Forensic Science International 261 (2016) 154–158

Contents lists available at ScienceDirect

Forensic Science International

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

Do group-specific equations provide the best estimates of stature?

John Albanese *, Stephanie E. Osley, Andrew Tuck

Department of Sociology and Anthropology, University of Windsor, Windsor, Ontario, Canada

ARTICLE INFO

Article history:
Received 20 July 2015
Received in revised form 7 February 2016
Accepted 9 February 2016
Available online 17 February 2016

Keywords:
Forensic anthropology
Stature estimation
Long bones
Fordisc
Secular changes
Sex
Race/ancestry

ABSTRACT

An estimate of stature can be used by a forensic anthropologist with the preliminary identification of an unknown individual when human skeletal remains are recovered. Fordisc is a computer application that can be used to estimate stature; like many other methods it requires the user to assign an unknown individual to a specific group defined by sex, race/ancestry, and century of birth before an equation is applied. The assumption is that a group-specific equation controls for group differences and should provide the best results most often. In this paper we assess the utility and benefits of using group-specific equations to estimate stature using Fordisc. Using the maximum length of the humerus and the maximum length of the femur from individuals with documented stature, we address the question: Do sex-, race/ancestry- and century-specific stature equations provide the best results when estimating stature? The data for our sample of 19th Century White males ($n = 28$) were entered into Fordisc and stature was estimated using 22 different equation options for a total of 616 trials: 19th and 20th Century Black males, 19th and 20th Century Black females, 19th and 20th Century White females, 19th and 20th Century White males, 19th and 20th Century any, and 20th Century Hispanic males. The equations were assessed for utility in any one case (how many times the estimated range bracketed the documented stature) and in aggregate using 1-way ANOVA and other approaches. This group-specific equation that should have provided the best results was outperformed by several other equations for both the femur and humerus. These results suggest that group-specific equations do not provide better results for estimating stature while at the same time are more difficult to apply because an unknown must be allocated to a given group before stature can be estimated.

1. Introduction

An estimate of stature, along with estimates of age and sex, can be used by a forensic anthropologist to assist in the identification of an unknown individual when human skeletal remains are recovered. After Trotter and Gleser [1–3] published a series of papers, the approach for stature estimation has been to develop equations that are group-specific where group membership is based on combinations of sex, race, ancestry, continental origin, nationality, year of birth, and other criteria (for example, [4–13,27–29]). However, there is some evidence that this group-specific approach can be problematic for various practical and theoretical reasons [14,31–33].

Fordisc is a computer application that can be used to estimate stature, as well as "race" or ancestry, and sex [15]. Fordisc, currently in version 3.1, is an automated version of many of these traditional methods with some changes to the reference samples used to generate equations for constructing a biological profile of an unknown individual. When estimating stature, the software requires the user to select sex-, race- and century-specific options. The rationale is that group-specific equations should provide the most accurate and most useful results most often for stature estimation. An option for non-specific equations is possible in Fordisc but it is described in the *Fordisc Help File* (Version 1.35, http://math.mercyhurst.edu/~sousley/Fordisc/) as a second-best approach that should be used when there are no other options. The most significant practical limitation of this approach is that an unknown individual must first be allocated to one of these sex-race-century groups before the "correct" equation is applied.

In this paper we test all the various parameters for group specificity required to estimate stature using humerus and femur data collected from individuals in the Terry Collection for whom stature is documented. We use Fordisc because it allows for the

* Corresponding author at: Department of Sociology, Anthropology, and Criminology, 401 Sunset Avenue, University of Windsor, Windsor, Ontario, N9B 3P4 Canada. Tel.: +1 519 253 3000x3973; fax: +1 519 971 3621.
E-mail address: albanese@uwindsor.ca (J. Albanese).

http://dx.doi.org/10.1016/j.forsciint.2016.02.019
0379-0738/

user to easily toggle through a number of group-specific equations, but the goal is not necessarily to test the utility of the software. The goal is to test whether sex-, race- and century-specific stature equations provide the **best** results when estimating stature.

2. Materials and methods

A sample ($n = 28$) was selected from the Terry Collection to include only White males with years of birth before 1870. The birth years were selected so that the entire growth and development period was completed before the start of 20th century. There are 134 individuals in the Terry Collection within these parameters for whom there are stature data. Approximately 20% of these 134 individuals were randomly selected to be included in this research. The maximum length of the humerus and the maximum length of the femur were collected from the left side by two of us on two separate occasions following the measurement description recommended for Fordisc [16]. Testing was conducted to ensure that data were collected consistently. Intra- and inter-observer measurement errors were assessed using the absolute differences between measurements. In over 95% of the cases, the absolute difference for each observer and between observers was less than 1 mm or 0.5%. In a few cases where the bone on the left side was damaged or was affected by trauma, data were collected from the right side. Both femur and humerus data were used to assess any differences between the upper and lower limb.

Data for this research were collected from the Terry Collection because it is one of the few identified skeletal collections with reliable, documented stature for a large number of individuals [17]. Stature data were collected by Robert Terry and his assistants using a standardized protocol for positioning, measuring and photographing the cadaver in a "standing" position that closely approximated living stature [1,34]. For some cases it was not always possible to accurately reproduce living stature from the cadavers. Those cases were easily excluded from our sample using Terry's detailed notes and photographs of cadavers.

The humerus and the femur data for our sample of 19th Century White males were entered into Fordisc and stature was estimated using 22 different equation options for a total of 616 trials: 19th and 20th Century Black males, 19th and 20th Century Black females, 19th and 20th Century White females, 19th and 20th Century White males, 19th and 20th Century Any, and 20th Century Hispanic males. There is no 19th Century Hispanic option. The "Any" option is better described as "all" because it includes all racial groups and both sexes in a century-specific equation. The 90% confidence interval was used for this analysis. If the race, century and sex assumptions are true, then the 19th Century White male equation should consistently provide the best predictions of stature for the test sample of 28 White males who were born and had their entire growth and development period in the 19th century. The latest version of Fordsic was used: Version 3.1, build 307, released October 16, 2015.

The accuracy and utility of the equations were assessed several ways. First, we assessed the equations using a simple count to determine the utility of the equations for providing useful information in a forensic investigation. We counted the number of times the actual stature was bracketed by the predicted range calculated using the 90% confidence interval, which provides the narrowest estimated range. We calculated the mean difference (MD) and mean absolute deviation (MAD). The MD is the average of the difference of the estimated stature minus the documented stature. One limitation of the MD is that some positive and negative errors may cancel each other out. One clear benefit of the MD is that it can be used to identify a tendency or bias to overestimate or underestimate stature using a specific equation. When calculated as the estimated stature minus the documented stature as in this research, a positive MD suggests a tendency to overestimate documented stature and a negative MD suggests a tendency to underestimate documented stature. The MAD is the mean of the absolute value of the documented stature minus predicted stature. In contrast to the MD, which is always equal to or lower than the MAD, the MAD is a better measure of overall error because it is the mean of the absolute difference, and positive and negative errors do not cancel each other out. Together the MAD and MD provide a good measure of precision (how close the estimate is to the documented stature), and bias (trend towards overestimation or underestimation), respectively. Both the MDs and MADs for each equation were tested using a one-sample *t*-test to assess whether they are significantly different than zero. The MADs for each equation were also tested for significant differences from each other using 1-way ANOVA. Finally, the estimated mean for each equation was tested for significant differences from the mean of the documented statures using 1-way ANOVA and Tukey HSD *post hoc* to group means into homogeneous subsets.

3. Results

The MAD and MD are listed for each equation for the humerus and the femur in Table 1. The equations are listed from best to

Table 1
Mean difference (MD) and mean absolute difference (MAD) for equations using the humerus and femur ranked by overall utility. The 19th Century White male equations calculated by Fordsic 3.1 do not provide the best results when tested on a sample ($n = 28$) of 19th Century White males from the Terry Collection.

	Humerus			Femur							Humerus and Femur		
	No	Yes	%	MAD*	MD	No	Yes	%	MAD*	MD	No	Yes	%
20th Century Black Male	0	28	100	3.58	0.78	0	28	100	4.31	−3.90**	0	56	100
20th Century Black Female	0	28	100	3.73	0.23	0	28	100	4.02	−3.78**	0	56	100
20th Century White Female	0	28	100	3.66	−1.97$^{\&}$	0	28	100	3.56	−3.08**	0	56	100
20th Century Any (All)	0	28	100	3.68	1.01	0	28	100	2.55	−0.93	0	56	100
20th Century White Male	1	27	96.4	3.89	2.32$^{\#}$	0	28	100	2.74	1.20$^{\&}$	1	55	98.2
20th Century Hispanic Male	1	27	96.4	3.58	1.23	1	27	96.4	2.66	0.39	2	54	96.4
19th Century Any (All)	1	27	96.4	3.95	−2.35$^{\wedge}$	1	27	96.4	3.68	−3.02**	2	54	96.4
19th Century White Male	2	26	92.9	3.64	−1.49	1	27	96.4	2.43	−0.54	3	53	94.6
19th Century Black Male	2	26	92.9	3.83	−2.10$^{\&}$	1	27	96.4	3.24	−2.64**	3	53	94.6
19th Century White Female	2	26	92.9	3.84	−1.95$^{\&}$	3	25	89.3	4.20	−3.95**	5	51	91.0
19th Century Black Female	8	20	71.4	4.84	−4.45**	12	16	57.1	5.96	−5.89**	20	36	64.3

* All MADs are significantly different from zero at $p < 0.0001$ level.
$^{\&}$ Significantly different from zero at $p < 0.05$ level.
$^{\wedge}$ Significantly different from zero at the $p < 0.01$ level.
$^{\#}$ Significantly different from zero at $p < 0.001$ level.
** Significantly different from zero at the $p < 0.0001$ level.