil Primates*, Firenze: Il Sedicesimo, 121-131.

Maclaughlin, S.M. and M.F. Bruce, 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. *Journal of Forensic Sciences* 35(6), 1384-1392.

Mays, S. and M. Cox, 2000. Sex determination in skeletal remains, in S. Mays and M. Cox (eds), *Human Osteology: In Archaeology and Forensic Science*, Cambridge: Cambridge University Press, 117-130.

Mays, S. and M. Fearman, 2001. Sex identification in some putative infanticide victims from Roman Britain using ancient DNA. *Journal of Archaeological Science* 28, 555-559.

Meyers, L.S., G. Gamst and A.J. Guarin, 2006. *Applied Multivariate Research: Design and Interpretation*. Thousand Oaks: Sage Publications.

Mittler, D.M. and S.G. Sheridan, 1992. Sex determination in subadults using auricular surface morphology: a forensic science perspective. *Journal of Forensic Sciences* 37(4), 1068-1075.

Molleson, T., K. Cruse and S. Mays, 1998. Some sexually dimorphic features of the human juvenile skull and their value in sex determination in immature skeletal remains. *Journal of Archaeological Science* 25, 719-728.

Okazaki, K., 2005. Sex assessment of subadults skeletons based on tooth crown measurements: an examination on the interpopulational variation of sex differences and an application to excavated skeletons. *Anthropological Science (Japanese Series)* 113, 139-159.

Perry, M.A., 2006. Redefining childhood through bioarchaeology: toward an archaeological and biological understanding of children in antiquity. *Archaeological Papers of the American Anthropological Association* 15, 89-111.

Phenice, T.W., 1969. A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology* 30(2), 297-301.

Pietrusewsky, M., 2008. Metric analysis of skeletal remains: methods and applications, in M. A. Katzenberg and S. R. Saunders (eds), *Biological Anthropology of the Human Skeleton*, New Jersey: John Wiley & Sons, 487-532.

Pretorius, E., M. Steyn and Y. Scholtz, 2006. Investigation into the usability of geometric morphometric analysis in assessment of sexual dimorphism. *American Journal of Physical Anthropology* 129(1), 64-70.

Reynolds, E.L., 1945. The bony pelvic girdle in early infancy. A roentgenometric study. *American Journal of Physical Anthropology* 3(4), 321-354.

Roberts, C. and K. Manchester, 2005. *The Archaeology of Disease*. New York: Cornell University Press.

Rösing, F.W., 1983. Sexing immature human skeletons. *Journal of Human Evolution* 12, 149-155.

Rösing, F.W., M. Graw, B. Marré, S. Ritz-Timme, M.A. Rothschild, K. Rötzscher, A. Schmeling, I. Schröder and G. Geserick, 2007. Recommendations for the forensic diagnosis of sex and age from skeletons. *Human Biology* 58, 75-89.

Saunders, S.R., 2008. Juvenile skeletons and growth-related studies, in M. A. Katzenberg and S. R. Saunders (eds), *Biological Anthropology of the Human Skeleton*, New Jersey: John Wiley & Sons, 117-148.

Scheuer, L., 2002. Brief communication: a blind test of mandibular morphology for sexing mandibles in the first few years of life. *American Journal of Physical Anthropology* 119, 189-191.

Scheuer, L. and S. Black, 2006. Osteology, in T. Thompson and S. Black (eds), *Forensic Human Identification: an Introduction*, Boca Raton: CRC Press, 199-220.

Schutkowski, H., 1987. Sex determination of fetal and neonate skeletons by means of discriminant analysis. *INternational Journal of Anthropology* 2(4), 347-352.

Schutkowski, H., 1993. Sex determination of infant and juvenile skeletons: I. morphognostic features. *American Journal of Physical Anthropology* 90, 199-205.

Sheuer, L. and S. Black, 2004. *The Juvenile Skeleton*. London: Elsevier Academic Press.

Steele, D.G., 1988. *The Anatomy and Biology of the Human Skeleton*. Texas: A&M University Press.

Sutter, R.C., 2003. Nonmetric subadult skeletal sexing methods: I. a blind test of the accuracy of eight previously proposed methods using prehistoric known-sex mummies from Northern Chile. *Journal of Forensic Sciences* 48(5), 927-935.

Thomson, A., 1899. The sexual differences of the foetal pelvis. *Journal of Anatomy and Physiology* 33(3), 359-380.

Verneau, R., 1875. *Le Bassin dans les Sexes et dans les Races*. Paris: Librairie J.-B. Bailliere.

Veroni, A., D. Nikitovic and M.A. Schillaci, 2010. Brief communication: sexual dimorphism of the juvenile basicranium. *American Journal of Physical Anthropology* 141(1), 147-151.

Veselka, B., 2012. Rural Rickets: Beemster, a Rural Farming Community in Post-Medieval Netherlands, Faculty of Archaeology, Leiden University, Leiden.

Veselka, B., M.L.P. Hoogland and A.L. Waters-Rist, 2013. Rural rickets: vitamin D deficiency in a Post-Medieval farming community from the Netherlands. *International Journal of Osteoarchaeology*.

Viciano, J., I. Alemán, R. D'Anastasio, L. Capasso and M.C. Botella, 2011. Odontometric sex discrimination in the Herculaneum sample (79 AD, Naples, Italy), with application to juveniles. *American Journal of Physical Anthropology* 145(1), 97-106.

Vlak, D., M. Roksandic and A. Schillaci, 2008. Greater sciatic notch as a sex indicator in juveniles. *American Journal of Physical Anthropology* 137, 309-315.

Waldron, T., M.G. Taylor and D. Rudling, 1999. Sexing of Romano-British baby burials from the Beddingham and Bignor villas. *Sussex Archaeological Collections* 137, 71-79.

Walker, P.L., 2005. Greater sciatic notch morphology: sex, age and population differences. *American Journal of Physical Anthropology* 127, 385-391.

Walker, P.L., 2008. Sexing skulls using discriminant function analysis of visually assessed methods. *American Journal of Physical Anthropology* 136(1), 39-50.

WEA, 1980. Recommendation for age and sex diagnoses of skeletons. *Journal of Human Evolution* 9(7), 517-549.

Weaver, D.S., 1980. Sex differences in the ilia of a known sex and age sample of fetal and infant skeletons. *American Journal of Physical Anthropology* 52, 191-195.

White, T.D. and P.A. Folkens, 2005. *The Human Bone Manual*. London: Elsevier Academic Press.

Wilson, L.A., H.F.V. Cardoso and L.T. Humphrey, 2011. On the reliability of a geometric morphometric approach to sex determination: a blind test of six criteria of the juvenile ilium. *Forensic Science International* 206, 35-42.

Wilson, L.A., N. MacLeod and L.T. Humphrey, 2008. Morphometric criteria for sexing juvenile human skeletons using the ilium. *Journal of Forensic Sciences* 53(2), 269-278.

Żądzińska, E., M. Karasińska, K. Jedrychowska-Dańska, C. Watala and H.W. Witas, 2008. Sex diagnosis of subadults specimens from Medieval Polish archaeological sites: metric analysis of deciduous dentition. *HOMO-Journal of Comparative Human Biology* 59, 175-187.

List of figures

1. Table with the five discriminant functions that Black has created from his study (Black 1978, 80) 11

2. Excavation of Middenbeemster seen from the church top (image taken by T. Vermaas). 50

3. Two female pelves with an elevated auricular surface. On the left S082 V0084 (17 days old) and on the right S344V0730 (21 years old) (images taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; figures created by H. Vermaas). 53

4. Two male pelves with a non-elevated auricular surface. On the left S035 V0031 (2 years old) and on the right S196V0437 (13 years old) (images taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; figures created by H. Vermaas). 53

5. A picture and a drawing of a female greater sciatic notch angle. On the left a picture of S373 V0798 (7 days old) and on the right a drawing (from Schutkowski 1993, 201) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 54

6. A picture and a drawing of a male greater sciatic notch angle. On the left a picture of S035V0031 (2 years old) and on the right a drawing (from Schutkowski 1993, 201) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 54

7. A picture and a drawing of a female greater sciatic notch depth. On the left a picture of S082 V0084 (17 days old) and on the right a drawing (from Schutkowski 1993, 201)(picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 55

8. A picture and a drawing of a male greater sciatic notch depth. On the left a picture of S035V0031 (2 years old) and on the right a drawing (from Schutkowski 1993, 201)(picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 55

9. Two pictures and a drawing of a female arch criterion. On the left a picture of S082V0084 (17 days old) and in the middle a picture of S286V0469 (10 years old). On the right a drawing (from Schutkowski 1993, 201) (pictures taken by T. Vermaas from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 56

10. Two pictures and a drawing of a male arch criterion. On the left a picture of S035V0031 (2 years old) and in the middle a picture of S350V0844 (19 years old). On the right a drawing (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 56

11. Two pictures and a drawing of a female iliac crest. On the left a picture of S165V0242 (2 years old) and on the right a picture of S047V0045 (21 years old). Below a drawing of the iliac crest (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 57

12. Two pictures and a drawing of a male iliac crest. On the left a picture of S364V1517 (6 years old) and on the right a picture of S350V0844 (19 years old). Below a drawing of the iliac crest (from Schutkowski 1993, 201) (pictures taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University; created by H. Vermaas). 57

13. A female protrusion of the chin. On the left a picture of S167V0270 (12 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 58

14. A male protrusion of the chin. On the left a picture of S384V0839 (8 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 58

15. A female anterior dental arcade. On the left a picture of S046V0023 (2 years old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 59

16. A male anterior dental arcade. On the left a picture of S037V0013 (3 months old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 59

17. A female eversion of the gonion region. On the left a picture of S082V0084 (17 days old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 60

18. A male eversion of the gonion region. On the left a picture of S400V0859 (24 days old) and on the right a drawing from Schutkowski (1993, 200) (picture taken by T. Vermaas; from the Laboratory for Human Osteoarchaeology, Leiden University). 60

19. A female and a male complete mandible. On the left a female (S328V0701; 3 years old) and on the right a male (S384V0839; 8 years old) (pictures taken by T. Vermaas; from the Laboratory for Human Osteology, Leiden University). 61

20. A schematic design of the complete mandible method from Loth and Henneberg (2001, 181). 61

21. Two female mandibular angles with on the left a less female mandibular angle and on the right a clear female mandibular angle (Molleson et al. 1998, 721). 62

22. Two male mandibular angles with on the left a less male mandibular angle and on the right a clear male mandibular angle (Molleson et al. 1998, 722). 62

23. Two female mentums with on the left a less female mentum and on the right a clear female mentum (Molleson et al. 1998, 721). 63

24. Two male mentums with on the left a less male mentum and on the right a clear male mentum (Molleson et al. 1998, 722). 63

25. A female orbital morphology (Molleson et al. 1998, 721). 63

26. A male orbital morphology (Molleson et al. 1998, 722). 64

27. An example of a logistic regression analysis with a clear S-shaped function (http://abel.ee.ucla.edu/cyxopt/examples/book/figures/fig-7-1). 67

28. All twelve methods with the amount of skeletons used for each method. 70

List of tables

1. Five functions from De Vito and Saunders with their accuracy rates (1990; 850-853). 11

2. Table with the various accuracies for the elevation of the auricular surface (based on Weaver 1980, 195; Mittler and Sheridan 1992, 1073 and Sutter 2003, 930). 25

3. The accuracies obtained by Sutter for the elevation of the auricular surface divided over four age groups (based on Sutter 2003, 930). 26

4. Table with the accuracy rates for the greater sciatic notch angle (based on Schutkowski 1993, 203; Sutter 2003 and Vlak et al. 2008, 311). 28

5. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311) for the greater sciatic notch angle divided over four age groups. 29

6. The various accuracy rates for the depth of the greater sciatic notch based on Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311). 30

7. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Vlak et al. (2008, 311) for the greater sciatic notch depth divided over four age groups. 30

8. The various accuracy rates for the arch criterion based on Schutkowski (1993, 203), Sutter (2003, 930) and Cardoso and Saunders (2008, 27). 31

9. The accuracies obtained by Schutkowski (1993, 203), Sutter (2003, 930) and Cardoso and Saunders (2008, 27) for the arch criterion divided over four age groups. 32

10. The various accuracy rates for the iliac crest based on Schutkowski (1993, 203) and Sutter (2003, 930). 33

11. The accuracies obtained by Schutkowski (1993, 203) and Sutter (2003, 930) for the iliac crest divided over four age groups. 34

12. The various accuracy rates for the protrusion of the chin based on Schutkowski (1993, 202) and Sutter (2003, 931). 36

13. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the protrusion of the chin divided over four age groups. 36

14. The various accuracy rates for the shape of the anterior dental arcade based on Schutkowski (1993, 202) and Sutter (2003, 931). 37

15. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the anterior dental arcade divided over four age groups. 37

16. The various accuracy rates for the eversion of the gonion region based on Schutkowski (1993, 202) and Sutter (2003, 931). 38

17. The accuracies obtained by Schutkowski (1993, 202) and Sutter (2003, 931) for the eversion of the gonion region divided over four age groups. 39

18. The various accuracy rates for the complete mandible based on Loth and Henneberg (2001, 183), Scheuer (2002, 190) and Coqueugniot et al. (2002, 136). 40

19. The accuracy rates for the complete mandible and the allocation in both session (Scheuer 2002, 190). 41

20. The accuracy rates for the complete mandible divided in two age groups (Loth and Henneberg 2001, 183; Scheuer 2002, 190). 42

21. The accuracy rates of Coqueugniot et al. (2002) for the complete mandible divided in two age groups. 42

22. The accuracy rates for the mandibular angle for the Spitalfields sample (Molleson et al. 1998, 724). 43

23. The accuracy rates for the mandibular angle for the Wharram Percy sample (Molleson et al. 1998, 724). 43

24. The accuracy rates for the mentum for the Spitalfields sample (Molleson et al. 1998, 724). 44

25. The accuracy rates for the mentum for the Wharram Percy sample (Molleson et al. 1998, 724). 45

26. The accuracy rates for the orbital morphology for the Spitalfields sample (Molleson et al. 1998, 724). 45

27. The accuracy rates for the orbital morphology for the Wharram Percy sample (Molleson et al. 1998, 724). 47

28. The accuracy rates for the CFS scores for both the Spitalfields sample as the Wharram Percy sample (Molleson et al. 1998, 724). 47

29. Table with the amount of females and males for each age group that were used within this research. 69

30. The accuracy rates of the pelvis, skull and overall. The number of skeletons is indicated with N. 70

31. The accuracy rates for the pelvis, skull and overall divided in five age groups and by female and male. 71

32. The three rates and their statistical results. 73

33. The accuracy rates of Weaver (1980), Mittler and Sheridan (1992), Sutter (2003) and the present study divided in female, male and combined for the auricular surface. The number of skeletons is indicated with N. 74

34. The accuracy rates divided in five age groups and divided by male and female for the auricular surface. For each age group the total number of skeletons (N) is given. 75

35. The accuracy rates of Schutkowski (1993), Sutter (2003), Vlak et al. (2008) and the present study divided in female, male and combined for the greater sciatic notch angle. The number of skeletons is indicated with N. 76

36. The accuracy rates divided in five age groups and divided by male and female for the greater sciatic notch angle. For each age group the total number of skeletons (N) is given. 77

37. The accuracy rates of Schutkowski (1993), Sutter (2003), Vlak et al. (2008) and this study divided in female, male and combined for the greater sciatic notch depth. The number of skeletons is indicated with N. 78

38. The accuracy rates divided in five age groups and divided by male and female for the greater sciatic notch depth. For each age group the total number of skeletons (N) is given. 79

39. The accuracy rates of Schutkowski (1993), Sutter (2003), Cardoso and Saunders (2008) and the present study divided in female, male and combined for the arch criterion. The number of skeletons is indicated with N. 81

40. The accuracy rates divided in five age groups and divided by male and female for the arch criterion. The number of skeletons for each age group is indicated with N. 82

41. The accuracy rates of Schutkowski (1993, Sutter (2003) and this study divided in female, male and combined for the iliac crest. The number of skeletons is indicated with N. 83

42. The accuracy rates divided in five age groups and divided by male and female for the iliac crest. For each age group the total number of skeletons (N) is given. 84

43. The accuracy rates of Schutkowski (1993), Sutter (2003) and this study divided in female, male and combined for the chin prominence. The number of skeletons is indicated with N. 85

44. The accuracy rates divided in five age groups and divided by male and female for the chin prominence. For each age group the total number of skeletons (N) is given. 86

45. The accuracy rates divided in five age groups and divided by male and female for the anterior dental arcade. The number of skeletons is indicated with N. 87

46. The accuracy rates divided in five age groups and by male and female for the anterior dental arcade. For each age group the total number of skeletons (N) is given. 88

47. The accuracy rates of Schutkowski (1993), Sutter (2003) and the present study divided in female, male and combined for the eversion of the gonion region. The number of skeletons is indicated with N. 89

48. The accuracy rates divided in five age groups and divided by male and female for the eversion of the gonion region. For each age group the total number of skeletons (N) is given. 90

49. The accuracy rates of Loth and Henneberg (2001), Scheuer (2002), Coqueugniot et al. (2002) and this study divided in female, male and combined for the complete mandible method. The number of skeletons is indicated with N. 91

50. The accuracy rates divided in five age groups and divided by male and female for the complete mandible method. For each age group the total number of skeletons (N) is given. 92
51. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the mentum. The number of skeletons is indicated with N. 94
52. The accuracy rates divided in five age groups and divided by male and female for the mentum. For each age group the total number of skeletons (N) is given. 94
53. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the mandibular angle. The number of skeletons is indicated with N. 95
54. The accuracy rates divided in five age groups and divided by male and female for the mandibular angle. For each age group the total number of skeletons (N) is given. 96
55. The accuracy rates of Molleson et al. (1998) and this study divided in female, male and combined for the orbital morphology. The number of skeletons is indicated with N. 97
56. The accuracy rates divided in five age groups and divided by male and female for the orbital morphology. 98

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (1)

Individuals	Age in years	Auricular surface	Greater sciatic notch angle	Greater sciatic notch depth	Arch criterion	Iliac crest	Chin prominence	Dental arcade	Archival sex
MBS018V0102	7	-	-	-	-	-	-	-	F
MBS035V0031	2	M	M	F	M	F	F	-	M
MBS037V0021	0	M	M	F	M	F	-	M	M
MBS046V0023	2	M	M	M	M	M	F	F	F
MBS047V0045	21	F	F	M	F	F	M	F	F
MBS050V0042	0	-	-	-	-	-	-	-	F
MBS062V0071	3	-	M	-	-	-	-	-	M
MBS072V0033	?	-	-	-	-	-	-	-	?
MBS082V0084	0	F	F	F	F	F	-	M	F
MBS089V0115	3	-	-	-	-	-	-	-	F
MBS099V0138	0	-	-	-	-	-	-	-	F
MBS131V0710	16	-	-	-	-	-	-	-	F
MBS145V0238	3	-	-	-	-	-	-	-	M
MBS152V0244	0	-	-	-	-	-	-	-	M
MBS154V0228	0	-	-	-	-	-	-	-	M
MBS158V0230	2	M	F	F	F	M	M	M	F
MBS164V0364	0	-	-	-	-	-	-	-	M
MBS165V0242	2	M	M	M	F	F	M	M	F
MBS167V0270	12	F	M	M	M	M	F	F	F
MBS177V0413	3	-	-	-	-	-	F	F	F

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (2)

Individuals	Gonion region	Complete mandible	Mentum	Mandibular angle	Orbital morphology	Pelvis	Skull	Overall	Archival sex
MBS018V0102	F	-	-	F	M	-	F	F	F
MBS035V0031	F	F	M	F	-	M	F	F	M
MBS037V0021	M	-	-	F	-	M	M	M	M
MBS046V0023	F	F	F	-	-	M	F	F	F
MBS047V0045	F	F	F	F	M	F	F	F	F
MBS050V0042	-	-	-	-	-	-	-	-	F
MBS062V0071	-	-	-	-	-	M	-	M	M
MBS072V0033	-	-	-	-	-	-	-	-	?
MBS082V0084	F	-	F	F	-	F	F	F	F
MBS089V0115	-	-	-	-	-	-	-	-	F
MBS099V0138	-	-	-	-	-	-	-	-	F
MBS131V0710	-	-	-	-	-	-	-	-	F
MBS145V0238	-	-	-	-	-	-	-	-	M
MBS152V0244	F	-	-	F	-	-	F	F	M
MBS154V0228	-	-	-	-	-	-	-	-	M
MBS158V0230	F	M	F	F	F	F	F	F	F
MBS164V0364	F	-	-	M	-	-	I	I	M
MBS165V0242	M	M	M	M	-	M	M	M	F
MBS167V0270	M	M	F	M	M	M	M	M	F
MBS177V0413	F	M	M	M	F	-	F	F	F

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (1)

Individual	Age	Auricular surface	Greater sciatic notch angle	Greater sciatic notch depth	Arch criterion	Iliac crest	Chin prominence	Dental arcade	Archival sex
MBS196V0437	13	M	F	F	F	M	M	F	M
MBS223V0275	17	-	-	-	-	-	-	-	M
MBS227V0297	0	M	F	F	M	M	-	M	F
MBS246V0396	19	F	M	M	M	-	M	F	M
MBS248V0393	9	M	M	M	M	M	F	F	F
MBS252V0443	3	M	F	F	M	M	-	-	F
MBS286V0469	10	M	M	M	F	M	F	F	F
MBS293V0495	4	M	M	F	F	M	M	F	?
MBS307V0591	21	M	M	M	F	F	M	F	F
MBS318V0643	2	-	-	-	-	-	-	-	M
MBS320V0662	0	M	F	F	M	M	-	M	?
MBS328V0701	3	M	F	F	F	F	M	F	F
MBS330V0706	?	-	-	-	-	-	-	-	F
MBS334V0716	9	M	M	M	M	M	-	F	F
MBS336V0709	?	M	F	F	F	M	M	F	F
MBS340V0724	19	M	M	M	F	M	M	F	M
MBS343V0732	3	M	M	F	F	-	M	F	F
MBS344V0730	20	F	M	M	F	M	M	F	F
MBS350V0844	19	M	M	M	M	M	F	F	M
MBS362V0770	?	M	F	M	F	M	F	F	?

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (2)

Individual	Gonion region	Complete mandible	Mentum	Mandibular angle	Orbital morphology	Pelvis	Skull	Overall	Archival sex
MBS196V0437	F	M	M	M	-	F	I	F	M
MBS223V0275	-	-	-	-	-	-	-	-	M
MBS227V0297	F	-	-	M	-	M	M	M	F
MBS246V0396	M	F	F	M	M	M	F	F	M
MBS248V0393	F	F	F	M	-	M	F	M	F
MBS252V0443	-	-	-	-	-	M	-	M	F
MBS286V0469	F	F	F	F	M	M	F	F	F
MBS293V0495	M	M	M	F	M	M	M	M	?
MBS307V0591	F	M	F	F	M	M	F	I	F
MBS318V0643	-	-	-	-	-	-	-	-	M
MBS320V0662	F	M	M	F	-	M	M	M	?
MBS328V0701	M	F	F	F	-	F	F	F	F
MBS330V0706	-	-	-	-	-	-	-	-	F
MBS334V0716	-	-	-	M	M	M	M	M	F
MBS336V0709	M	F	M	M	-	F	M	F	F
MBS340V0724	-	F	F	F	M	M	F	M	M
MBS343V0732	M	M	M	F	-	I	M	M	F
MBS344V0730	F	M	F	F	F	M	F	F	F
MBS350V0844	M	M	F	M	-	M	I	M	M
MBS362V0770	M	-	-	M	-	M	I	M	?

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (1)

Individual	Age	Auricular surface	Greater sciatic notch angle	Greater sciatic notch depth	Arch criterion	Iliac crest	Chin prominence	Dental arcade	Archival sex
MBS364V1517	6	M	F	F	M	M	-	F	M
MBS365V0773	8	M	F	M	F	-	M	M	F
MBS367V0803	11	M	F	F	F	F	M	F	F
MBS373V0798	0	F	F	F	F	F	-	M	F
MBS376V0900	?	-	-	-	-	-	-	-	?
MBS384V0839	8	M	M	F	F	F	M	F	M
MBS388V0952	21	F	M	M	F	M	F	F	F
MBS389V0857	6	M	F	F	F	F	-	-	F
MBS396V0877	8	M	M	M	M	M	M	M	F
MBS400V0859	0	-	-	-	-	-	-	-	M
MBS452V0985	17	M	M	M	F	M	F	F	F
MBS454V0963	21	-	-	-	-	-	M	F	M
MBS462V0987	16	M	M	M	M	M	M	F	M
MBS465V1001	15	M	M	M	F	M	-	M	M
MBS471V1020	9	M	F	F	F	M	M	F	F
MBS488V0000	45	-	-	-	-	-	-	-	-
MBS522V1127	13	M	M	M	F	M	F	F	F
MBS540V1172	?	M	M	F	M	M	-	-	M
MBS549V1181	11	M	M	M	F	F	F	F	M

Appendix A

Overview of all the individuals with their scoring for each method and the archival sex and age (2)

Individual	Gonion region	Complete mandible	Mentum	Mandibular angle	Orbital morphology	Pelvis	Skull	Overall	Archival sex
MBS364V1517	F	-	-	F	M	M	F	F	M
MBS365V0773	F	M	F	M	-	I	M	M	F
MBS367V0803	M	M	F	F	-	F	I	F	F
MBS373V0798	M	F	M	F	-	F	M	F	F
MBS376V0900	F	-	-	F	-	-	F	F	?
MBS384V0839	F	M	F	F	F	F	F	F	M
MBS388V0952	F	M	F	F	F	M	F	F	F
MBS389V0857	-	-	-	-	-	F	-	F	F
MBS396V0877	M	F	F	M	M	M	M	M	F
MBS400V0859	M	-	-	F	-	-	I	I	M
MBS452V0985	M	F	F	F	M	M	F	I	F
MBS454V0963	F	F	F	F	M	-	F	F	M
MBS462V0987	F	F	M	F	M	M	F	M	M
MBS465V1001	M	M	M	M	-	M	M	M	M
MBS471V1020	F	M	F	F	M	F	F	F	F
MBS488V0000	-	-	-	-	-	-	-	-	-
MBS522V1127	F	F	F	F	F	M	F	F	F
MBS540V1172	-	-	-	-	-	M	-	M	M
MBS549V1181	M	F	F	M	M	M	F	I	M

Appendix B

Scoring form

Elevation of the auricular surface	
Number	**Score (female/male)**
MBS018V0102	
MBS035V0031	
MBS037V0021	
MBS046V0023	
MBS047V0045	
MBS050V0042	
MBS062V0071	
MBS072V0033	
MBS082V0084	
MBS089V0115	
MBS099V0138	
MBS131V0710	
MBS145V0238	
MBS152V0244	
MBS154V0228	
MBS158V0230	
MBS164V0364	
MBS165V0242	
MBS167V0270	
MBS177V0413	
MBS196V0437	
MBS223V0275	
MBS227V0297	
MBS246V0396	
MBS248V0393	
MBS252V0443	
MBS286V0469	
MBS293V0495	
MBS307V0591	
MBS318V0643	
MBS320V0662	
MBS328V0701	
MBS330V0706	
MBS334V0716	
MBS336V0709	
MBS340V0724	
MBS343V0732	
MBS344V0730	