Forensic Science International 277 (2017) 259.e1–259.e8

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

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

Article history:
Received 7 December 2016
Received in revised form 26 April 2017
Accepted 1 May 2017
Available online 8 May 2017

Keywords:
Sexual dimorphism
Validation study
Recalibration
Sex estimation
Logistic regression
Forensic Anthropology Population Data

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.

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

☆ This research was partially funded by National Institute of Justice Grant 2015-DN-BX-K014. Opinions or points of view expressed in this research represent a consensus of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or the National Institute of Justice. Any products and manufacturers discussed are presented for informational purposes only and do not constitute product approval or endorsement by the U.S. Department of Justice or the National Institute of Justice.

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

http://dx.doi.org/10.1016/j.forsciint.2017.05.001
0379-0738/

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.

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

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 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