Utility of the Sex Assessment Method, DSP, on a Nubian Skeletal Sample

Mathilde Sehested Thormann

Abstract

In this study, the validity of a new method for sex determination, Diagnose Sexuelle Probabiliste (DSP), was tested using ten coxal measurements obtained from 43 ancient Nubian individuals. Using statistical analyses, differences in means between females and males and the discriminant power of each metric variable were tested. The inter-observer agreement between the sex determinations by the author and the DSP2 program was also tested. The sex of 34 individuals was determined using the DSP2 program with a sexing accuracy of 100% according to the sex estimation of the author. Metric variables with significant difference in means between sexes displayed sexing accuracies between 89,7-97,1%, while, metric variables with no significant difference in means between sexes displayed sexing accuracies between 72,2-78,3%. The inter-observer agreement between the results of the author and the DSP2 program were of very good strength. Conclusively, the DSP method is highly recommended to be used in the sex estimation of skeletal remains and is assessed to be an accurate method to be used on ancient Nubian skeletal material as well.

Mathilde Sehested Thormann

MA-student in Near Eastern Archaeology, UCPH

Contact: Mathilde.thormann @sund.ku.dk

Research interests: Human remains, osteoarchaeology, dental anthropology, health/ disease, biomolecular archaeology

صلاحية المنهج (دسب) لتقييم الجنس لبقايا عظام نوبية

في هذه الدراسة أختبرت صلاحية المنهج الجديد (Diagnose Sexuelle Probabiliste), (دسب) لتحديد نوع الجنس باستخدام مقاييس عظام حرقفية (ten coxal measurements) أُخذت من 43 فردا نوبيا أثريا. من خلال تحليل إحصائي تم اختبار الاختلافات بين مقاييس الإناث ومقاييس الذكور والدالة المميزة لكل المتغيرات المترية كما تم اختبار ما يسمى ب(the inter-observer agreement) الذي هو مدى التوافق بين نتائج هذه الدراسة ونتائج البرنامج (دسب2) فيما يتعلق بتحديد الجنس. تم تحديد جنس 34 فردا من بقايا العظام باستخدام البرنامج (دسب2) وذلك بتوافق مع تقييمات الكاتب في نسبة 100% من الحالات. بينت متغيرات مترية لها اختلافات ذات أهمية بين الجنسين دقة تتراوح ما بين 89,7 – 97,1% في تحديد الجنس, بينما بينت متغيرات مترية ليست لها اختلافات ذات أهمية بين الجنسين دقة تتراوح ما بين 72,2 – 78,3% في تحديد الجنس. لقد كان التوافق (the inter-observer agreement) ما بين نتائج الكاتب ونتائج (دسب2) واضحا جدا. ختاما نستطيع أن نستنتج أن استخدام المنهج (دسب) موصى به إلى حد كبير لتقييم نوع الجنس لبقايا عظمية كما أنه قُيِّم كمنهج دقيق يمكن استخدامه لبقايا عظام نوبية أثرية.

Abstract translated by Rasmus Mortensen, MA-student, University of Copenhagen

Introduction

Traditionally, the first step in an osteological analysis is the determination of sex and age-at-death of an individual, gathering key variables for the creation of a demographic profile if the individual is part of a skeletal population. Sexual dimorphism is often more pronounced in the pelvis, which is regarded the most reliable part of the skeleton for sex determination with a sexing accuracy of 98% as reported by Mays and Cox (2000; Brůžek and Murail 2006). Here, the male pelvis is generally narrower and more robust as opposed to the wider and more gracile female pelvis, which reflects the gestation and parturition abilities of females. When reading about the sexual dimorphic traits and features of the os coxae, such as the subpubic angle, greater sciatic notch, ventral arc, subpubic concavity, and preauricular sulcus, one would think that the determination of sex is a rather brief and easy process, where the sex is immediately given away by the pelvic morphology. Unfortunately, to the frustration of many bioarchaeologists and especially less experienced analysts, this is not always the case. Challenging our sometimes-unintended binary perception of the world, individuals do not always display unambiguous female or male traits, and sometimes an individual may display both female and male morphological traits and features. Additionally, fragmented material can complicate the process. In order to overcome such obstacles when determining the sex, new methods and techniques drawing on statistics have been developed over the past two decades and have been proposed to ensure a more accurate and reliable determination of sex. One of such methods and techniques is the focus of this study, in which the applicability of a new technique for sex determination, DSP, will be tested and evaluated using skeletal material from a Nubian collection. This paper is an adapted version of an unpublished exam project in the independent self-study course Archaeological Analysis at the University of Copenhagen (Thormann 2022).

The Scandinavian Joint Expedition to Sudanese Nubia

The bioarcheological interest in Nubia has since its beginning been rooted in the geographical location, where the Sudanese Nile Valley was seen as a natural gateway between southern and northern Africa. The trade routes established in the pre-dynastic period facilitated contact between Mediterranean and Sub-Saharan cultures, and the movement between populations of different genetic backgrounds was of key interest to the early anthropological studies of ancient Nubia (Binder 2019). With the construction of the Aswan High Dam between 1960-1971, the Nile Valley was flooded to a great extent, creating the water reservoir, Lake Nasser, between Aswan and the Second Cataract. With no financial resources required for archaeological salvation projects along the Nile Valley, where archaeological sites and monuments located in the reservoir area would be submerged, Egypt and Sudan reached out to UNESCO, asking for international help. Teams from 21 countries helped excavate more than a third of around 1000 recorded sites. Here, large-scale multi-period collections of human remains, which would later serve as main sources for bioarcheological information of ancient Nubian populations, were uncovered. One of these collections were uncovered by Denmark, Finland, Norway, and Sweden, who financed the Scandinavian Joint Expedition to Sudanese Nu-

bia. The excavations of the expedition were concentrated to the east banks of the Nile Valley, stretching from the Egyptian border and 60 km south, covering districts including Faras, Serra, Debeira, Ashkeit, Sahaba, Wadi Halfa, Abka, and Gamai. For his doctoral thesis, O. Vagn Nielsen conducted thorough examinations of the material and studied metrical and non-metrical anatomical variations between samples of different periods (Vagn Nielsen 1970). Today, the collection is curated at the Laboratory of Biological Anthropology at the University of Copenhagen.

The Nubian skeletal material

For this project, the skeletal collection from the Scandinavian Joint Expedition to Sudanese Nubia was used with the courtesy of the Laboratory of Biological Anthropology at the University of Copenhagen. The complete collection excavated during the Scandinavian Joint Expedition comprises 1546 individuals from seven chronological periods spanning from 3000 BCE to around 1100-1200 CE. In chronological order, these are the so-called A-group, C-group, Pharaonic, Meroitic, X-group, Christian, and Muhammadan (Islamic) periods, with the majority (n=653, 42,2%) of the individuals belonging to the C -group (Vagn Nielsen 1970). For this project, a total of 43 individuals with preserved os coxae were analyzed and measured. Four of these individuals were excavated in undated contexts, and the remaining 39 individuals dated to the C-group (n=4), Meroitic (n=18), Pharaonic (n=16), and X-group (n=1) periods.

Methods: Diagnose Sexuelle Probabiliste (DSP)

In 2005, Murail et al. developed the Diagnose Sexuelle Probabiliste (DSP) method, based on the pattern of sexual dimorphism observed world-wide in the os coxae of males and females. Using 17 metric variables of the os coxae, the examination of 2.040 individuals of known sex from 12 metapopulation samples formed the basis of the development of the method. By identifying a "*common sexual dimorphism pattern among modern human populations with the help of discriminant analyses*", ten out of the 17 initial metric variables were selected, based on their discriminant power and their preservation rate, ensuring the applicability of the DSP method on skeletal populations from all over the world (Tab. 1, Fig. 1) (Murail et al. 2005, 169). The first eight variables in the table were selected because of their discriminant power and are ranked accordingly with PUM being the most discriminant variable and SA being the least discriminant variable. The last two variables, SIS and VEAC, were selected because of their preservation rate in archaeological contexts and should only be used as part of the DSP method if only three or less of the other variables (PUM, SPU, DCOX, IIMT, ISMM, SCOX, SS, and SA) are available for proper measurement. In 2017, an article introducing a new software program facilitating the DSP method, DSP2, was published, using the same previous sample and two new samples to test the software (Brůžek et al. 2017). Results showed that the application of DSP2 could determine 90.98% of the total number of individuals with a 99,65% accuracy, if the eight variables of high discriminant power (PUM, SPU, DCOX, IIMT, ISMM, SCOX, SS, and SA) were used. Even the usage of the four most discriminating variables could determine 87,17%

of the total number of individuals with a 99,53% accuracy. The usages of the four least discriminating variables could only determine 41,49% of the individuals, however with an accuracy of 98,67%.

In this project, the individuals from the Nubian sample were first examined and their sex determined by the author. All bones were used in the determination of sex, but features of the os coxae, which were assessed based on the methods of Buikstra and Ubelaker (1994), were the overruling factors in the final sex determination. Second, the ten ox coxae measurements previously presented were obtained for each individual with a sliding caliper, pelvimeter or an osteometric board and entered into the DSP2

Abbreviation	Name of measurement	Reference
PUM (M14)	Acetabulo-symphyseal pubic length	Bräuer 1998
SPU	Cotylo-pubic breadth	Gaillard 1960
DCOX (M1)	Maximum pelvic height	Bräuer 1998
IIMT (M15.1)	Depth of the great sciatic notch	Bräuer 1998
ISMM	Post-acetabular ischium length	Schulter-Ellis et al. 1983
SCOX (M12)	Iliac breadth	Bräuer 1998
SS	Spino-sciatic length	Gaillard 1960
SA	Spino-auricular length	Gaillard 1960
SIS (M14.1)	Cotylo-sciatic breadth	Bräuer 1998
VEAC (M22)	Vertical acetabular diameter	Bräuer 1998

Table 1:
List of the ten coxal measurements used in the DSP method with description and original reference.

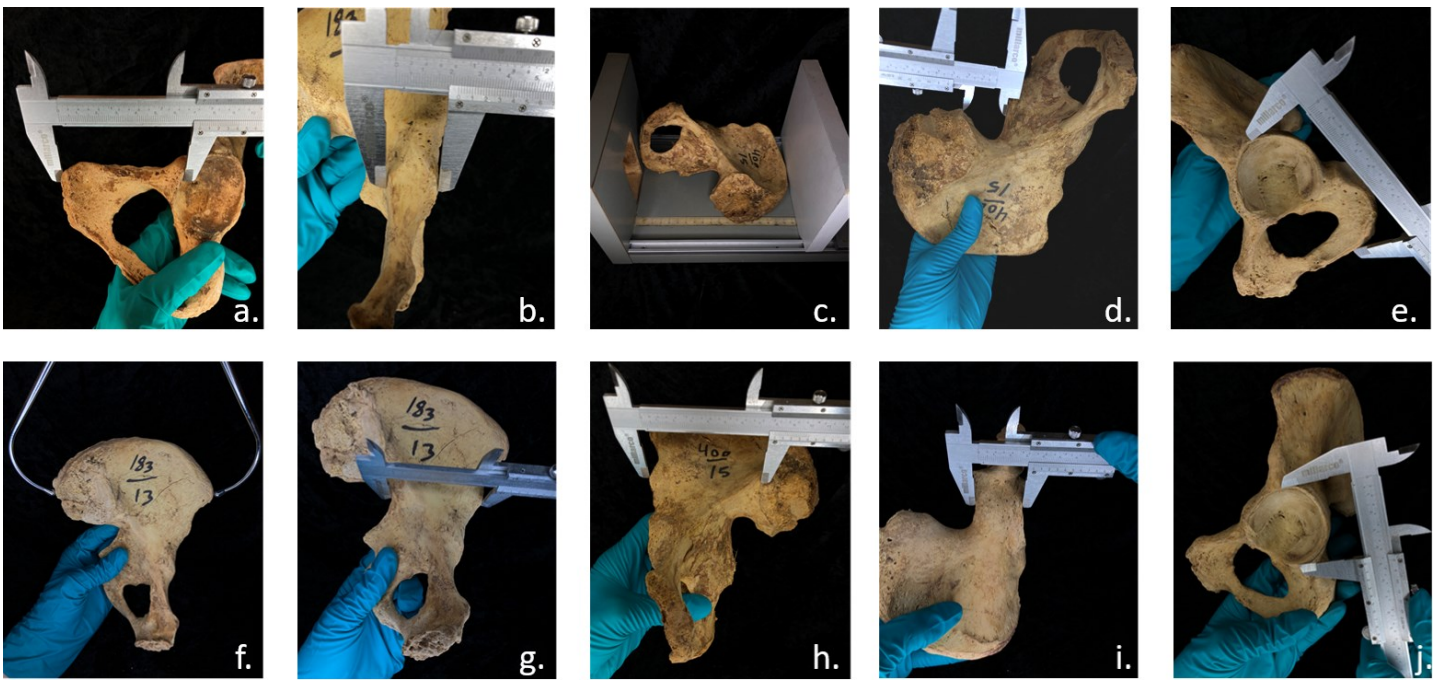

Figure 1: Ten coxal measurements used in the DSP method. From top left: a. PUM, b. SPU, c. DCOX, d. IIMT, e. ISMM, f. SCOX, g. SS, h. SA, i. SIS, and j. VEAC. Photos © Mathilde Sehested Thormann, from Thormann 2022.

program. Having run the datasets of the os coxae metric variables through the DSP2 program, the program will determine the sex of the individuals based on Fisher's linear discriminant analysis. Additionally, the program would calculate the posterior probability of the individual being male (PM=1, PF=0) or female (PF=1, PM=0), based on the prior probability that a metric observation would indicate either male or female (0,5). Sex could not be determined reliably if the posterior probability was below the threshold of 0,95, indicating "N/A" as a result instead of either "Male" or "Female".

For data analysis, the parametric independent samples t-test and binary logistic regression as well as the non-parametric Cohen's Kappa test were conducted. The independent samples t-test was used to detect potential statistically significant differences in means and variances between the sexes for each metric variable, and the binary logistic regression was used to test if sex can be determined based on the measurement of each metric variable, thus testing the discriminant power of each metric variables using predictions. The Cohen's Kappa test was used to measure the inter-observer agreement between the sex determinations by the author and the DSP2 program. All statistical analyses were conducted using the SPSS (v. 28) program.

Results

Out of 43 individuals, the sex of 34 individuals could be determined and was determined correctly in accordance with the subjective sex estimation of the author. The sex of nine individuals could not be determined and were labeled as N/A, because the posterior probability was less than 0,95. In average, females were sexed with 99,81% accuracy, while males were sexed with 99,18% accuracy (Tab. 2).

Independent samples t-test

The differences in the distribution of PUM, SPU, DCOX, IIMT, ISMM, SCOX, SS, and SA between males and females are illustrated as boxplots (Fig. 2a-j). Considering the different distribution patterns of the os coxae measurements, there is a clear difference between the distribution of some measurements observed between males and females (SPU, DCOX, IIMT, ISMM, and VEAC), where other measurements display no clear difference in distribution (PUM, SCOX, SS, SA, and SIS). These observations of clear differences between the distribution of some measurements observed between males and females versus other measurements were tested to see if they were significant, using an independent samples t-test. Results showed that SPU and VEAC display p-values under 0,05 in the Levene's Test for Equality of Variances, which means that equal variances in SPU and VEAC are assumed, whereas the remaining metric variables displayed p-values over 0,05, which means that equal variances are not assumed. Looking at the results from the t-test of equality of means, where the p-value is read based on the results from Levene's Test for Equality of Variances, several of the variables displayed p-values under 0,05, including SPU, DCOX, IIMT, ISMM, and VEAC. In this case, it is not assumed that males and females have equal means. The same cannot be said about PUM, SCOX, SS, SA, and SIS, where a p-valued over 0,05 suggests that males and females have equal means (Tab. 3). The t-test for equality of means

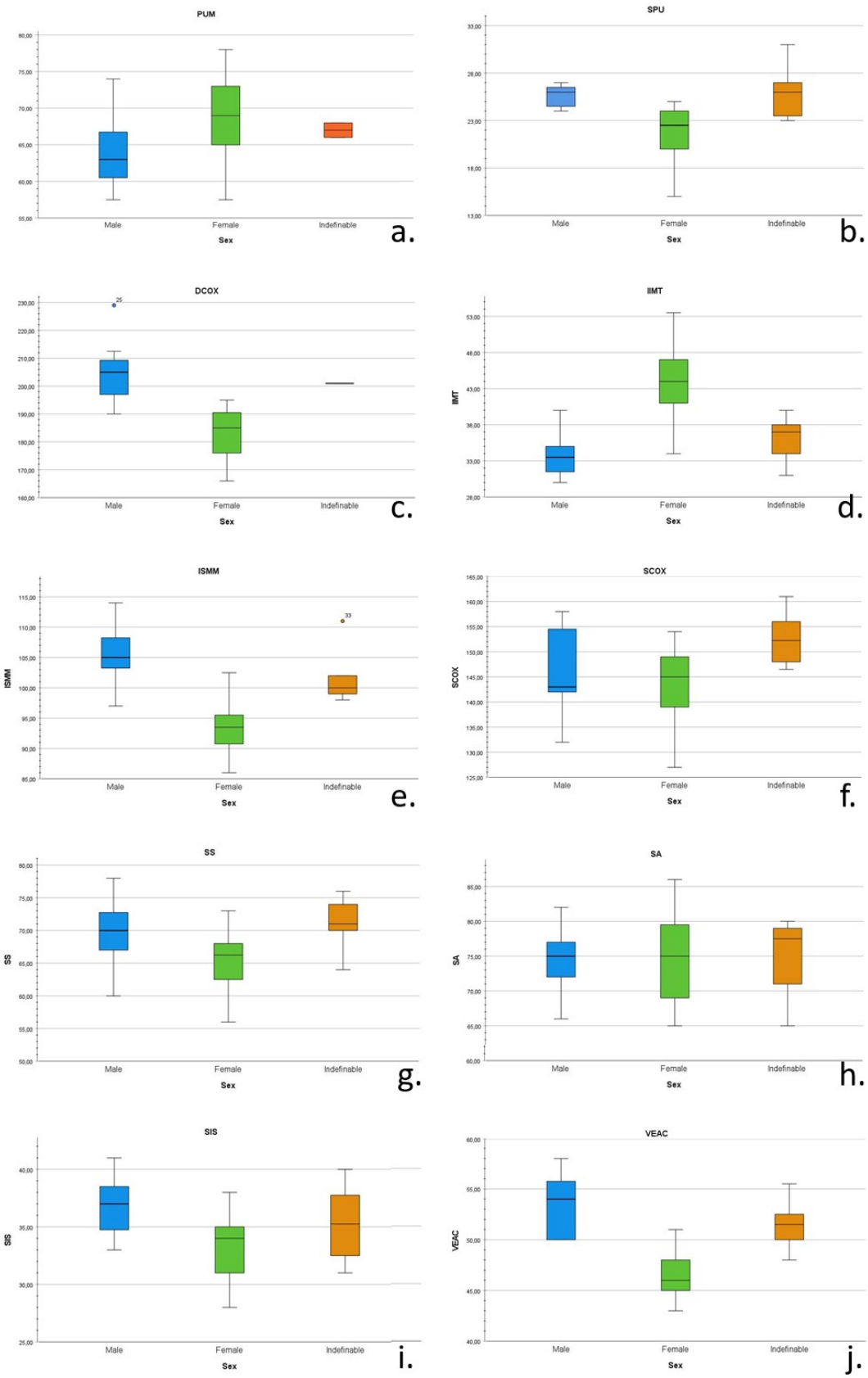

Figure 2: Boxplots displaying the distribution of the coxal measurements observed in males and females of the Nubian sample. From top left: a. PUM, b. SPU, c. DCOX, d. IIMT, e. ISMM, f. SCOX, g. SS, h. SA, i. SIS, and j. VEAC. From Thormann 2022.

SITE	GRAVE	PERIOD	Sex (Thormann 2022)	Sex (Vagn Nielsen 1970)	PUM	SPU	DCOX	IIMT	ISMM	SCOX	SS	SA	SIS	VEAC	P_F	P_M	DSP2 Result
183	13	undat.	Female	Female	64,50	21,00	170,00	44,00	91,00	132,00	64,00	74,50	34,00	47,00	1	0	Female
400	11B	Pharaonic	Female	Female	73,00	25,00	189,00	44,50	93,00	141,50	63,00	78,00	33,50	46,00	1	0	Female
246	20	C-group	Female	Female	-	25,00	195,00	38,00	102,50	154,00	68,00	84,00	34,00	51,00	0.986	0.014	Female
246	26	C-group	Female	Female	-	-	-	41,00	-	149,00	66,00	69,00	-	46,00	0.989	0.011	Female
183	8	undat.	Male	Male	-	-	-	37,00	-	156,00	67,00	77,50	-	-	0.746	0.254	N/A
185	637	Pharaonic	Female	Male	70,00	21,00	-	34,00	98,00	-	67,00	71,00	32,50	47,50	0.999	0.001	Female
220	35:03	C-group	Female	Female	-	-	178,00	42,00	92,00	141,00	64,00	77,00	35,50	46,50	1	0	Female
246	17	C-group	Female	N/A	-	23,50	-	33,00	-	148,00	70,00	79,00	34,50	52,00	0.775	0.225	N/A
401	5	undat.	Female	Male	65,00	24,00	185,00	39,00	95,00	145,00	71,00	83,00	38,00	47,00	0.997	0.003	Female
185	646	Pharaonic	Male	Male	-	-	-	35,00	-	-	64,00	65,00	31,00	48,00	0.864	0.136	N/A
185	672	Pharaonic	Female	Female	-	-	-	34,00	-	148,00	68,00	69,00	34,00	44,00	0.983	0.017	Female
185	675	Pharaonic	Male	Male	-	-	-	37,00	-	153,50	76,00	80,00	34,00	50,00	0.753	0.247	N/A
185	687	Pharaonic	Male	Male	-	-	-	30,00	-	-	72,50	74,00	37,00	53,00	0.049	0.951	Male
185	687N	Pharaonic	Female	N/A	66,00	23,00	176,00	48,00	90,50	-	68,00	74,00	37,00	45,00	1	0	Female
185	692	Pharaonic	Female	N/A	-	23,00	-	-	98,00	-	71,00	70,00	31,00	50,00	0.927	0.073	N/A
400	17b	Pharaonic	Female	Female	-	24,00	185,00	-	95,00	-	68,00	84,00	35,00	48,00	1	0	Female
400	15	Pharaonic	Male	Male	59,00	24,50	195,00	33,00	103,00	142,00	69,00	79,00	38,00	55,00	0.001	0.999	Male
400	17a	Pharaonic	Male	Male	-	26,00	-	-	102,00	-	71,50	79,00	36,00	52,00	0.948	0.052	N/A
400	16	Pharaonic	Female	N/A	-	22,00	193,00	46,00	94,00	-	73,00	86,00	37,00	46,00	1	0	Female
400	12	Pharaonic	Female	Female	-	25,00	-	53,50	93,00	150,00	68,00	78,00	**67,00**	46,00	1	0	Female
400	11	Pharaonic	Male	Male	66,00	31,00	-	39,00	100,00	151,00	70,00	71,00	36,50	53,00	0.062	0.938	N/A
400	18b	Pharaonic	Female	Female	72,50	24,00	190,00	47,50	95,00	153,00	**98,50**	79,50	34,50	46,00	1	0	Female
400	18a	Pharaonic	Female	Female	75,00	23,00	-	47,00	94,00	152,00	68,00	83,00	34,00	48,00	1	0	Female
250	30	Meroitic	Male	Male	63,00	26,00	199,00	40,00	105,00	143,00	69,00	76,00	33,00	50,00	0.005	0.995	Male
250	31	Meroitic	Male	Male	74,00	24,50	229,00	34,00	114,00	158,00	78,00	76,00	41,00	58,00	0.001	0.999	Male
250	28 (1 of 2)	Meroitic	Female	Female	65,00	18,00	176,00	36,00	93,00	145,00	62,50	71,50	28,00	48,00	1	0	Female
250	28 (2 of 2)	Meroitic	Female	Female	78,00	24,00	193,00	43,00	96,00	-	67,00	83,00	33,00	49,00	1	0	Female
250	33 (1 of 3)	Meroitic	Female	Female	68,00	21,00	177,50	43,00	87,00	-	64,00	76,00	31,00	43,50	1	0	Female
250	33 (2 of 3)	Meroitic	Male	Male	68,00	27,00	201,00	31,00	99,00	146,50	74,00	77,50	39,00	51,00	0.072	0.928	N/A
250	33 (3 of 3)	Meroitic	Male	Male	62,00	27,00	206,00	35,00	103,50	-	73,00	72,00	39,00	50,00	0	1	Male
250	29 (1 of 3	Meroitic	Female	Female	-	19,00	177,00	48,00	**86,00**	139,00	58,00	66,00	30,00	45,00	1	0	Female
250	29 (2 of 3	Meroitic	Female	Female	66,50	20,00	191,00	43,00	95,00	141,00	66,50	75,00	31,00	46,00	1	0	Female
250	29 (3 of 3	Meroitic	Male	Male	-	-	-	40,00	111,00	161,00	74,00	80,00	40,00	55,50	0.112	0.888	N/A
250	21	Meroitic	Female	Female	76,00	25,00	194,00	48,00	96,00	-	72,00	79,00	35,00	46,00	1	0	Female
250	19 (1 of 3)	Meroitic	Male	Female	57,50	24,00	190,00	30,00	97,00	132,00	60,00	66,00	33,00	56,50	0.002	0.998	Male
250	19 (2 of 3)	Meroitic	Male	Male	65,00	26,00	212,50	33,00	106,50	-	65,00	72,00	37,00	50,00	0	1	Male
250	19 (3 of 3)	Meroitic	Male	Male	68,50	27,00	205,00	35,00	110,00	154,50	71,00	82,00	36,50	55,00	0.008	0.992	Male
250	17 (1 of 2)	Meroitic	Female	Female	69,00	**15,00**	175,00	45,00	90,00	128,50	59,00	69,00	28,00	44,00	1	0	Female
250	17 (2 of 2)	Meroitic	Female	Female	62,00	18,00	176,00	44,00	88,00	127,00	60,00	68,00	28,00	44,00	1	0	Female
250	43	Meroitic	Female	Female	72,50	20,00	187,00	45,50	96,00	149,00	64,00	73,00	32,00	47,00	1	0	Female
250	38	Meroitic	Female	Female	-	-	-	36,50	91,00	-	56,50	68,00	28,00	48,00	0.996	0.004	Female
401	1:I	X-group	Female	Female	57,50	18,00	**166,00**	48,00	**86,50**	127,00	56,00	65,00	34,00	43,00	1	0	Female
401	1:II	undat.	Female	N/A	75,00	23,00	188,00	47,00	98,00	145,00	62,00	75,00	33,00	48,00	1	0	Female

Table 2: Data recorded for each individual and results from the DSP2 analysis. PF and PM are the posterior probability of an individual being either sexed as female (F) or male (M) based on the combination of coxal measurements. Data left blank means that a variable was unable to be measured due to fragmentation of os coxa. Data in bold are scored either higher or lower than the DSP2 range variation.

supported the inferences of difference in the SPU, DCOX, IIMT, ISMM, and VEAC metric variables among males and females. To no surprise, the variables with significant difference in means between males and females are the ones with higher discriminant power according to the DSP ranking, whereas the variables with no significant difference in means between males and females are the ones with lower discriminant power according to the DSP ranking. Surprisingly though, is the fact that PUM, which is supposed to be the most discriminant variable, did not display a significant difference in means between males and females, and likewise that VEAC, which is not a discriminating variable but only included because of its preservation rate, did display a significant difference in means (Tab. 3).

Variable	Levene's Test for equality of variances	t-test for equality of means
PUM	*0,849*	*0,059*
SPU	***0,027***	***<0,001***
DCOX	*0,559*	***<0,001***
IIMT	*0,168*	***<0,001***
ISMM	*0,422*	***<0,001***
SCOX	*0,642*	*0,486*
SS	*0,725*	*0,261*
SA	*0,221*	*0,731*
SIS	*0,497*	*0,355*
VEAC	***0,013***	***<0,001***

Table 3: Results from the independent samples t-test. P-values in bold are below the level of significance (0,05) and suggests that equal variances or means between sexes are not assumed.

Binary logistic regression

When conducting a binary logistic regression, b-coefficients are automatically calculated and can be used to calculate the probability of an individual being either female or male based on the measurement of a given variable. If the probability is calculated to be below 0,50 (50%), the baseline model simply predicts that every individual is male based on the measurement, and on the other hand, if the probability was calculated to be above 0,50 (50%), the baseline model simply predicts that every individual is female based on the measurement. Wanting to test how well the new model performs compared to a baseline model, an omnibus test of model coefficients is used to test if the model is significantly better than the baseline model. If the p-value is not below the level of significance of 0,05, the new model is not significantly improved from the baseline model. Similar to the results from the t-test of equality of means, several of the variables displayed p-values under 0,05, including PUM, SPU, DCOX, IIMT, ISMM, and VEAC, meaning that the new models were significantly improved from the baseline models. The same cannot be said about SCOX, SS, SA, and SIS, where p>0,05 means that the new models were not significantly improved from the baseline model. In other words, the discriminant power of each metric variables was tested using predictions, and sex could accurately be determined in between 89,7-97,1% of the cases based respectively on the SPU, DCOX, IIMT, ISMM, and VEAC measurements. On the other hand, sex could accurately be determined in between 72,2-78,3% of the cases based respectively on the PUM, SCOX, SS, SA, and SIS measurements (Tab. 4).

Variable	Baseline model	Omnibus test	Accuracy
PUM	*70,80%*	***0,048***	***75%***
SPU	*75,90%*	***<0,001***	***89,70%***
DCOX	*74,10%*	***<0,001***	***92,60%***
IIMT	*75,80%*	***<0,001***	***90,90%***
ISMM	*77,40%*	***<0,001***	***93,50%***
SCOX	*78,30%*	*0,453*	*78,30%*
SS	*76,50%*	*0,271*	*73,50%*
SA	*76,50%*	*0,72*	*76,50%*
SIS	*75,80%*	*0,371*	*72,20%*
VEAC	*76,50%*	***<0,001***	***97,10%***

Table 4: Results from the binary logistic regression. P-values in bold are below the level of significance (0,05) and means that the sexing accuracy of the new model is significantly improved from the baseline model.

Cohen's Kappa test

One of the pitfalls of accurate sex determination is that the chance for determining correctly is already 50%, and thus the probability of determining sex accurately by chance is also high. Therefore, a Cohen's Kappa test will measure the inter-observer agreement, and thus test the agreement between the results obtained by the author based on visual and qualitative sex determination, and the results obtained by the DSP2 program based on metric and quantitative sex determination, by factoring out agreement which happened due to chance. The level of agreement is measured with the Kappa value and can be assessed as either poor (k<0,20), fair (0,21≤k≤0,40), moderate (0,41≤k≤0,60), good (0,61≤k≤0,80), and very good (0,81≤k≤1,00). Having excluded

results where the sex could not be reliably determined, comparison of the sex determination made by the author and DSP2 program produces a Kappa value of 1,0 which suggests a very good strength of agreement. Similarly, the agreement between the sex determination results obtained by the author and the results obtained by O. Vagn Nielsen back in 1970 was also tested. Here, a Kappa value of 0,837 also suggests a very good strength of agreement albeit not as strong as the agreement between the author and the DSP2 program (Tab. 5).

Observators	Kappa value	N of valid cases	Agreement
Thormann (2022) - DSP2	*1*	*34*	Very good
Vagn Nielsen (1970) – DSP2	*0,757*	*31*	Good
Thormann (2022) – Vagn Nielsen (1970)	*0,837*	*38*	Very good

Table 5: Results from Cohen's Kappa test of inter-observer agreement. Sex estimation results where sex could not be reliably determined have been excluded.

Discussion and conclusions

The sex of 43 individuals from ancient Nubia was determined using the DSP method based on ten different measurements of the os coxae. Through statistical analysis, results confirmed the magnitude of sexual dimorphism as expressed in the os coxae. Differences in the distribution of measurements observed between males and females were significant for the variables with higher discriminant power, whereas variables with lower discriminant power displayed no significant difference in means between males and females. Additionally, statistics were used to test the discriminating power of each metric variable based on the probability of an individual being determined as either male or female based on the measurement of each metric variable. In general, variables of high discriminant power were able to classify between 89,7-97,1% of the results correctly, whereas variables of low discriminant power were able to classify between 72,2-78,3% of the results correctly. As a final statistical test, the inter-observer agreement was measured for the results obtained from sex determination by the author and the DSP2 program. Here, very good strengths of agreement were measured.

All of these results suggests that the DSP method is a very reliable tool for sex determination. The DSP method for sex determination of skeletal material is advantageous for several reasons. It can be applied by everyone, also people with less experience in classifying os coxae into sexes, and it can be applied on fragmented os coxae (to a certain extent) as well.. When sex cannot be reliably determined due to fragmented coxal traits essential to traditional sexing methods, the DSP method can be applied, requiring only a minimum of four traits to be preserved. Another advantage lies in the method's ability to accurately classify individuals as either female or males independently of the geographic location of the sample, because the method was based on a heterogeneous sample consisting of metapopulations from different continents. The success of the method is rooted in the accuracy, which is a result of the high threshold, where sex cannot be determined if the posterior probability is below 0,95. Applying such a high threshold instead of the standard 0,5, limits issues of incorrect classification, but addi-

tionally it creates a high percentage of individuals, where sex cannot be reliably determined (Quatrehomme et al. 2017). Compared to the traditional method of sex determination which relies on the subjective scoring of visual traits and attributes, the DSP method offers an objective alternative, where metric data is recorded based on standardized measurements (Rodriquez et al. 2019; Brůžek et al. 2017).

However, even though the collection of metric data is considered an objective process, the usage of sliding calipers and interpretation of measurement description can lead to mistakes in the measuring process. The main disadvantage of the DSP method can thus be found in the identification of the different measurements and their landmarks on the os coxae, where a misreading of the measurements could lead to the calculation of incorrect results. For instance, the SPU and IIMT measurements are at times quite difficult to obtain correctly, and additionally, the arcuate line and acetabular rim are sometimes difficult to identify properly. As suggested by Rodriguez et al. (2019), the method could be improved by adding new measurements concerning other sexual dimorphic features, such as the subpubic angle and pubic arch. This would help the sex determination of the high percentage of individuals, where sex cannot be determined because of the high threshold for discriminant analysis, where either lack of relevant coxal features or varying degree of sexual dimorphism makes the sexing unreliable.

Future research on the application of the DSP method on ancient Nubian material could include testing of sexing accuracy using different combinations of variables, such as the eight and four most discriminating variables, and the four least discriminating variables. Such tests of different combinations of variables requires a sample of complete os coxae. The Nubian sample used in this study included fragmented os coxae, and thus different combinations of variables were difficult to test. Another element of this study which should be drawn awareness to, is the basis of comparison for accurate sexing. The sex estimation results by the DSP2 program are in this study compared with the sex estimation results by the author because the true sex of the individuals is unknown. In order to get a more accurate picture of the DSP2 program and its ability to determine sex accurately, samples of known sex should be used in these kinds of studies.

But despite its premises and preliminary nature, this study presents one of the few engagements with the Nubian skeletal collection from the Scandinavian Joint Expedition since the extensive studies by O. Vagn Nielsen (1970), and to the extent of the author's knowledge, it is the first of its kind to test the validity of the DSP method on ancient skeletal material from this region. In conclusion, the DSP method and its affiliated software program is highly recommended to be used when estimating the sex of skeletal remains, and additionally, it is assessed to be an accurate method to use on ancient Nubian skeletal material as well.

Bibliography

Binder, Michaela. 2019. “The Role of Physical Anthropology in Nubian Archaeology”. In *Handbook of Ancient Nubia*, edited by Dietrich Raue, 105-127. Berlin: De Gruyter Reference. https://doi-org.ep.fjernadgang.kb.dk/10.1515/9783110420388

Bräuer, Günter. 1988. “Osteometrie”. In *Anthropologie, Handbuch des vergleichenden Biologie des Menschen: Band 1,* edited by Rainer Knußmann, 160-232. New York: Gustav Fischer Verlag.

Brůžek, Jaroslav and Pascal Murail. 2006. “Methodology and reliability of sex determination from the skeleton”. In *Forensic Anthropology and Medicine*, edited by Aurore Schmitt, Eugénia Cunha, and João Pinheiro, 225–242. Totowa: Humana Press Inc. https://doi.org/10.1007/978-1-59745-099-7_9

Brůžek, Jaroslav, Frédéric Santos, Bruno Dutailly, Pascal Murail, & Eugénia Cunha. 2017. “Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology”. *American journal of physical anthropology* 164 (2): 440–449. https://doi.org/10.1002/ajpa.23282

Buikstra, Jane E. and Douglas H. Ubelaker. 1994. *Standards for data collection from human skeletal remains*. *Research series no. 44*. Fayetteville: Arkansas Archeological Survey.

Gaillard, Jean. 1960. “Détermination sexuelle d’un os coxal fragmentaire”. *Bulletins et Mémoires de la Société d’Anthropologie de Paris,* XIe série*, 2:* 255-267. https://doi.org/10.3406/bmsap.1960.1145

Mays, Simon and Margaret Cox. 2000. “Sex determination in skeletal remains”. In Human Osteology in Archaeology and Forensic Science, edited by Margaret Cox and Simon Mays, 117–130. London: Greenwich Medical Media Ltd.

Murail, Pascal, Jaroslav Brůžek, Francis Houët, and Eugénia Cunha. 2005. “DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements”. *Bulletins et Mémoires de la Société d’Anthropologie de Paris* 17: 167–176. https://doi.org/10.4000/bmsap.1157

Quatrehomme, Gérald, Irina Radoman, Luísa Nogueira, Philippe du Jardin, & Véronique Alunni. 2017. “Sex determination using the DSP (probabilistic sex diagnosis) method on the coxal bone: Efficiency of method according to number of available variables”. *Forensic Science International* 272: 190–193. https://doi.org/10.1016/j.forsciint.2016.10.020

Rodriguez Paz, Adrià, Jytte Banner and Chiara Villa. 2019. “Validity of the probabilistic sex diagnosis method (DSP) on 3D CT-scans from modern Danish population”. *La Revue de Médecine Légale* 10 (2): 43-49. https://doi.org/10.1016/j.medleg.2018.08.002.

Schulter-Ellis, Frances P., Dwight J. Schmidt, Lee-Ann C. Hayek, and James Craig. 1983. "Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I". *Journal of Forensic Sciences* 28 (1): 169-180.

Thormann, Mathilde Sehested. 2022. "Validity of the Diagnose Sexuelle Probabiliste (DSP) method on Sudanese Nubians from the Pharaonic period". Unpublished project, University of Copenhagen.

Vagn Nielsen, Ole. 1970. "Human Remains. Metrical and Non-Metrical Anatomical Variations". In *The Scandinavian Joint Expedition to Sudanese Nubia Vol. 9*, edited by Torgny Säve-Söderbergh, 1-139. Oslo: Scandinavian University Books.