The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

| | |
|---|---|
| Document Title: | An Interactive Morphological Database for Estimating Sex in Modern Adults |
| Author(s): | Alexandra R. Klales |
| Document Number: | 252746 |
| Date Received: | March 2019 |
| Award Number: | 2015-DN-BX-K014 |

**NIJ Award: Award No. 2015-DN-BX-K014**

# *An Interactive Morphological Database for Estimating Sex in Modern Adults*

**Final Summary Overview Report: August 26, 2018**

*Principle Investigator:*
Alexandra R. Klales
Washburn University
alexandra.klales@washburn.edu
484-844-3329

*Recipient Organization:*
Mercyhurst University
501 E. 38th Street
Erie, PA 16546
POC: Matthew Sanfilippo

This project is supported by **Award No. 2015-DN-BX-K014**, awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice.

Submission: April 2015

Award Date: September 16, 2015

Grant Period: January 2016 – June 30, 2018 (extension)

## Introduction and Purpose

Estimating an unidentified person's biological profile (age, sex, ancestry, stature) from their skeletal remains is a vital component of the forensic anthropologist's role in medico-legal investigations and is necessary to narrow the list of potential victims. Sex estimation is the most important aspect of this profile because it essentially reduces the list by half and the analytical methods for the other parameters are often sex-specific. This research examined validity, reliability, temporal change, population variation, and the effects of asymmetry on sex classification using the eight most popular morphological traits of the skull (Walker 2008) and pelvis (Klales et al. 2012), in order to develop a free, interactive morphological database for sex estimation. Within the MorphoPASSE database program, practitioners can enter morphological traits of their unknown case and compare them with a large sample of known demographic data in order to more accurately and more easily estimate sex in a manner compliant with the *Daubert* criteria (Daubert vs. Merrell Dow Pharmaceuticals, Inc. 1993).

## Project Design and Methods

Score data of the five Walker (2008) traits (nuchal crest-NC, mastoid process-MP, glabella-G, supra-orbital margin-SO, mental eminence-ME) and the three Klales et al. (2012) traits (ventral arc-VA, subpubic contour-SPC, medial aspect of the ischio-pubic ramus-MA) were collected from 2,026 individuals of known sex and ancestry (Asian-A, Black-B, Hispanic-H, Native American-NA, White-W) from multiple national and international skeletal collections. Score data were also sourced from other researchers for an additional 696 individuals after the evaluation of observer agreement was assessed and determined to be acceptable. Validity of the original methods was assessed by entering each individual's score data into the original equations provided by both methods (Table 1). Recalibrated logistic regression equations were also calculated for each population group and for the global sample for the skull and pelvis.

Table 1. Sex classification equations provided by Walker (2008) and Klales et al. (2012).

| **Method** | **Equation** |
|---|---|
| Walker (2008) | Y=glabella x -1.375 + mastoid x -1.185 + mental x -1.151 + 9.128 |
| | Y=glabella x -1.568 + mastoid x -1.459 + 7.434 |
| | Y=glabella x -1.525 + mental x -1.485 + 7.372 |
| | Y=mental x -1.629 + mastoid x -1.415 + 7.382 |
| | Y=orbit x -1.007 + mental x -1.850 + 6.018 |
| | Y=nuchal x -0.7 + mastoid x -1.559 + 5.329 |
| Klales et al. (2012) | Y=ventral arc x 2.726 + medial aspect x 1.214 + subpubic contour x 1.073 - 16.312 |

The recalibration accuracy results were compared to the validation results, as well as, to the classification accuracies generated using the global equation to determine if population specific equations were necessary. Reliability was evaluated by comparing trait scores between seven observers with varying levels of experience (expert, experience, inexperienced) and within observers, for intraobserver agreement, using the intraclass correlation coefficient (ICC) and Cohen's (1968) weighted Kappa (w*K*). To assess population variation, trait score frequencies and trait score means were calculated for each ancestry group by sex. A Fisher's exact test with Monte Carlo simulation was used to test for sexual dimorphism in trait score frequencies between females and males for each trait within each group. To examine the differences between the geographic regions and ancestry groups a Kruskal–Wallis test was used. Temporal variation was assessed by comparing trait score frequencies between contemporary and historic samples using the Mann-Whitney U test. To assess asymmetry, bilateral traits from the skull (MP, SO) and pelvis (VA, SPC, MA) were tested for agreement between left and right sides using the ICC. Trait frequency calculations were tabulated to determine how common asymmetry is in each of the five bilateral traits and to determine directionality in asymmetric individuals. Classification accuracy between the left and right sides were compared to determine how asymmetry impacted correct classification and application of the two methods. Finally, after the completion of all these analyses, the data were incorporated into a free user database, known as MorphoPASSE.

## Results

### *Validity*

For the skull, validation of the Walker method produced high classification accuracy for males (average 94.1% all groups), but poor accuracy for females (58.5% all groups) (Table 2), suggesting that population specific recalibration was necessary. Equation one produced the highest overall classification accuracy (84.1%) utilizing the G, MP, and ME. Conversely, validation of the pelvis equation remained high for all groups, with females achieving higher accuracy overall (Table 2). Accuracy was above 90% for all groups except Asian (87.1%) and Hispanic (84.4%) males, which suggested a lower level of sexual dimorphism in these groups and that population specific equations would likely increase accuracy in these ancestry groups. Recalibration of the equations using the worldwide sample (all ancestry groups) increased accuracy to 83.0% overall (78.8% females; 86.0% males) using all five traits of the skull. Recalibrating the pelvis using the worldwide sample achieved comparable accuracy to the validation test (96.2% vs. 94.9%). Details of the pelvis analysis using this data and additional sourced data from the grant can be found in Kenyhercz et al. (2017).

Table 2. Validity of the six Walker (2008) and the one Klales et al. (2012) equations based on ancestry/population group.

| | AM* (n=88) | AF* (n=46) | BM (n=399) | BF (n=364) | HM (n=107) | HF (n=57) | WM (n=625) | WF (n=453) | Average All Groups |
|---|---|---|---|---|---|---|---|---|---|
| Eq1 | 100.0 | 100.0 | 85.4 | 70.7 | 93.2 | 69.6 | 94.4 | 61.9 | 84.1 |
| Eq2 | 100.0 | 100.0 | 84.8 | 61.3 | 85.2 | 60.7 | 95.0 | 59.3 | 81.4 |
| Eq3 | 100.0 | 100.0 | 90.1 | 58.5 | 95.2 | 77.5 | 97.2 | 42.4 | 81.7 |
| Eq4 | 100.0 | 100.0 | 91.4 | 46.4 | 99.0 | 23.4 | 93.4 | 35.5 | 74.3 |
| Eq5 | 100.0 | 0.0 | 94.7 | 39.2 | 99.0 | 22.5 | 95.8 | 25.5 | 61.3 |
| Eq6 | 100.0 | 100.0 | 84.5 | 56.6 | 92.5 | 40.4 | 86.5 | 59.9 | 76.9 |
| Skull Avg. | 100.0 | 83.3 | 88.5 | 55.4 | 94.0 | 49.0 | 93.7 | 47.4 | 76.6 |
| Klales | 87.1 | 100.0 | 94.2 | 97.7 | 84.4 | 98.2 | 90.9 | 96.9 | 94.9 |

*n=4 for AM skulls and n=1 for AF skulls, so scores are inflated due to small sample size.

### ***Reliability***

Intraobserver agreement was substantial to almost perfect with the exception of the ME in one observer and the SO in another observer (Table 3). Interobserver agreement was excellent for all traits, except the SO, when three observers (expert, experienced, and inexperienced) were compared (Table 4). Agreement was excellent between the three expert observers, with the exception of the ME (Table 5). Results from this portion of the research suggested that multiple observers, with varied backgrounds and experience levels, can reliably score the traits utilized in Walker (2008) and Klales et al. (2012) with the exception of the mental eminence (ME), which has shown considerable variation between studies and individuals. Because trait scoring is reliable, data can be sourced from other researchers for inclusion into the database in order to increase the sample sizes and population inclusiveness of the database program. Expanded observer error results that include data from the grant and additional sourced data can be found in Walls et al. (2018).

Table 3. Intraobserver error using linear w*K* for Observer 1, 2, and 3.

| Trait | Expert 1 | Experienced | Inexperienced |
|---|---|---|---|
| VA | 0.87 | -- | -- |
| SPC | 0.89 | -- | -- |
| MA | 0.84 | -- | -- |
| NC | 0.72 | 0.81 | 0.76 |
| G | 0.74 | 0.83 | 0.86 |
| SO | 0.67 | 0.69 | 0.60 |
| MP | 0.66 | 0.73 | 0.83 |
| ME | 0.72 | 0.44 | 0.62 |

Table 4. Interobserver agreement (ICC) between three observers of varying experience levels (Observer 1-3).

| | ICC Absolute | Agreement Level |
|---|---|---|
| VA | 0.89 | Excellent |
| SPC | 0.85 | Excellent |
| MA | 0.79 | Excellent |
| NC | 0.85 | Excellent |
| G | 0.83 | Excellent |
| SO | 0.70 | Good |
| MP | 0.77 | Excellent |
| ME | 0.75 | Excellent |

Table 5. Interobserver agreement using the ICC between three expert observers (Observer 1,4,5).

| | Expert 1 and 2 | | Expert 1 and 3 | |
|---|---|---|---|---|
| **Trait** | **ICC Absolute** | **Agreement Level** | **ICC Absolute** | **Agreement Level** |
| VA | 0.94 | Excellent | - | - |
| SPC | 0.91 | Excellent | - | - |
| MA | 0.83 | Excellent | - | - |
| NC | 0.94 | Excellent | 0.82 | Excellent |
| G | 0.89 | Excellent | 0.88 | Excellent |
| SO | 0.79 | Excellent | 0.86 | Excellent |

| | | | | |
|---|---|---|---|---|
| MP | 0.89 | Excellent | 0.82 | Excellent |
| ME | 0.58 | Fair | 0.64 | Good |

***Population Variation***

For the skull, significant differences were found in traits for all groups ($p=0.016$ or lower) between males and females indicating sexually dimorphic differences. Population differences were also found between all groups for all five traits at the $p<0.001$ level. Population variation was more varied for the pelvis; however, as with the skull, significant differences ($p < 0.001$) in score frequencies were found between males and females in all three traits indicating the presence of sexual dimorphism. A score of 1 was most frequent for females, while scores of 4 or 5 were most frequent for males and significant group level differences were observed in the expression of VA for both the males ($p < 0.001$) and the females ($p = 0.03$). Females mostly had a SPC score of 1 or 2 (88%), whereas males tended to have a SPC score of 4 or 5 (87%). Males do not show any significant level differences in the distribution of trait scores ($p = 0.08$) between geographic groups, but females did demonstrate significant differences in trait expression ($p < 0.001$). Finally, for the MA, females exhibited more score variation than males, including more robust scores (4/5) than any other trait. There were significant, geographic group level differences in the expression of the MA for both females ($p < 0.001$) and males ($p < 0.001$). More detailed information on pelvis population variation using the grant data and additionally sourced data can be found in Kenyhercz et al. (2017).

***Temporal Variation***

Contemporary and historic sample sizes were only large enough to compare for whites and blacks of both sexes. Significant differences between the modern and historic samples were found for the G ($p=0.012$) and left SPC ($p=0.039$) in black females and G ($p=0.008$), left MP ($p=0.012$), and the SPC ($p=0.037$ left; $p=0.002$ right) in black males. In all cases, the contemporary sample had significantly lower (i.e., more gracile) scores. In whites, significant differences between temporal periods were found for the MP ($p=0.000$ left and right), G ($p=0.000$), and NC ($p=0.03$) in females and the NC ($p=0.000$) in males. All three traits of the pelvis were significantly different at the $p>0.05$ level for both white males and females. All traits exhibited significantly lower scores (i.e., more gracile) in the modern sample except for the NC which has more robust/higher scores in the modern white samples.

***Asymmetry***

Asymmetry was found to be present in all bilateral traits. Using parameters outlined by Cicchetti (1994), the ICC results indicate good to excellent consistency between scores from the left and right sides of the same individual for the skull and pelvic traits. The frequency of asymmetry ranged from 21.6% to 36.0% for females and 23.4% to 41.0% for males. Despite asymmetries being common, the degree of asymmetry was relatively low, with the vast majority of asymmetries consisting of only one score point difference between left and right sides for all traits. Directional asymmetries were detected, with the majority of traits exhibiting right dominance (i.e., the right side received the higher/more robust score) for both sexes. The trend towards right dominance led to increased classification accuracies for females when traits from the left side were used for analysis, with the reverse being true for males. Asymmetric individuals were generally found to render significantly ($p < 0.05$) lower classification accuracies than their symmetric counterparts; however, classification differences were minimal between the symmetric and asymmetric groups. The results of the asymmetry portion of the grant are presented in more detail in Cole et al. (2017) and Cole (2017).

Table 6. Agreement between the left and right sides of the bilateral traits using the ICC.

| Trait | Females | Agreement | Males | Agreement |
|---|---|---|---|---|
| MP | 0.690 | Good | 0.672 | Good |
| SO | 0.729 | Good | 0.739 | Good |
| VA | 0.754 | Excellent | 0.615 | Good |
| SPC | 0.744 | Good | 0.671 | Good |

| | | | | |
|---|---|---|---|---|
| MA | 0.743 | Good | 0.785 | Good |

***Database***

The data collected as part of this grant and the data sourced from additional researchers have been integrated into the program MorphoPASSE: the Morphological Pelvis and Skull Sex Estimation Database (available here: https://morphopasse.shinyapps.io/morphoPASSE/). The database contains access to all data that was collected as part of this grant and sourced data that already has dissemination approval from the researcher that collected it. A website has been created to host the database and all information related to it: www.MorphoPASSE.com. The website also contains an accompanying manual that includes 1) revised trait scoring information and photos based on this research, 2) information on the samples and populations, and 3) instructions for operating the database program, as well as, how to properly interpret the statistical output. The website also contains dissemination completed thus far, including conference papers and peer-reviewed journal publications. Finally, the website also contains instructions and a form that researchers may use to request access to the data collected for this grant.

## Project Finding Summary

This research has tested the validity and reliability of the popular Walker (2008) and Klales et al. (2012) morphological methods in a variety of ancestry populations. The myriad of factors impacting applications of these methods (e.g. population differences, asymmetry, temporal change) have been evaluated and the results suggest that these methods can be confidentially applied to active forensic anthropological casework in both the U.S. and abroad. In addition, these methods can now be more easily applied using the free MorphoPASSE database, which takes into consideration the results from all the analyses listed above.

## Implications for Criminal Justice Policy and Practice in the U.S.

The creation of the database and the standardization of techniques will have a noted impact on the field. The factors impacting the application of these methods have been thoroughly tested and both methods are now compliant with the *Daubert* criteria (Daubert vs. Merrell Dow Pharmaceuticals, Inc. 1993). Each method can also be more easily applied in a standardized fashion to active forensic casework using the free MorphoPASSE database. Because the database and data are both freely available to the public, it has already and will continue to facilitate collaboration. The free access to the MorphoPASSE data collection software also provides a teaching and practical resource to the academic and professional forensic anthropological community.

## References

Cicchetti, D.V. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 4, 284–290.*

Cohen, J. (1968) Weighted kappa: nominal scale agreement with provisions for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213–220.

Cole, S.J. (2017). The effects of cranial and pelvic asymmetry on accurate sex classification, Master's thesis [Mercyhurst University, Erie, PA].

Cole, S.J., Cabo, L.L., & Klales, A.R. (2017). The effects of cranial and pelvic asymmetry on accurate sex classification, *Proceedings of the 69$^{th}$ Annual Meeting of the American Academy of Forensic Sciences* (pp.118). New Orleans, Louisiana, United States.

Daubert v. Merrell Dow Pharmaceuticals. (1993). US Supreme Court 509.U.S.579,113S.Ct.2786, 125L. Ed.2d 469.

Kenyhercz, M.W., Klales, A.R., Stull, K.E., McCormick, K.A., & Cole, S.J. (2017). Worldwide population variation in pelvic sexual dimorphism: a validation and recalibration of the Klales et al. method. *Forensic Science International*, *277*, 259-259.

Klales, A.R., Ousley, S.D., & Vollner, J.M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology, 149*, 104-114.

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

Walls, M.M., Klales, A.R., Lesciotto, K.M., Gocha, T.P., & Garvin, H.M. (2018) Reliability and validity of the Walker (2008) and Klales et al. (2012) methods, *Proceedings of the70th Annual Meeting of the American Academy of Forensic Sciences* (pp.43). Seattle, Washington, United States.