Testing the validity of the DSP2 sex assessment method on two European populations

Clara Van Melle[a] BSc, Henrietta Stover[a] MSc, Mark Grabowski[a] PhD, Matteo Borrini[a] PhD, Constantine Eliopoulos[a] PhD

[a] School of Biological and Environmental Sciences, Liverpool John Moores University, United Kingdom

E-mails of authors Clara Van Melle clara.vanmelle@gmail.com; Henrietta Stover vita.heni99@gmail.com; Mark Grabowski M.W.Grabowski@ljmu.ac.uk; Matteo Borrini M.Borrini@ljmu.ac.uk; Constantine Eliopoulos C.Eliopoulos@ljmu.ac.uk

Corresponding author Constantine Eliopoulos, PhD
School of Biological and Environmental Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, United Kingdom
E-mail: C.Eliopoulos@ljmu.ac.uk

Acknowledgements The authors want to thank Dr Kevin Cootes for his technical assistance with the Poulton collection and Dr Fabio Di Vincenzo for providing us access to and technical assistance with the Syracuse collection.

Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ABSTRACT: Sex estimation from human skeletal remains can be achieved using a diverse range of skeletal elements, methodologies, and statistical approaches. While population specificity presents a challenge for most metric sex assessment methods, the os coxae seems to bypass this issue due to its universal pattern of sexual dimorphism. The DSP2 (Diagnose Sexuelle Probabiliste, version 2) is a non-population-specific metric sex estimation method that employs a combination of at least four of ten measurements of the os coxae to estimate sex probabilistically. Despite the method being based on a worldwide meta-database, many populations remain unrepresented. The present study aims to test the DSP2 method on two previously untested European populations. A sample of 51 modern Italian os coxae and a sample of 53 British historical os coxae were used. The validity of the DSP2 method was evaluated through the sexing rate (88.24% for the Italian sample and 86.79% for the British sample) and sexing accuracy (ranging from 95.83% to 100%). The sex estimates were obtained by DSP2 with a 0.95 posterior probability threshold. Intra- and inter-observer error analyses revealed close to excellent reliability for most measurements (with the depth of the great sciatic notch being the only exception). The results of this study support the use of DSP2 on populations not represented in the original datasets, including modern Italian and historical British populations.

Keywords Forensic anthropology, DSP2, metric sex assessment, sex estimation, osteometry, os coxae

Running title Testing DSP2 on Italian and British Skeletal Samples

Highlights

- The DSP2 method was validated in both modern and historical populations.
- DSP2 achieved high sexing accuracy and robust sexing rates.
- Highest Italian sexing accuracy with the “all ten” and “best-four” combination.
- Highest British sexing accuracy with the “worst-four” combination.
- Only one measurement exceeded the 5% rTEM threshold for intra/inter-observer error.

1. INTRODUCTION

Building a biological profile is the primary task for practitioners and researchers when presented with skeletonised human remains in both forensic and bioarchaeological contexts.[1, 2] In recent years, there has been a growing interest in developing metric methods to estimate various components of the biological profile, as they have the potential to yield more robust and unbiased results than traditional morphological methods.[2-4] Sex estimation is the first and arguably most crucial step of the profile, as it informs the rest of the examination; thus, an accurate assessment is paramount.[1, 5] In forensic investigations, it is particularly important to employ methods that adhere to the highest standards of scientific rigour to ensure admissibility in court.[1, 6] In bioarchaeological contexts, sex estimation allows researchers to reconstruct the demography of past populations.[1, 7-9]

The os coxae is the most consistently sexually dimorphic element of the skeleton.[9, 10] The female pelvis is constructed to facilitate parturition, whereas the male pelvis is not.[9-11] The universal nature of the os coxae enabled the development of metric standards applicable for estimating the sex of individuals of unknown origin.[10] The *Diagnose Sexuelle Probabiliste* (DSP), or probabilistic sex diagnosis, is a metric sex assessment method developed by Murail et al.[11] and improved by Brůžek et al. (DSP2)[9], utilising ten measurements of the os coxae to estimate sex probabilistically. The software, based on Fisher's linear discriminant analysis, calculates the probability of an individual being male or female from a combination of os coxae measurements by comparing them against a reference database of 2040 os coxae from Europe, Africa, North America, and Asia.

Since its development, the DSP method has been validated by several studies from populations worldwide.[12-20] Numerous studies have also tested the application of DSP to CT images,[2, 21-23] and they all reported high accuracy and reliability.

This study aims to evaluate the validity of the DSP2 method and software for estimating sex in the modern Syracuse population from Italy and the historical Poulton Collection from the United Kingdom. It does so by estimating intra- and inter-observer reliability, calculating the sexing rate and accuracy of the DSP2 method, and comparing these results with previous validation studies.

2. MATERIALS AND METHODS

2.1. Samples

The samples used in the present study are derived from a modern Italian collection from Syracuse, Sicily, housed at the Department of Anthropology and Ethnology of the Natural History Museum of Florence, Italy, and the Poulton Collection of British historical remains, housed at Liverpool John Moores University, UK.

The Sicilian remains all belong to individuals who died between 1900 and 1902 and were excavated from a cemetery in 1909. It is considered a reference collection with known sex, age at death, and exact date of death for all individuals, while cause of death and occupation are also documented in most cases.[24] The individuals from the Poulton Collection were excavated as part of the Poulton Research Project from the Poulton Chapel, located in the rural village of Poulton, west Cheshire, in the United Kingdom. The Chapel is part of a multi-period site from the 10th to the 15th century, which sits at the border with Wales.[25, 26] This is not a reference collection, meaning that there are no records for those originally buried on the Chapel's grounds. However, sex and approximate age at death have been assessed by

LJMU staff and/or postgraduate researchers using common anthropological methods and compiled into catalogues used for this study.

The os coxae of a total of 51 adult individuals, 35 males and 16 females, from the Syracuse Collection were included in this validation study. The os coxae of 53 adults, 25 males and 28 females, from the Poulton Collection were also used.

2.2. DSP2 software and posterior probability calculations

For this study, the graphical user interface of the DSP2 method was employed.[27] The posterior probability of an individual being male or female (respectively PM or PF) is calculated using the Mahalanobis distances between the individual's measurements and the mean vectors for females (DF) and males (DM) in the meta-database.[9, 28] Sex estimates are then assigned based on the posterior probability of the individual being female: if PF ≥ 0.95, the estimate will be female; if PF ≤ 0.05, it will be male; and if neither condition is satisfied, it will be undetermined.[27]

2.3. DSP2 measurements

The measurements used in this study are the ten outlined for the DSP2 software. These measurements were selected by the developers of the *Diagnose Sexuelle Probabiliste* as they provide good coverage of the three morpho-functional areas of the os coxae, namely the sacro-iliac, acetabular, and ischio-pubic sections.[9] The abbreviation, full name, definitions, and references for each measurement are provided in Table 1.

Table 1 *Abbreviations and definitions for the DSP2 measurements taken on the os coxae.*[9]

Abbreviation	Measurement	Definition	Reference
PUM	Acetabulo-symphyseal pubic length	Minimum distance from the superior and medial point of the pubic symphysis to the nearest point on the acetabular rim at the level of the lunate surface	M 14,[29]
SPU	Cotylo-pubic breadth	Pubic breadth between the most lateral acetabular point and the medial aspect of the pubis. Measurement is perpendicular to the major axis of the os pubis. Arms of the sliding caliper are thus parallel to the plane of the obturator foramen	[30]
DCOX	Maximum pelvic height	Maximum height of os coxae measured from the inferior border of the os coxae to the most superior portion of the iliac crest. Can be taken with sliding calipers or osteometric board	M 1,[29]
IIMT	Depth of the great sciatic notch	Distance from the postero-inferior iliac spine (defined as the point of intersection between the auricular surface and the posterior portion of the sciatic notch) to anterior border of the great sciatic notch. Axis of the measurement must be perpendicular to the anterior border. Because of the configuration of hip bone, it is easier to use small arms of sliding caliper	M 15.1,[29]
ISMM	Post-acetabular ischium length	Distance from the most anterior and inferior point of the ischial tuberosity to the furthest point on the acetabular border	[31]
SCOX	Iliac breadth	Distance between the anterior-superior iliac spine and the postero-superior iliac spine	M 12,[29]
SS	Spino-sciatic length	Minimum distance between the antero-inferior iliac spine and the deepest point in the greater sciatic notch	[30]
SA	Spino-auricular length	Distance between the antero-inferior iliac spine and the auricular point. Auricular point is defined as the intersection of the arcuate line with the auricular surface	[30]
SIS	Cotylo-sciatic breadth	Distance between the lateral border of the acetabulum and the midpoint of the anterior portion of the great sciatic notch. Fixed arm of the sliding caliper is parallel to the acetabular plane	M 14.1,[29]
VEAC	Vertical acetabular diameter	Maximum vertical diameter of the acetabulum, measured on the acetabular rim, as a prolongation of the longitudinal axis of the ischium	M 22,[29]

2.4. Data collection

All ten measurements from the DSP2 sex estimation method were taken according to Brůžek et al.[9] using the specified equipment: an osteometric board, sliding calipers, or spreading calipers. In most cases, all ten measurements were available; however, in some instances, one to five measurements could not be measured due to postmortem damage to the landmark, which would have resulted in inaccurate measurements. These were recorded as missing data. To assess intra-observer reliability, one observer (HS) recorded all ten measurements for the entire sample size across three separate sessions for each collection. For interobserver reliability, the ossa coxae of 16 randomly chosen cases from both collections were measured independently by second observers.

Throughout all rounds of data collection, the observer(s) remained unaware of the individual's sex to minimise confirmation bias. Only after all the data had been collected were the catalogues consulted to ascertain the sexes of the individuals selected for this study. All collected data were processed using the DSP2 software to estimate sex once the entire data collection was completed.

2.5. Statistical analyses

Descriptive statistics were computed for each variable and sex category across both samples. The entire dataset was tested for normality and homogeneity of variance. To determine whether there were statistically significant differences between male and female measurements for each variable, independent sample t-tests (or Mann-Whitney U-tests) were employed. All tests were performed with a 95% confidence interval.

To assess how effectively DSP2 performed on these two samples, the sexing rate and sexing accuracy were calculated. Sexing rate is the percentage of individuals classified as either male or female, excluding those classified as undetermined. Sexing accuracy is the percentage of individuals sexed by DSP2 who were correctly assessed, where the DSP2 estimate matches the individual's true sex in the Syracuse population and the catalogued sex in the Poulton population. These metrics were calculated for the full sample and by sex within each population. Additionally, they were calculated for the estimates obtained using all available DSP2 variables, the best four variables (PUM, SPU, DCOX, IIMT) and the worst four variables (SS, SA, SIS, VEAC). These combinations of variables are used in numerous validation studies[13, 15, 19, 20, 22, 32] and are recommended by Murail et al.[11] and Brůžek et al.[9] in their original articles. The terms "best" and "worst" were given to these combinations by the creators of the DSP method, not because of the accuracy they yield, but rather their ability to estimate sex, i.e. the percentage of sexed individuals they produce.[9, 11, 32]

Intra- and inter-observer error rates were estimated using the Technical Error of Measurement (TEM), the relative Technical Error of Measurement (rTEM), and the reliability coefficient (R). An rTEM of 5% was adopted as the threshold for acceptable error.[33, 34] All calculations were performed in Microsoft Excel (version 16.97.2) and IBM SPSS (version 27.0).

3. RESULTS

The sexing rate and sexing accuracy results using all available variables are presented in Tables 2 and 3 (see Supplementary Material for results of the other combinations tested). When all available DSP2 variables were used, the Syracuse

sample showed higher overall sexing accuracy (100% vs. 97.83%) and higher male sexing accuracy (100% vs. 95.83%), while female sexing accuracy was 100% in both samples. The pattern observed between the two samples was the same for the sexing accuracy obtained with the best-four variable combination (PUM, SPU, DCOX, IIMT). However, when using the worst-four variable combination (SS, SA, SIS, VEAC), the Poulton sample yielded 100% sexing accuracy for the entire sample and for males and females separately. In the Syracuse sample, this combination resulted in an overall sexing accuracy of 78.95%, a male sexing accuracy of 20%, and a female sexing accuracy of 100%. It is also noteworthy that one catalogued male from the Poulton Collection was misclassified as female by DSP2 with all available variables and with the best four variables. The worst four variables resulted in an undetermined classification.

When all available variables were used, DSP2 yielded sexing rates of 88.24% in the Syracuse sample and 86.79% in the Poulton sample. Regarding sexing accuracy, the Syracuse sample maintained a high sexing rate with the best-four combination and performed markedly worse with the worst-four combination (82.35% vs. 37.25%). However, the Poulton sample yielded the same sexing rate for both variable combinations (43.40%). Additionally, for the Poulton sample, the sexing rate for females was consistently lower than that for males.

The results for intra- and inter-observer error are in the Supplementary Material. In the Syracuse sample, none of the variables had intra- or inter-observer rTEM exceeding 5%. However, for both intra- and inter-observer reliability, the SPU variable yielded the highest rTEM (1.72% and 2.15%, respectively). In the Poulton sample, IIMT resulted in an inter-observer rTEM of 8.91%. This was the only value exceeding 5%. The worst-performing variable for intra-observer reliability in the

Poulton sample was again SPU, at 1.54%. For both populations and both intra- and inter-observer reliability, the best-performing variable was DCOX.

Tables 4 and 5 present descriptive statistics for all 10 variables. Independent sample t-tests were performed to assess whether there was a statistically significant difference between the sexes for each variable. All but the SCOX (p = 0.146) variable in the Syracuse sample exhibited statistically significant differences between sexes. In the Poulton sample, three variables, namely PUM (p = 0.920), IIMT (p = 0.441), and SA (p = 0.450), showed no statistically significant differences between male and female measurements. In the Syracuse sample, IIMT (p = 0.045), SS (p = 0.014), and SIS (p = 0.040) were not normally distributed, and the variances of SS (p = 0.024) were not equal between the sexes. In the Poulton sample, PUM (p = 0.023) was not normally distributed, and the variances of SCOX (p = 0.038) were not equal between sexes.

4. DISCUSSION

This study assessed the validity of DSP2 for estimating sex in two populations not represented in the original reference dataset, using the sexing rate and sexing accuracy, as defined above and widely used in other studies[13, 14, 19] With all available variables, the Syracuse sample yielded slightly higher sexing rate (88.24%) and sexing accuracy (100%) than the Poulton sample (86.79% and 97.83%, respectively) (Tables 2 and 3). The difference in results between the two populations likely reflects the reduced preservation of the Poulton remains, which led to a higher incidence of missing variables, and the fact that the sex estimates in the Poulton catalogues were based on prior anthropological assessments rather than documented sex.

The Syracuse sample still performed very well with the best-four combination for both sexes and the whole sample, although the female sexing rate was lower than the male's (75% vs. 85.71%), unlike with all available variables. This difference may be accentuated by the disparity in sample sizes between males and females. Overall, the Syracuse sample performed rather poorly with the worst-four combination. Although the female sexing rate and sexing accuracy were best with this combination, and the female sexing rate even surpassed the best-four combination. Male sexing accuracy was 20%, and the male sexing rate was 14.29%, both of which were the study's worst results.

The Poulton sample performed similarly across both the best-four and worst-four combinations, particularly with respect to sexing rate. The sexing accuracy of this sample was highest with the worst-four combination (100% for males, females, and the entire sample). The excellent performance of the worst-four combination in the Poulton sample is likely due to the fact that SS, SA, SIS, and VEAC were the best-preserved variables in this sample. This means that, when using the worst-four combination, there were fewer missing data points than with the other variable combinations. This is in line with the recommendation of Brůžek et al.[9] to use the last two variables (SIS and VEAC) only for "incomplete bones", as they tend to be better preserved but are less diagnostic for sex estimation.

Across validation studies of the DSP (only in Chapman et al.[22]) and DSP2 on populations different from those tested in the present study, reported sexing rates ranged from 71.25%[23] to 97.44%[22]. Variation across studies may reflect differences in preservation rates and/or sample sizes. The Syracuse and Poulton samples both yielded sexing rates within that range.

The lowest, although still reasonable, sexing accuracy (88.34%) was reported by Machado et al.,[13] which may reflect the complexity of mixed populations such as the one tested in this study. In most existing validation studies, the worst-four combination yielded the lowest results, as in the Syracuse sample.[9, 15, 19, 20, 22] These studies were all based on modern collections, from which the Syracuse sample also stems. Nevertheless, Jerković et al.[32] showed a pattern similar to that of the Poulton sample, in which the worst-four combination yielded the highest sexing accuracy. Jerković et al.[32] used a medieval collection, as does the Poulton sample, which likely explains the similarity. However, Jerković et al.[32] yielded higher sexing rates with the best-four combination, whereas the Poulton sample yielded higher sexing accuracy with the same combination.

These findings suggest that the DSP2 method can achieve robust results when applied to modern Italian and historical British populations, and that the main sources of variability across studies are preservation rates and sample heterogeneity rather than population origin per se.

Numerous validation studies have been conducted on DSP and DSP2, generating substantial data on intra- and inter-observer reliability in the literature. However, the statistical methodologies vary, which complicates direct comparison.[20] The most commonly used tests (TEM, rTEM, and R) were adopted in the present study. The most frequently cited cut-off point for acceptable rTEM values is <5%.[35] Only one variable exceeded this threshold: the inter-observer rTEM value for IIMT in the Poulton sample (rTEM = 8.91%). All intra-observer R values from either population were ≥ 0.97, indicating excellent repeatability. Inter-observer R values fell below 0.95 for SPU in the Syracuse sample and for IIMT, SIS, and VEAC in the Poulton sample.

These results align with prior research identifying IIMT as one of the worst-performing variables for intra- and inter-observer reliability across populations and study formats, including virtual measurements.[2, 13-15, 19-21] As outlined by Lesciotto and Klales,[20] the lack of reliability in measuring IIMT may be due in part to the difficulty of positioning the calipers, the need to use the internal jaws of sliding calipers or the discrepancy in definitions of the posterior inferior iliac spine between DSP2 and osteological reference material.[36]

One central claim of DSP2 is that, because it is built on a large, geographically diverse meta-database and focuses on a highly sexually dimorphic element, it can be applied to individuals of unknown ancestry without loss of accuracy.[9] The high sexing accuracy and promising sexing rates observed in both the modern Italian (Syracuse) and historical British (Poulton) samples lend further support to this claim. Importantly, these populations differ not only in modernity and geography but also in preservation state and the amount of information available.

The slightly lower performance in the Poulton sample appears to be driven by preservation-related missing data and the indirect nature of the “true” sex, rather than by any systematic mismatch between the DSP2 model and the skeletal morphology of this historical British group. Indeed, the Poulton collection is not a reference collection; thus, the true sex of the individuals is unknown. Very few, if any, documented collections from the medieval period exist.[32] As a result, for the Poulton sample, this study evaluated the concordance between DSP2 estimates and prior morphological sex estimates recorded in the catalogues, rather than validating DSP2 against known sex. This distinction should be borne in mind when interpreting the high sexing accuracy values in the Poulton sample, as they reflect

agreement with previous assessments rather than necessarily indicating true biological sex.

Despite this limitation, high sexing accuracy can be achieved when morphological traits from both the pelvis and skull,[37] as well as metric methods, are combined, as was the case with the Poulton skeletons. The strong concordance between DSP2 estimates and catalogued sex, therefore, suggests that DSP2 is, at minimum, in good agreement with established morphological protocols.

Another limitation of this study is the small sample size in both populations, driven by the need for predominantly intact os coxae. Fragmented bones constitute the majority of osteological collections, particularly historical ones, making a study such as this difficult.[8, 32] Sex distribution within each sample was uneven, potentially influencing the results. However, within the limitations imposed by the lack of documented sex and the fragmentation present in the Poulton collection, the findings suggest that DSP2 can be used effectively in both modern and historical European contexts, provided pelvic preservation is sufficient to obtain at least four reliable measurements, as required by the method.

The authors recommend integrating osteometric methods with ancient DNA (aDNA) analysis when working with “undocumented” historical collections. This interdisciplinary strategy would improve the accuracy of validation studies, enable more robust testing of DSP2’s population independence, and facilitate the creation of known-sex reference collections for periods when such collections are rare.[32, 38] Robust contextual data, combined with metric analysis, would increase both the research validity and the forensic applicability of skeletal assessments.[32, 38] However, destructive analyses should be considered carefully when working with already fragmentary historical collections.[32]

5. CONCLUSION

DSP2 showed high accuracy and acceptable reliability across a modern Italian and a British historical collection, neither of which was represented in the original meta-database. Despite differences in preservation and documentation, sexing and accuracy rates matched or exceeded previous DSP/DSP2 validation studies. These results support DSP2 as a population-independent method for forensic and bioarchaeological sex estimation.

The study highlights the importance of preservation quality, measurement protocol, and reference data (known vs estimated sex) in interpreting performance metrics. Future work should validate against genetically sexed samples and expand testing across different times and locations. Accurate sex estimation is critical to the biological profile and must be reliable. Validation studies are essential for new methods, and DSP2 offers a robust, probabilistic framework to improve reliability and transparency in sex estimation.

References

1. Klales AR. Sex estimation of the human skeleton: history, methods, and emerging techniques. 2020.
2. Mestekova S, Bruzek J, Veleminska J, et al. A test of the DSP sexing method on CT images from a modern French sample. *Journal of forensic sciences* 2015; 60: 1295-1299.
3. Walker PL. Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 2008; 136: 39-50.

4. Kiskira C, Eliopoulos C, Vanna V, et al. Biometric sex assessment from the femur and tibia in a modern Greek population. *Legal Medicine* 2022; 59: 102126.
5. Scheuer L. Application of osteology to forensic medicine. *Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists* 2002; 15: 297-312.
6. Melnick RL. A Daubert motion: a legal strategy to exclude essential scientific evidence in toxic tort litigation. *American journal of public health* 2005; 95: S30-S34.
7. Klales AR, Ousley SD and Vollner JM. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American journal of physical anthropology* 2012; 149: 104-114.
8. Agarwal SC and Wesp JK. *Exploring sex and gender in bioarchaeology*. University of New Mexico Press, 2017.
9. Brůžek J, Santos F, Dutailly B, et al. 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* 2017; 164: 440-449.
10. Steyn M and Patriquin M. Osteometric sex determination from the pelvis—does population specificity matter? *Forensic science international* 2009; 191: 113. e111-113. e115.
11. Murail P, Brůžek J, Houët F, et al. DSP: a tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et mémoires de la Société d'Anthropologie de Paris BMSAP* 2005; 17: 167-176.

12. Thormann MS. Utility of the Sex Assessment Method, DSP, on a Nubian Skeletal Sample. *Chronolog* 2023; 1.
13. Machado MPS, Costa ST, Freire AR, et al. Application and validation of Diagnose Sexuelle Probabiliste V2 tool in a miscegenated population. *Forensic science international* 2018; 290: 351. e351-351. e355.
14. Kranioti EF, Šťovíčková L, Karell MA, et al. Sex estimation of os coxae using DSP2 software: a validation study of a Greek sample. *Forensic science international* 2019; 297: 371. e371-371. e376.
15. de Almeida SM, de Carvalho MVD, de Lyra Menezes MCT, et al. Validation of the DSP2 tool in a contemporary identified skeletal collection from northeastern Brazil. *Adv Anthropol* 2020; 10: 169-180.
16. Lopes ARdO, Silva EML, Nascimento MMdS, et al. DSP2 for sex determination of miscegenated contemporary hip bones. *Anatomia, Histologia, Embryologia* 2024; 53: e12979.
17. Sánchez-Mejorada G, Gómez-Valdés J, Herrera P, et al. Valoración del método de diagnóstico sexual probabilístico (DSP) en una colección osteológica mexicana. *Estudios de Antropología Biológica* 2011; 15.
18. Quatrehomme G, Radoman I, Nogueira L, et al. 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* 2017; 272: 190-193.
19. San-Millán M, Warrier V, Carrera A, et al. Application of DSP2 for biological sex estimation in a Spanish sample: analysis based on sex and side. *International Journal of Legal Medicine* 2025; 139: 847-862.

20. Lesciotto KM and Klales AR. Sex estimation using metrics of the innominate: A test of the DSP2 method. *Journal of Forensic Sciences* 2025; 70: 249-257.
21. Abegg C, Hoxha F, Campana L, et al. Measuring pelvises in 3D surface scans and in MDCT generated virtual environment: Considerations for applications in the forensic context. *Forensic science international* 2023; 352: 111813.
22. Chapman T, Lefevre P, Semal P, et al. Sex determination using the Probabilistic Sex Diagnosis (DSP: Diagnose Sexuelle Probabiliste) tool in a virtual environment. *Forensic science international* 2014; 234: 189. e181-189. e188.
23. Stan E, Muresan C-O, Dumache R, et al. Sex estimation from computed tomography of os coxae—validation of the Diagnose Sexuelle Probabiliste (DSP) software in the Romanian population. *Applied Sciences* 2024; 14: 4136.
24. Papini A, Galli M, Di Martino N, et al. Studio delle simmetrie e asimmetrie ossee mascellari in collezioni italiane presenti nella cranioteca del Museo di Storia Naturale dell'Università di Firenze tramite CBCT. *Museologia Scientifica* 2015; 14: 87-90.
25. Burrell CL, Emery MM and Gonzalez S. Paget's disease of bone in two medieval skeletons from Poulton Chapel, Cheshire, UK. *International Journal of Osteoarchaeology* 2019; 29: 922-933.
26. Canavan SM, Burrell CL, Emery MM, et al. A case of projectile trauma from the medieval Poulton Chapel, Cheshire, the United Kingdom. *International Journal of Osteoarchaeology* 2019; 29: 1079-1090.
27. Brůžek J, Santos F and Dutailly B. DSP V2. 2017.
28. Mahalanobis PC. On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)* 2018; 80: S1-S7.

29. Knußmann R. Anthropologie. *Handbuch der vergleichenden Biologei des Menschen* 1988; 1: 251-255.
30. Gaillard J. Détermination sexuelle d'un os coxal fragmentaire. *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 1960; 1: 255-267.
31. Schulter-Ellis FP, Schmidt DJ, Hayek L, et al. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I. *Journal of Forensic Sciences* 1983; 28: 169-180.
32. Jerković I, Bašić Ž, Kružić I, et al. Creating reference data on sex for ancient populations using the Probabilistic Sex Diagnosis method: A validation test using the results of aDNA analysis. *Journal of Archaeological Science* 2018; 94: 44-50.
33. Ulijaszek SJ and Kerr DA. Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition* 1999; 82: 165-177.
34. Perini TA, Oliveira GLd, Ornellas JdS, et al. Technical error of measurement in anthropometry. *Revista Brasileira de Medicina do Esporte* 2005; 11: 81-85.
35. Weinberg SM, Scott NM, Neiswanger K, et al. Intraobserver error associated with measurements of the hand. *American Journal of Human Biology: The Official Journal of the Human Biology Association* 2005; 17: 368-371.
36. White TD, Black MT and Folkens PA. *Human osteology*. 3rd ed. ed. Amsterdam: Elsevier/Academic press, 2012.
37. White TD and Folkens PA. *The human bone manual*. Elsevier, 2005.
38. Inskip S, Scheib CL, Wohns AW, et al. Evaluating macroscopic sex estimation methods using genetically sexed archaeological material: The medieval skeletal collection from St John's Divinity School, Cambridge. *American journal of physical anthropology* 2019; 168: 340-351.

Table 2 *Sexing rate and sexing accuracy by DSP2 using all available variables in the Syracuse sample.*

	Known sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	29	0	**29**
Female	0	16	**16**
Undetermined	6	0	**6**
Total	**35**	**16**	**51**
Sexing rate (%)	82.86 (29/35)	100 (16/16)	88.24 (45/51)
Sexing accuracy (%)	100 (29/29)	100 (16/16)	100 (45/45)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Table 3 *Sexing rate and sexing accuracy by DSP2 using all available variables in the Poulton sample.*

	Catalogued sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	23	0	**23**
Female	1	22	**23**
Undetermined	1	6	**7**
Total	**25**	**28**	**53**
Sexing rate (%)	96 (24/25)	78.57 (22/28)	86.79 (46/53)
Sexing accuracy (%)	95.83 (23/24)	100 (22/22)	97.83 (45/46)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Table 4 *Descriptive statistics on the Syracuse sample data for each variable and each sex category, with the significance of the independent sample t-test and Mann-Whitney U-test.*

Descriptive						*p*-value
Variables	Sex	N	Mean (mm)	SD	[Min. - Max.] (mm)	
PUM	M	31	66.98	4.39	[61.00 - 75.66]	0.003
	F	10	71.93	4.16	[65.33 - 77.33]	
SPU	M	33	26.71	2.31	[21.00 - 31.33]	<0.001
	F	15	22.22	2.22	[17.66 - 26.00]	
DCOX	M	33	206.39	8.68	[188.00 - 228.00]	<0.001
	F	14	195.61	9.63	[180.00 - 215.33]	
IIMT	M	31	36.22	4.58	[26.66 - 46.00]	<0.001*
	F	16	45.70	4.75	[37.33 - 53.00]	
ISMM	M	35	105.07	5.63	[90.66 - 118.66]	<0.001
	F	15	96.77	5.07	[86.00 - 105.66]	
SCOX	M	29	148.31	7.01	[136.00 - 164.00]	0.146
	F	13	152.02	8.57	[136.00 - 165.00]	
SS	M	34	71.36	4.59	[63.00 - 81.00]	0.003*
	F	16	67.62	2.67	[64.33 - 75.00]	
SA	M	35	74.30	5.81	[65.00 - 87.66]	0.007
	F	15	78.35	4.92	[71.33- 94.66]	
SIS	M	35	35.71	2.96	[30.00 - 41.66]	<0.001*
	F	16	32.29	2.30	[29.00 - 38.33]	
VEAC	M	35	53.23	3.14	[44.00 - 59.66]	<0.001
	F	16	48.20	2.37	[44.33. - 52.66]	

N = number of os coxae on which the variable was measured.

SD = standard deviation.

p-value = significance of the independent sample t-test and Mann-Whitney U-test.

*Mann-Whitney U-test *p*-value.

Table 5 *Descriptive statistics on the Poulton sample data for each variable and each sex category, with the significance of the independent sample t-test and Mann-Whitney U-test.*

Descriptive						*p*-value
Variables	Sex	N	Mean (mm)	SD	[Min. - Max.] (mm)	
PUM	M	14	73.83	3.45	[69.00 - 80.00]	0.920*
	F	18	73.31	2.38	[69.33 - 78.67]	
SPU	M	20	29.91	2.18	[23.33 - 33.00]	<0.001
	F	22	24.56	1.82	[21.00 - 28.00]	
DCOX	M	6	225.61	8.27	[215.67 - 238.33]	<0.001
	F	13	204.20	9.17	[191.67 - 216.33]	
IIMT	M	20	38.13	5.12	[28.00 - 46.33]	0.441
	F	26	39.30	5.03	[32.00 - 49.67]	
ISMM	M	20	115.86	6.33	[106.00 - 128.00]	<0.001
	F	17	101.58	4.60	[94.00- 108.00]	
SCOX	M	9	157.07	7.85	[145.00 - 170.33]	0.010*
	F	14	153.00	6.44	[145.00 - 168.33]	
SS	M	24	73.43	6.27	[64.33 - 87.00]	0.012
	F	28	69.40	4.91	[61.33 - 77.67]	
SA	M	23	74.69	5.39	[67.00 - 90.33]	0.450
	F	27	75.30	5.80	[64.33 - 84.67]	
SIS	M	25	39.00	2.95	[35.00 - 44.67]	<0.001
	F	26	34.88	3.09	[29.67 - 41.67]	
VEAC	M	23	58.00	2.46	[53.67 - 63.67]	<0.001
	F	25	50.02	2.01	[45.00 - 54.00]	

N = number of os coxae on which variable was measured.

SD = standard deviation.

p-value = significance of the independent sample t-test and Mann-Whitney U-test.

*Mann-Whitney U-test *p*-value.

Supplementary Material

Syracuse best-four, worst-four tables:

Sexing rate and sexing accuracy by DSP2 using the best four variables (PUM, SPU, DCOX, IIMT) in the Syracuse sample.

	Known sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	30	0	**30**
Female	0	12	**12**
Undetermined	5	4	**9**
Total	**35**	**16**	**51**
Sexing rate (%)	85.71 (30/35)	75 (12/16)	82.35 (42/51)
Sexing accuracy (%)	100 (30/30)	100 (12/12)	100 (42/42)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Sexing rate and sexing accuracy by DSP2 using the worst four variables (SS, SA, SIS, VEAC) in the Syracuse sample.

	Known sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	1	0	**1**
Female	4	14	**18**
Undetermined	30	2	**32**
Total	**35**	**16**	**51**
Sexing rate (%)	14.29 (5/35)	87.50 (14/16)	37.25 (19/51)
Sexing accuracy (%)	20 (1/5)	100 (14/14)	78.95 (15/19)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Poulton best-four, worst-four tables:

Sexing rate and sexing accuracy by DSP2 using the best four variables (PUM, SPU, DCOX, IIMT) in the Poulton sample.

	Catalogued sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	13	0	**13**
Female	1	9	**10**
Undetermined	11	19	**30**
Total	**25**	**28**	**53**
Sexing rate (%)	56 (14/25)	32.14 (9/28)	43.40 (23/53)
Sexing accuracy (%)	92.86 (13/14)	100 (9/9)	95.65 (22/23)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Sexing rate and sexing accuracy by DSP2 using the worst four variables (SS, SA, SIS, VEAC) in the Poulton sample.

	Catalogued sex		
	Male	Female	**Total**
Sex estimation via DSP2			
Male	15	0	**15**
Female	0	8	**8**
Undetermined	10	20	**30**
Total	**25**	**28**	**53**
Sexing rate (%)	60 (15/25)	28.57 (8/28)	43.40 (23/53)
Sexing accuracy (%)	100 (15/15)	100 (8/8)	100 (23/23)

Sexing rate = percentage of individuals for whom sex was estimated.

Sexing accuracy = percentage of correctly sexed individuals from the sexed sample.

Intra- and inter-reliability rates for both samples:

Intra- and inter-reliability rates in the Syracuse sample.

Variables	TEM intra (mm)	rTEM intra	R intra	TEM inter (mm)	rTEM inter	R inter
PUM	0.48	0.71	0.98	0.60	0.84	0.97
SPU	0.43	1.72	0.97	0.58	2.15	0.94
DCOX	0.55	0.27	0.99	0.40	0.19	0.99
IIMT	0.60	1.52	0.99	0.65	1.69	0.97
ISMM	0.56	0.55	0.99	0.93	0.90	0.98
SCOX	0.51	0.34	0.99	0.77	0.51	0.99
SS	0.37	0.53	0.99	0.54	0.74	0.98
SA	0.49	0.64	0.99	0.42	0.55	0.99
SIS	0.51	1.47	0.97	0.35	0.96	0.99
VEAC	0.48	0.93	0.98	1.06	2.08	0.96

TEM = technical error of measurement.

rTEM = relative technical error of measurement.

R = reliability coefficient.

Intra- and inter-reliability rates in the Poulton sample.

Variables	TEM intra (mm)	rTEM intra	R intra	TEM inter (mm)	rTEM inter	R inter
PUM	0.52	0.71	0.97	1.31	1.81	0.99
SPU	0.41	1.54	0.98	0.80	3.07	0.98
DCOX	0.57	0.27	0.99	1.37	0.66	0.99
IIMT	0.59	1.53	0.98	3.48	8.91	0.94
ISMM	0.51	0.47	0.99	1.19	1.13	0.99
SCOX	0.71	0.45	0.99	1.45	0.93	0.99
SS	0.43	0.61	0.99	0.99	1.43	0.99
SA	0.47	0.64	0.99	1.17	1.60	0.99
SIS	0.38	1.05	0.98	1.39	3.81	0.79
VEAC	0.49	0.91	0.98	1.71	3.23	0.86

TEM = technical error of measurement.

rTEM = relative technical error of measurement.

R = reliability coefficient.