Forensic Anthropology Population Data

Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method☆

Michael W. Kenyhercz[a,b,*], Alexandra R. Klales[c], Kyra E. Stull[d,b], Kyle A. McCormick[a], Stephanie J. Cole[e]

[a] Department of Defense POW/MIA Accounting Agency, Central Identification Laboratory, 590 Moffet Street, Bldg 4077, JBPHH, HI 96853, United States
[b] Department of Anatomy, University of Pretoria, Private Bag x323, 0007, Arcadia, South Africa
[c] Forensic Anthropology Program, Sociology & Anthropology Department, Washburn University, 1700 SW College Avenue, Topeka, KS 66621, United States
[d] Department of Anthropology, University of Nevada, Reno, 1664 North Virginia Street, Stop 0096, Reno, NV 89557, United States
[e] Department of Applied Forensic Sciences, Mercyhurst University, 501 E 38th St., Erie, PA 16546, United States

ARTICLE INFO

ABSTRACT

Sex estimation is an integral aspect of biological anthropology. Correctly estimating sex is the first step to many subsequent analyses, such as estimating living stature or age-at-death. Klales et al. (2012) [6] provided a revised version of the Phenice (1969) [3] method that expanded the original three traits (ventral arc, subpubic concavity/contour, and medial aspect of the ischio-pubic ramus) into five character states to capture varying degrees of expression within each trait. The Klales et al. (2012) [6] method also provided associated probabilities with each sex classification, which is of particular importance in forensic anthropology. However, the external validity of this method must be tested prior to applying the method to different populations from which the method was developed. A total of 1915 innominates from four diverse geographic populations: (1) U.S. Blacks and Whites; (2) South African Blacks and Whites; (3) Thai; and (4) unidentified Hispanic border crossers were scored in accordance with Klales et al. (2012) [6]. Trait scores for each innominate were entered into the equation provided by Klales et al. (2012) [6] for external validation. Additionally, recalibration equations were calculated with logistic regression for each population and for a pooled global sample. Validation accuracies ranged from 87.5% to 95.6% and recalibration equation accuracies ranged from 89.6% to 98% total correct. Pooling all samples and using Klales' et al. (2012) [6] equations achieved an overall validation accuracy of 93.5%. The global recalibration model achieved 95.9% classification accuracy and can be employed in diverse worldwide populations for accurate sex estimation without the need for population specific equations.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Sex estimation is of particular importance within biological anthropology, as subsequent analyses are often sex specific, such as age and stature [1]. In practical terms, accurate sex estimations allow bioarchaeologists to estimate past population demographics and forensic anthropologists to develop a biological profile that is *Daubert* compliant [2].

Many nonmetric sex estimation techniques are available for the skull and postcranial elements; however, the innominate has long been regarded as the most sexually dimorphic element within humans [3–5]. Noting the differences in the human innominate morphology, specifically the pubic bone, Phenice [3] suggested that the presence or absence of the ventral arc (VA), subpubic concavity (SPC), and ridge along the medial aspect of the ischio-pubic ramus (MA) could accurately estimate sex with a 96% accuracy rate. Klales et al. [6] revised the Phenice [3] method by expanding the scoring of each of the three traits from presence or absence into five ordered character states. Using modified trait descriptions, Klales et al. [6] achieved classification rates ranging

* Corresponding author.
  E-mail address: michael.kenyhercz@gmail.com (M.W. Kenyhercz).

from 93.5–95.5% total correct for experienced observers. Besides providing a *Daubert* compliant method, expanding Phenice's [3] binary system into five ordinal character states accounts for a greater range of variation than simply the presence or absence of a particular trait and more broadly, allows for studies of sexual dimorphism through time or among populations [6].

In order for a method to be confidently employed in a population, it must be validated with different samples. While Klales et al. [6] included a validation sample in their original article, both the calibration and validation samples were derived from U.S. collections. Given the different levels of sexual dimorphism among populations, the original equation provided by Klales et al. [6] needs to be validated on populations that were not included in the original sample and on populations with geographic and genetic diversity. If necessary, the equations need to be re-calibrated to account for the greater degree of human variation. Recent validation studies in Hispanic populations have suggested that recalibration improves accuracy and can decrease sex bias [8,9].

The aims of the current research are two-fold: (1) test the original Klales et al. [6] logistic regression equation on a large sample of various geographic populations (i.e., test the external validity of the method) and (2) determine if population specific equations (i.e., recalibration of the original logistic regression equation for each geographic group) are necessary or if a global equation, combining all temporal and geographic groups, can be accurately applied instead.

2. Materials and methods

The first four authors scored 2019 innominates from four geographic regions following the descriptions and illustrations in Klales et al. [6]. Each of these authors have previous experience and training in the Klales et al. [6] method and all have advanced degrees in biological anthropology. Previous research has shown inter- and intra-observer error for the Klales et al. method is minimal [6]. Only innominates with all three traits available for scoring were used for the current study, resulting in a total sample size of 1915. The geographic populations included are: (1) U.S. Blacks and Whites; (2) South African Blacks and Whites; (3) Thai; and (4) unidentified Hispanic border crossers recovered in the U.S. Southwest (Table 1). When all four of the geographic populations are included (n = 1195) in model creation, it is hereafter referred to as the "global" sample. For the U.S. and South African samples, the Black and White ancestry groups were analyzed separately and also as pooled geographic groups. The U.S. population is comprised of individuals from the Hamann-Todd Human Osteological Collection housed at the Cleveland Museum of Natural History, the Robert J. Terry Anatomical Skeletal Collection from the Smithsonian Institute, the William M. Bass Donated Skeletal Collection at the University of Tennessee, Knoxville, identified forensic cases from the Department of Applied Forensic Sciences at Mercyhurst University, and the Hartnett–Fulginiti Pubic Bone Collection at the Maricopa County Office of the Medical Examiner in Phoenix, Arizona. The South African sample was collected from

the Pretoria Bone Collection at the University of Pretoria, South Africa. The Thai sample was collected at Khon Kaen University, Thailand. Lastly, the unidentified border crosser (UBC) sample was collected at the Forensic Anthropology Center at Texas State University. The UBC sample consists of unidentified migrants who died crossing the U.S.–Mexico border. The demographic information for these individuals was inferred based on a number of variables. Ancestry was estimated based on a combination of factors including artifacts and metric analyses using FORDISC [7], while sex was determined via DNA or visual assessment of genitalia. For the remaining individuals, sex was estimated based on artifacts and metric analyses (see Ref. [9] for a more in depth discussion of sex and ancestry estimation for these individuals).

Trait score frequencies and means were tabulated for each of the four geographic regions for each trait by sex. Additionally, the trait score distributions by geographic group were visualized with ggplot2 [10]. To test for sexual dimorphism in trait frequencies, the count data for each trait was subjected to a Fisher–Freeman–Halton test for each population group and the pooled sample. Lastly, a Kruskal–Wallis test was used to examine differences between the four geographic regions for each sex by trait. If the Kruskal–Wallis test was rejected, then a Dunn's test, a nonparametric pairwise multiple-comparison procedure, was employed using a Holm's stepwise adjustment.

All three trait scores for each innominate were entered into the original logistic regression formula supplied by Klales et al. [6] and the classification accuracy was recorded to examine external validity of the original method. After each innominate was classified, total correct classifications were tallied. Sex bias was calculated by subtracting the male total correct classification from the female total correct classification, which means that any positive value indicates a sex bias in favor of females and any negative value indicates a sex bias in favor of males. Next, logistic regression recalibration equations were calculated in the statistical program R [11] for each geographic population, as well as, for the global sample. Classification accuracy of the recalibrated models were then compared to the classification accuracy of those geographic populations using the global equation to determine if population specific equations are necessary.

3. Results

3.1. Frequency distributions

Frequency distributions of each trait are shown in Tables 2–7 and Figs. 1–3. Significant differences (p < 0.001) in score frequencies were observed for all traits between males and females with regard to each population (i.e., ancestry groups for the U.S. and South African samples), pooled sample (i.e., combined ancestry groups for the U.S. and South African samples), and the global pooled sample (i.e., all individuals from all geographic groups).

Table 1
Sample composition by geographic region, ancestry group (where appropriate), and sex.

Geographic group	Females	Males	Total
South Africa Black	50	50	100
South Africa White	50	50	100
Thai	45	96	141
Hispanic UBC	24	24	48
U.S. Black	254	294	548
U.S. White	401	579	980
Total	823	1092	1915

Table 2
Frequency distributions (%) of the expressions of the VA and mean trait score for females. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	**72.0**	22.0	6.0	0.0	0.0	1.34
South Africa White	**80.0**	16.0	2.0	0.0	2.0	1.28
South Africa Pooled Ancestry	**76.0**	19.0	4.0	0.0	1.0	1.31
Thai	**55.6**	37.8	2.2	2.2	2.2	1.58
Hispanic UBC	**83.3**	4.2	4.2	4.2	4.2	1.52
U.S. Black	**54.5**	36.4	7.1	0.4	1.6	1.58
U.S. White	**71.8**	19.7	6.7	0.7	1.0	1.39
U.S. Pooled Ancestry	**65.1**	26.1	6.9	0.6	1.2	1.47
Global	**66.5**	25.3	6.2	0.7	1.3	1.46

Table 3
Frequency distributions (%) of the expressions of the VA and mean trait score for males. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	0.0	2.0	20.0	**40.0**	38.0	4.14
South Africa White	2.0	2.0	16.0	36.0	**44.0**	4.18
South Africa Pooled Ancestry	1.0	2.0	18.0	38.0	**41.0**	4.16
Thai	1.0	5.3	29.5	**36.8**	27.4	3.84
Hispanic UBC	4.2	4.2	12.5	**45.8**	33.3	4.00
U.S. Black	0.0	1.4	12.9	**43.2**	42.5	4.27
U.S. White	0.5	4.3	13.1	38.7	**43.4**	4.20
U.S. Pooled Ancestry	0.3	3.3	13.1	40.2	**43.1**	4.22
Global	0.6	3.4	14.9	39.8	**41.3**	4.18

Table 4
Frequency distributions (%) of the expressions of the SPC and mean trait score for females. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	**78.0**	16.0	6.0	0.0	0.0	1.28
South Africa White	**76.0**	20.0	2.0	2.0	0.0	1.30
South Africa Pooled Ancestry	**77.0**	18.0	4.0	1.0	0.0	1.29
Thai	**64.5**	31.1	4.4	0.0	0.0	1.80
Hispanic UBC	**58.3**	25.0	8.3	8.3	0.0	1.67
U.S. Black	**50.6**	35.6	8.7	4.3	0.8	1.69
U.S. White	**58.4**	29.7	7.7	3.7	0.5	1.58
U.S. Pooled Ancestry	**55.4**	32.0	8.1	4.0	0.6	1.63
Global	**58.6**	30.0	7.4	3.5	0.5	1.58

Table 5
Frequency distributions (%) of the expressions of the SPC and mean trait score for males. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	0.0	2.0	20.0	**44.0**	34.0	4.10
South Africa White	0.0	4.0	8.0	20.0	**68.0**	4.52
South Africa Pooled Ancestry	0.0	3.0	14.0	32.0	**51.0**	4.31
Thai	0.0	3.2	16.8	**47.4**	32.6	4.09
Hispanic UBC	4.2	0.0	16.7	**50.0**	29.2	3.96
U.S. Black	0.3	4.4	5.4	**55.8**	34.0	4.14
U.S. White	0.2	3.1	8.8	**51.1**	36.8	4.02
U.S. Pooled Ancestry	0.2	3.6	7.7	**52.7**	35.9	4.06
Global	0.3	3.4	9.2	**50.3**	36.8	4.20

Table 6
Frequency distributions (%) of the expressions of the MA and mean trait score for females. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	34.0	**50.0**	14.0	2.0	0.0	1.84
South Africa White	**48.0**	40.0	8.0	4.0	0.0	1.68
South Africa Pooled Ancestry	41.0	**45.0**	11.0	3.0	0.0	1.76
Thai	**46.7**	44.4	8.9	0.0	0.0	1.62
Hispanic UBC	12.5	29.2	**54.2**	4.2	0.0	2.60
U.S. Black	7.9	34.8	**48.2**	7.1	2.0	2.60
U.S. White	11.2	30.9	**46.1**	10.0	1.7	2.60
U.S. Pooled	9.9	32.4	**46.9**	8.9	1.8	2.60
Global	15.8	34.5	**40.7**	7.5	1.5	2.45

Table 7
Frequency distributions (%) of the expressions of the MA and mean trait score for males. Highest frequency for each expression per group is in bold.

Geographic group	1	2	3	4	5	Mean
South Africa Black	0.0	0.0	14.0	**44.0**	42.0	4.28
South Africa White	0.0	4.0	20.0	**38.0**	**38.0**	4.10
South Africa Pooled Ancestry	0.0	2.0	17.0	**41.0**	40.0	4.19
Thai	1.1	7.4	26.3	**50.5**	14.7	3.65
Hispanic UBC	0.0	0.0	**54.2**	20.8	29.2	3.71
U.S. Black	0.0	0.7	14.6	34.4	**50.3**	4.34
U.S. White	0.2	1.2	18.5	**42.7**	37.5	4.16
U.S. Pooled Ancestry	0.1	1.0	17.2	39.9	**41.8**	4.22
Global	0.2	1.6	18.8	**40.5**	38.9	4.16

expression of scores 4 and 5 (combined 64.2%), and also the lowest mean VA score (3.84). Contrastingly, both South African and U.S. males have the highest incidences of scores of 5.

3.1.2. Subpubic contour

The relative frequencies of SPC expression are shown in Tables 4 and 5 and Fig. 2. Overall, more than 88% of all females show an SPC score of 1 or 2, whereas over 87% of all males have an SPC score of 4 or 5. Males do not show any significant level differences in the distribution of trait scores (p = 0.08) between geographic groups, but females do demonstrate significant differences in trait expression (p < 0.001). The pairwise comparisons indicate that the primary difference in trait expression is between U.S. and South African females (p < 0.001). The U.S. females, specifically U.S. Black females, have the lowest frequency of an SPC score of 1 (55.4% for U.S. pooled females), while South African females have the highest frequency of an SPC score of 1 (77.0% for South African pooled females).

3.1.3. Medial aspect of the ischio-pubic ramus

The frequencies of MA expressions are shown for females in Table 6, males in Table 7, and graphically in Fig. 3. Overall, females show a much wider distribution of MA scores than males, which includes the highest incidence of scores 4 and 5 than any other trait observed in females and the lowest incidence of a score of 1 than any other trait by 40% or more. Further, there are significant, geographic group level differences in the expression of MA for both females (p < 0.001) and males (p < 0.001).

The Dunn's test indicates that there are pairwise differences between South African females and both UBC and U.S. females (both p < 0.001) and also between Thai females and both UBC and U.S. females (both p < 0.001). The UBC and U.S. females have the highest mean MA scores (both 2.60) while the Thai and South African females have the lowest mean MA scores (1.62 and 1.76, respectively). Significant pairwise differences are also found between South African males and both Thai and UBC males (p = 0.003 and p = 0.012, respectively). Similarly, U.S. males show significant differences in MA expression with both Thai and UBC males (p < 0.001 and p = 0.005, respectively). South African and U. S. males show the highest mean MA scores (4.19 and 4.22, respectively) while Thai and UBC males demonstrate the lowest mean MA scores (3.65 and 3.71, respectively).

3.2. Classification accuracies

3.2.1. Validation

Total correct classifications using the logistic regression equation provided by Klales et al. [6] for each population group range from 87.5% (UBC) to 95.6% (U.S. Black) (Table 8). The total correct classification for the global pooled sample is 93.5%. Sex biases range from low (3.7% for U.S. Black) to high (17.7% for Thai). For all geographic groups and the combined global sample, females
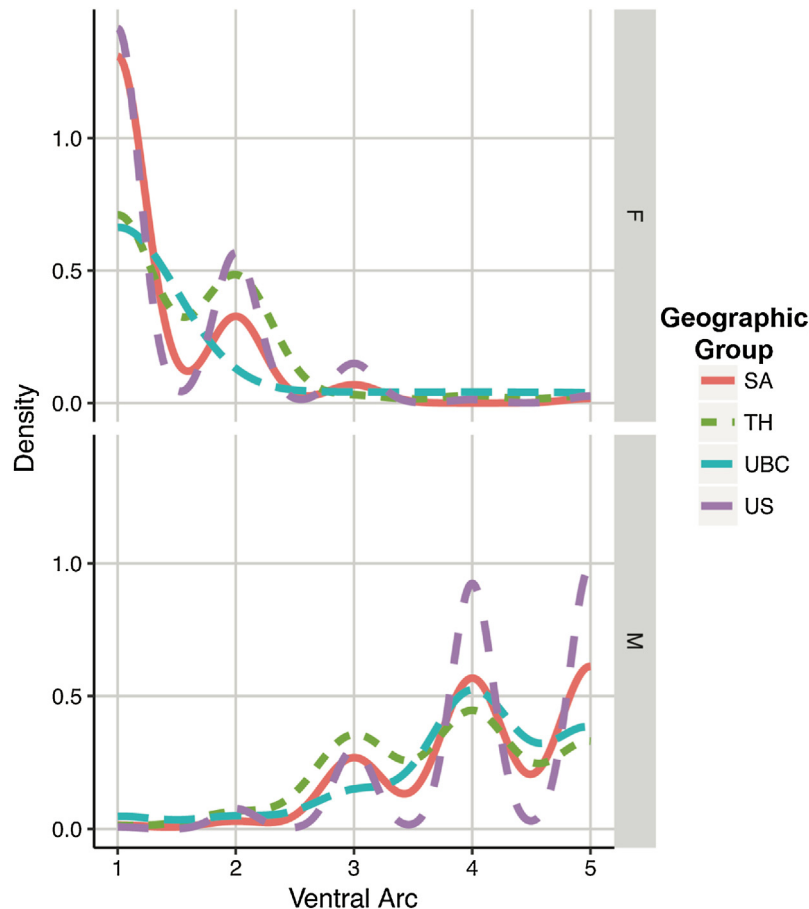
3.1.1. Ventral arc

In all geographic groups, a score of 1 was most frequent for females, while scores of 4 or 5 were most frequent for males (Tables 2 and 3; Fig. 1). Significant geographic group level differences are observed in the expression of VA for both the males (p < 0.001) and the females (p = 0.03). The Dunn's test indicates that there are pairwise differences between Thai males and both South African and U.S. males. Thai males show the highest expression of scores 3 and below (combined 35.8%), the lowest

Fig. 1. Density plots of VA scores for each geographic group by sex where SA = South Africa, TH = Thailand, UBC = undocumented Hispanic border crossers, and US = United States.

have higher correct classifications than males using the original Klales et al. [6] equation.

3.2.2. Recalibration

The results of the logistic regression recalibration equations for each geographic group are provided in Table 9. Population specific correct classifications range from 91.7% (UBC) to 98.6% (Thai). Sex biases range from 0 (South African White) to 8.3% (UBC). Contrary to the results obtained by the original equation, each geographic group, with the exception of South African Whites, shows a male sex bias. In all instances, the sex bias is less pronounced in each comparison than with the original equation.

Further, each individual geographic group was subjected to a logistic regression created from the global model, or all pooled samples (Table 10). The global recalibrated model achieves a total correct classification of 95.9% with an absolute sex bias of 2% in favor of males. When individual geographic groups are subjected to the global equation, correct classifications range from 89.6% (UBC) to 98.0% (South African Black). The pattern of sex bias is similar to the individual recalibration equations with the exception of lowering the bias of UBC from 8.3% male bias to 4.2% male bias.

3.2.3. Recalibration and validation comparison

A comparison of total correct classifications from the original Klales et al. [6] validation (Section 3.2.1), the population specific recalibration by geographic group (Section 3.2.2), and the global recalibration equations are shown in Table 11. In all instances the population specific and global recalibration equations outperform the original Klales et al. [6] equation. Generally, global recalibration

performed similar to the population specific equations. Only the Thai, UBC, and U.S. Black groups show lower classification rates using the global compared to the population specific equation; however, classification accuracy is still at approximately 90% or higher for these groups.

The recalibration equation for the global pooled sample is as follows:

$$Y = 1.4296(VA) + 1.0415(SPC) + 0.9752(MA) - 10.0139$$

To derive the probability that a case is female, the following equation may be used:

$$p_f = 1/(1 + e^{(Y)})$$

To calculate the probability of being male use:

$$p_m = 1 - p_f$$

where p_f is the probability of being female, p_m is the probability of being male, and $e^{(Y)}$ is the exponential constant risen to the power of the logistic regression equation score [12]. The sectioning point is 0 and a positive number indicates a classification of male and a negative number indicates a classification of female.

4. Discussion

Overall, the original method (i.e., the external validity of the method) performs well when applied to samples and populations other than the ones from which it was developed. The validation accuracies, obtained using multiple independent geographic groups, exceed the results obtained from the original Klales
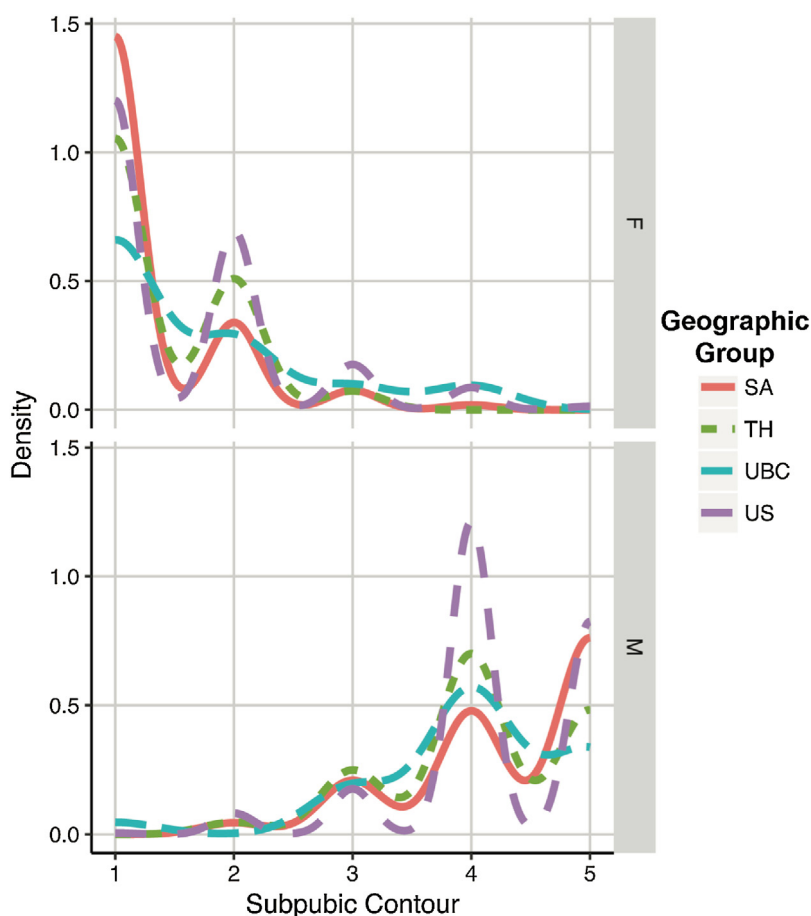
Fig. 2. Density plot of SPC scores for each geographic group by sex where SA = South Africa, TH = Thailand, UBC = undocumented Hispanic border crossers, and US = United States.

et al. [6] validation using the William M. Bass Donated Skeletal Collection (86.2% accuracy). Not surprisingly, the geographic groups that included the same ancestry groups as the original publication outperformed those that did not. For example, the U.S. pooled group achieved the highest classification accuracy, followed by the South African pooled group. Each of these samples contain individuals of African (Black) and European (White) ancestry. Both the Thai (87.9%) and Hispanic UBC (87.5%) groups achieved the lowest classification accuracies, which is not surprising given that neither an Asian nor Hispanic sample was included in the original Klales et al. [6] research. It is surprising, however, that both the Thai and UBC samples performed better or nearly as well as the validation included in Klales et al. [6] using the modern U.S. Bass Collection.

Combined sex classification accuracies in the validation were all above the 85% threshold for acceptable levels of sex classification [1]. These results are encouraging, as it is not always possible to broadly apply a method that was developed on a specific sample to different groups. The population specific recalibration equations resulted in a total correct classification mean increase of 4.1% with a range of increase between 1.1% (U.S. Black) and 10.7% (Thai). The dramatic increase in the correct classification of the Thai males is likely due to the consistently lower trait scores compared to the males included in the original Klales et al. [6] study. Sex bias was reduced with the population specific equations (mean = −2.64 and sd = 2.56) compared to the Klales et al. [6] original equation (mean = 9.0 and sd = 5.0).

When the global recalibration equation was applied to each individual population, classification accuracies had a mean improvement of 3.7% with a range of increase between 0.9% (U.S. Black) to 10.0% (Thai), which is comparable to the population specific equations. Furthermore, the magnitude and variability of sex bias was greatly reduced using a global recalibration (mean = −2.75 and sd = 1.42) compared to the Klales et al. [6] original equation (mean = 9.0 and sd = 5.0), showing that the global equation can be applied to a wide range of populations. Interestingly, sex bias shifted from favoring females to males with the global recalibration, however this is likely a byproduct of larger male sample sizes.

While some geographic groups achieved comparably lower classification rates, it is important to note that all groups still achieved high accuracies. For example, the Thai sample performed poorly when using the original Klales et al. [6] method equation, but once analyzed using the global recalibration, achieved high classification rates that are comparable to the other groups and to results from craniometric analyses from the same country [13]. Interestingly, the UBC sample still yielded the lowest classification rates even when recalibrated using a population specific equation and when analyzed using the global recalibration. This is likely due to the fact that both UBC males and females display more scores of 3 than any other score for the MA and the overlap in score frequencies between males and females resulted in reduced correct classifications. Additionally, the smaller sample size and unconfirmed sex assignment in some of these unidentified individuals may also be impacting the lower classification accuracy (see Ref. [8]). Despite these challenges, classification accuracy for the UBCs was still well above the 85% threshold. Unlike craniometric comparisons between South Africans and North
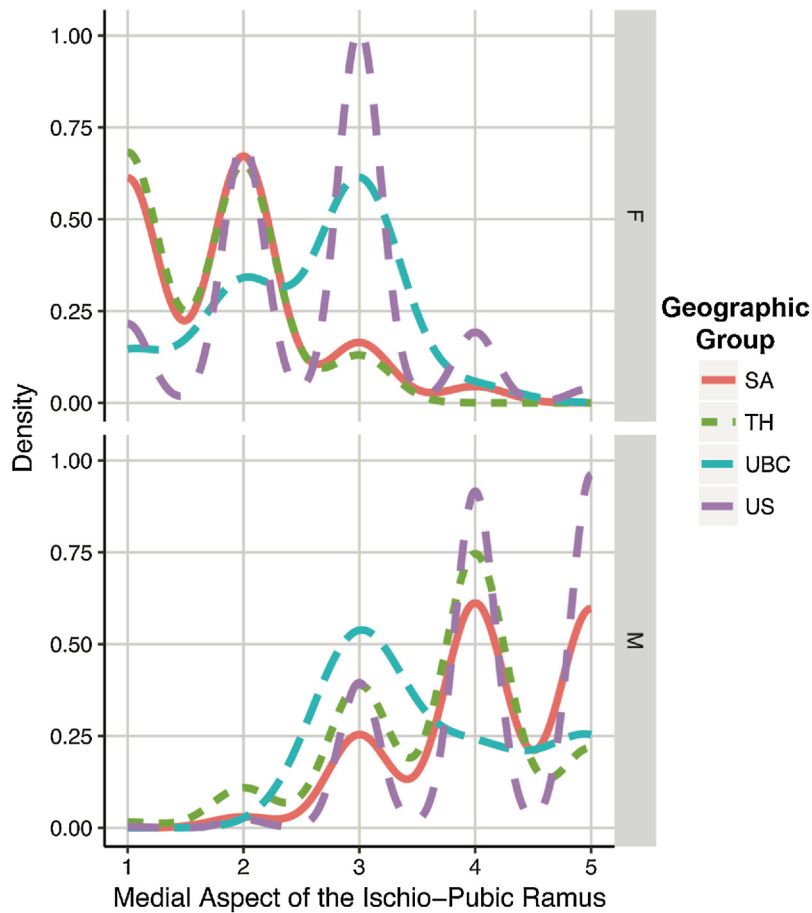
Fig. 3. Density plot of MA scores for each geographic group by sex where SA = South Africa, TH = Thailand, UBC = undocumented Hispanic border crossers, and US = United States.

Table 8
Total correct classifications (%) and sex bias of each geographic group using the original logistic regression equation supplied by Klales et al. [6].

Geographic group	N		Correct female Klales equation	Correct male Klales equation	Combined correct Klales equation	Sex bias
	F	M				
South Africa Black	50	50	100.0	86.0	93.0	14.0
South Africa White	50	50	98.0	90.0	94.0	8.0
South Africa Pooled Ancestry	100	100	99.0	88.0	93.5	11.0
Thai	45	96	100.0	82.3	87.9	17.7
Hispanic UBC	24	24	91.7	83.3	87.5	8.4
U.S. Black	254	294	97.6	93.9	95.6	3.7
U.S. White	401	579	97.3	93.4	90.7	3.9
U.S. Pooled Ancestry	654	873	97.4	91.8	94.2	5.6
Global	823	1092	97.6	90.4	93.5	7.2

Table 9
Total correct classifications (%) and sex bias of each geographic group using recalibrated population specific logistic regression equations.

Geographic group	N		Correct female recalibration	Correct male recalibration	Combined correct recalibration	Sex bias
	F	M				
South Africa Black	50	50	96.0	100.0	98.0	−4.0
South Africa White	50	50	96.0	96.0	96.0	0.0
South Africa Pooled Ancestry	100	100	96.0	98.0	97.0	−2.0
Thai	45	96	97.8	98.9	98.6	−1.1
Hispanic UBC	24	24	87.5	95.8	91.7	−8.3
U.S. Black	254	294	95.7	97.6	96.7	−1.9
U.S. White	401	579	94.5	96.0	95.4	−1.5
U.S. Pooled Ancestry	654	873	94.5	96.8	95.8	−2.3

Table 10

Total correct classifications (%) and sex bias of each geographic group using the global model (i.e., the model created using the entire sample).

Geographic group	N		Correct female global recalibration	Correct male global recalibration	Combined correct global recalibration	Sex bias
	F	M				
South Africa Black	50	50	96.0	100.0	98.0	−2.0
South Africa White	50	50	96.0	96.0	96.0	0.0
South Africa Pooled Ancestry	100	100	96.0	98.0	97.0	−2.0
Thai	45	96	100.0	96.8	97.9	3.2
Hispanic UBC	24	24	87.5	91.7	89.6	−4.2
U.S. Black	254	294	94.5	98.3	96.5	−3.8
U.S. White	401	579	94.5	96.0	95.4	−1.5
U.S. Pooled Ancestry	654	873	94.5	96.8	95.8	−2.3
Global	823	1092	94.8	96.8	95.9	−2.0

Table 11

Comparison of total correct classifications (%) from the original Klales et al. [6] equation, the population specific recalibration equations, and the global recalibration equation.

Geographic group	Klales et al. [6] external validation	Pop. specific recalibration	Global recalibration
South African Black	93.0	98.0	98.0
South African White	94.0	96.0	96.0
South African Pooled Ancestry	93.5	97.0	97.0
Thai	87.9	98.6	97.9
Hispanic UBC	87.5	91.7	89.6
U.S. Black	95.6	96.7	96.5
U.S. White	90.7	95.4	95.4
U.S. Pooled Ancestry	94.2	95.8	95.8
Global	93.5	–	95.9

Americans (see Ref. [14]), pelvic dimensions appear to be equally sexually dimorphic between these two populations.

It may be the case that the expression of nonmetric pelvic traits correlates with gross body size. If so, then populations with markedly less sexual size dimorphism, such as the Thai and Hispanic groups, would show more overlap in trait score expressions. While this trend is observed in the expression of nonmetric pelvic trait scores above, it is impossible to explicitly test the effect of gross size on the expression of the pelvic traits given the current dataset. Further, the Klales et al. [6] traits are to be judged in reference to the overall size of the specimen. It has been suggested by Guégan et al. [15] that sexual size dimorphism, as defined through statures, are correlated with maternal mortality and fecundity across different geographic populations. Populations that have greater maternal mortality-x-fecundity rates have lower values of sexual size dimorphism. Given that the expression of pelvic traits contribute to the biomechanical aspects of childbirth, the same trend in decreasing sexual dimorphism may hold.

The greater classification accuracies obtained in the current study as compared to those initially reported by Klales et al. [6] may be due to several reasons: (1) population specific equations; (2) larger sample sizes, and; (3) and mostly modern collections. Applying population specific equations allows higher accuracy rates to be achieved by reducing the effects introduced by less dimorphic populations or by populations that follow different trends in the expression of the suite of nonmetric traits. However, as demonstrated, the global recalibration equation performs just as well as population specific equations. The larger sample sizes reduce the potential impact of outliers on classification. Klales [16] demonstrated that these nonmetric pelvic traits have undergone secular changes wherein modern females, in particular, are showing greater expressions of lower trait scores in comparison to historic females, and modern males are showing somewhat higher trait scores in comparison to historic males. These shifting trait frequencies therefore allow for greater correct classification using modern samples.

The classification accuracies obtained by the population specific recalibrated equations are virtually identical to the classification accuracies obtained by the global recalibrated equation for most groups (Table 11). However, this fact is not especially surprising for two main reasons. The first is that the population specific individuals are also included in the global dataset. The second is that, while there may be geographic differences in the frequency distributions of trait scores among groups, these differences do not necessarily impact classification. For example, the mean VA score for Thai males is 3.84 and for U.S. Blacks is 4.27 and while these result in significant differences among trait frequencies, the individuals still classify as male. Other studies examining the applicability of the Klales et al. [6] method on different populations have also shown improvements in classification using recalibrated equations as opposed to the original equation [8,9], yet the present research shows that a general global equation can be used instead. Given the high classification accuracy of the global population regression equation, population specific equations are not necessary for the Klales et al. [6] pelvic traits. Practitioners can accurately predict sex in diverse worldwide populations using the new global equation provided. However, if a practitioner prefers to only include data from specific populations, they can do so an upcoming, freely available database being developed [17].

5. Conclusions

The Klales et al. [6] revision of the Phenice [3] method is externally valid and the original equation can be used to correctly estimate sex at rates much greater than chance. However, the recalibration global equation that takes into account many geographic areas based on larger sample sizes improves classification accuracy and reduces sex bias. Given that the global equation accuracy rates are as high as the population specific equations and sex bias is low, the global equation can be used for sex estimation in diverse, worldwide populations.

Acknowledgments

the curators at each of the institutions from which data were collected.

References

[1] Research Methods in Human Skeletal Biology, in: E.A. DiGangi, M.K. Moore (Eds.), Elsevier Academic Press, Oxford, 2013.

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

[3] T.W. Phenice, A newly developed visual method for sexing the os pubis, Am. J. Phys. Anthropol. 30 (1969) 297–302.

[4] T.D. Stewart, Essentials of Forensic Anthropology, Especially as Developed in the United States, first ed., CC Thomas, Springfield, 1979.

[5] W.M. Krogman, M.Y. Isçan, The Human Skeleton in Forensic Medicine, second ed., CC Thomas, Springfield, 1986.

[6] A.R. Klales, S.D. Ousley, J.M. Vollner, A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods, Am. J. Phys. Anthropol. 149 (2012) 104–114.

[7] R.L. Jantz, S.D. Ousley, FORDISC 3.0. Computer Program, University of Tennessee Knoxville, Knoxville, TN, 2005.

[8] J.A. Gómez-Valdés, A.M. Garmendia, L. García-Barzola, G. Sánchez-Mejorada, C. Karam, J.P. Baraybar, A.R. Klales, Recalibration of the Klales et al. 2012 method of sexing the human innominate for Mexican populations, Am. J. Phys. Anthropol. 162 (3) (2017) 600–604.

[9] A.R. Klales, S.J. Cole, Improving nonmetric sex classification for Hispanic individuals, J Forensic Sci (2016), doi:http://dx.doi.org/10.1111/1556-4029.13391.

[10] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2009.

[11] R Development Core Team, R: A Language and Environment for Statistical Computing Version 3.3.2., R Foundation for Statistical Computing, Vienna Austria, 2016.

[12] S.J. Press, S. Wilson, Choosing between logistic regression and discriminant Analysis, J. Am. Stat. Assoc. 73 (1978) 699–705.

[13] P. Mahakkanukrauh, A. Sinthubua, S. Prasitwattanaseree, S. Ruengdit, P. Singsuwan, S. Praneatpolgrang, P. Duangto, Craniometric study for sex determination in a Thai population, Anat. Cell Biol. 48 (2015) 275–283.

[14] E.N. L'Abbé, M.W. Kenyhercz, K.E. Stull, Application of Fordisc 3.0 to explore differences among crania of North American and South African blacks and whites, J. Forensic Sci. 58 (2013) 1579–1583.

[15] J.F. Guégan, A.T. Teriokhin, F. Thomas, Human fertility variation, size-related obstetrical performance and the evolution of sexual stature dimorphism, Proc. R. Soc. Lond. 267 (2000) 2529–2535.

[16] A.R. Klales, Secular change in morphological pelvic traits used for sex estimation, J. Forensic Sci. 61 (2016) 295–301.

[17] A.R. Klales, An Interactive Morphological Database for Estimating Sex in Modern Adults, National Institute of Justice, 2015 grant # 2015-DN-BX-K-014.