Performance of FORDISC 2.0 using inaccurate measurements

Carlos J Zambrano, BA, Adam Kolatorowicz, BS*, Stephen P. Nawrocki, PhD, Archeology and Forensics

ABSTRACT

Once decomposition has progressed beyond the point of recognizing the soft tissue indicators of an individual's sex and ancestry ("racial" affiliation), it becomes the job of the forensic anthropologist to make these determinations. The determination of sex and ancestry are part of the biological profile that the anthropologist constructs during a skeletal analysis. The anthropologist uses both non-metric and metric analysis to determine these characteristics. However, the determination of ancestry is not always a clear or simple task. In recent years the use of metric analysis has become more prominent due to the advancement of computers and use of statistical software. FORDISC 2.0, distributed by the University of Tennessee at Knoxville, is a program that facilitates the collection and use of metric data in discriminant function analysis to determine sex and ancestry from the skeleton. The program was developed using the Forensic Data Bank, which is an electronic accumulation of data from modern forensic cases from around North America. The use of FORDISC 2.0 is prevalent in the field of forensic anthropology and at times it may be used by individuals having only a cursory knowledge of measurement techniques and morphometrics. This poster will examine the outcome of poor data input into FORDISC 2.0 to determine how well the program performs with inaccurate data.

The 24 standard FORDISC 2.0 cranial measurements were taken on 4 carefully-selected crania that had been positively identified or had enough soft tissues at autopsy to make a determination of sex and ancestry. The measurements were entered into the program to classify each specimen using the "White" male and "Black" male reference groups. Two of the specimens were chosen because they were classified as strongly belonging to one of the reference groups, while the others were weakly classified. The measurements with the highest relative weights in the discriminant functions were then selected and manipulated by the addition and subtraction of 1 to 3 mm from the original measurement, and the changes in probabilities and classification were recorded. Measurements were changed all at once and in isolation. Results demonstrate that individuals who are classified strongly into a reference group remain strongly classified in that same group even with significantly altered measurements. Individuals with a weak classification into a reference group can be subject to a significant change by the addition or subtraction of even as little as 1 mm to a single measurement. It is generally accepted that interobserver error among trained individuals can reach 2 to 3 mm. Depending on the measurement and the morphology of the subject, errors of this magnitude can have a significant influence on the discriminant function analysis. These results demonstrate the necessity of receiving proper training in the collection of metric data as well as the need to check instrumentation used to collect metric data. These findings in no way suggest the abandonment of the use of FORDISC 2.0 or other forms of discriminant function analysis, however, we do suggest that only individuals who have received a substantial amount of training should be using FORDISC 2.0 in a professional arena.

INTRODUCTION

The advent of the personal computer has had a significant impact on forensic anthropology, permitting a degree of data storage and analysis that only a few decades ago was unavailable to the typical practitioner. As such, anthropology's use of metric methods in the determination of sex, ancestry, stature, and age has increased substantially. While discrete (non-metric) approaches still have their value, metric methods now provide an independent and (perhaps) more rigorous approach, permitting a quantification of the likelihood of accurate assessment and a higher level of replicability between researchers and studies.

The broad availability of software such as FORDISC 2.0 to analyze human skeletal remains, however, may have its drawbacks. One of us (SPN) has been approached on a number of occasions by police, coroners, and pathologists wanting to obtain this or other computer programs so that they could, ostensibly, "save the anthropologist some work" and do their own analyses. Some of these individuals may not have had a qualified forensic anthropologist in their regions, or may have been forensic scientists with medical or clinical specialties (such as craniofacial reproduction or biomechanical analysis) and who felt that their familiarity with the human skeleton would facilitate their use of anthropological software.

The obvious question raised by this well-intentioned interest in quantitative anthropological methods is whether a passing or cursory interest and knowledge of human skeletal anatomy is sufficient background for a non-anthropologist to use these programs effectively. Anthropometric measurement looks easy but requires significant training and experience to produce valid and replicable results. Even among trained anthropologists there will likely be slight differences in measurements obtained on the same specimen, reflecting random interobserver error, individual stylistic differences in using the instruments and defining anthropometric landmarks, and possibly variations in the skeletal material itself due to temperature and humidity fluctuations.

This study was designed to determine how robust metric methods are to minor errors in measurement. We define "minor errors" as those that would be expected to occur occasionally both within and between experienced anthropometrists: 1 to 2 millimeters for straightforward measurements, or perhaps 1 to 3 millimeters for more complex measurements. How will the results of the discriminant analyses conducted by FORDISC 2.0 be influenced by minor errors?

MATERIALS & METHODS

The 4 specimens used in this study were chosen according to the following criteria:

(1) all had been analyzed as part of the forensic anthropology caseload of the UIAFL during the last 10 years;

(2) all were positively identified, fully-adult males with known ancestry (3 European-American & 1 African-American);

(3) each specimen was in excellent condition, preserving all 24 cranial measurements (excluding the mandible) specified by FORDISC 2.0;

(4) each measurement for each specimen was taken directly or checked by a trained anthropometrist (SPN);

(5) 2 specimens were **strongly classified** into their correct ancestral category by FORDISC 2.0, with posterior probablilties of 1.00 in the 2-way (white male vs. black male) discriminant analysis using the correct measurements (classification data is presented in Table 1);

(6) 2 specimens were **weakly or incorrectly classified** as to ancestry by FORDISC 2.0, with posterior probabilities of <0.65 in the 2-way (white male vs. black male) discriminant analysis using the correct measurements.

TABLE 1: Demographic Data on the Study Specimens, with Initial FORDISC 2.0 Results Using the Correct Measurements in a 2-way (white male vs. black male) Discriminant Analysis.

Specimen Number	Ancestry/ Sex	Age	Predicted Group	Posterior Prob.	Typicality Prob.	Other Group	Posterior Prob.	Typicality Prob.
UI-10-96	WM	19	BM*	0.511**	0.209	WM	0.489	0.206
UI-07-01	WM	28	WM	1.000	0.493	BM	0.000	0.017
UI-31-01	WM	73	WM	1.000	0.687	BM	0.000	0.028
UI-06-03	BM	55	BM	0.637**	0.083	WM	0.363	0.065

* denotes an incorrect initial classification by FORDISC 2.0
** denotes a "weak" initial classification by FORDISC 2.0

In each of the 2-way discriminant analyses, 3 variables (Figure 1) consistently loaded as the most important discriminators, as reflected by relative weights higher than 13.0:

-- basion-nasion length (BNL)
-- basion-prosthion length (BPL)
-- nasal breadth (NLB)

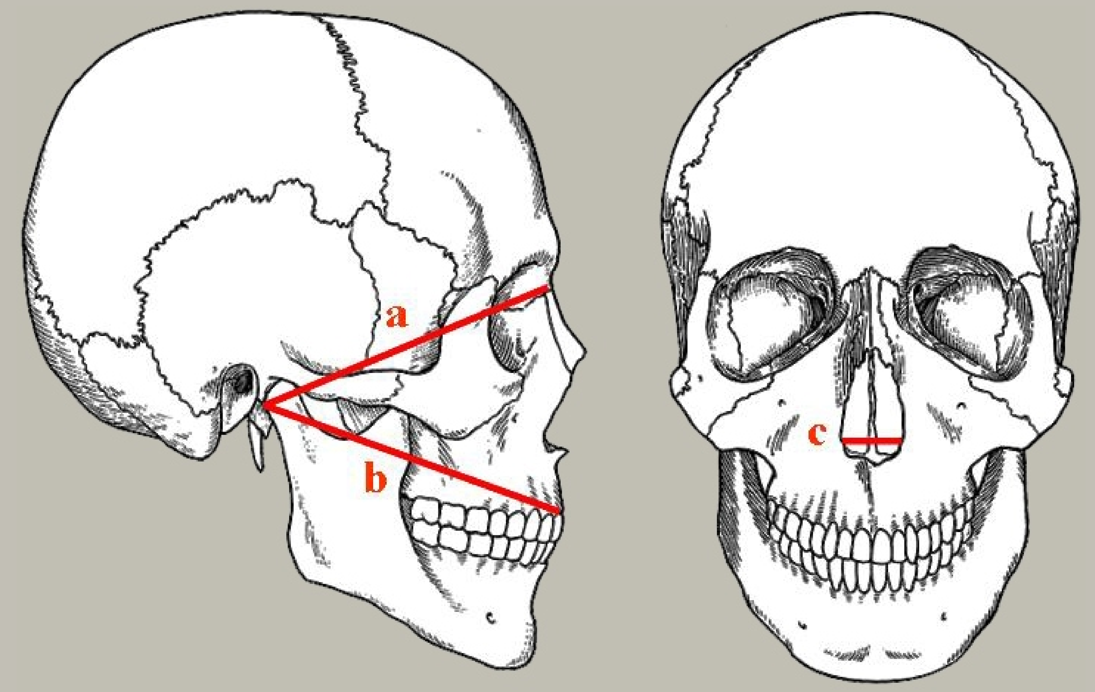

Figure 1: Measurements Manipulated in Study: a = BNL, b = BPL c = NLB.

None of the other 21 measurements ever had relative weights higher than 8.5. Therefore, we manipulated these 3 measurements in the following fashion for each specimen:

(1) BNL was increased by 1mm increments up to +3 mm. A 2-way discriminant analysis was run after each change in value and the posterior and typicality probabilities were recorded.

(2) BNL was decreased by 1mm increments down to -3 mm. A 2-way discriminant analysis was run after each change in value and the posterior and typicality probabilities were recorded.

(3) Steps 1and 2 were repeated with BPL and NLB.

(4) BNL, BPL, and NLB were increased by 1mm increments as a group up to +3 mm. A 2-way discriminant analysis was run after each change in value and the posterior and typicality probabilities were recorded.

(5) BNL, BPL, and NLB were decreased by 1mm increments as a group down to -3 mm. A 2-way discriminant analysis was run after each change in value and the posterior and typicality probabilities were recorded.

A total of 24 different numerical combinations were subjected to separate discriminant analyses for each of the 4 specimens.

RESULTS

TABLE 2: FORDISC 2.0 Classification Results For UI-10-96 Using the Correct Measurements and the Manipulated Measurements in a 2-way Discriminant Analysis

	BNL				BPL				NLB			
	White Male		Black Male		White Male		Black Male		White Male		Black Male	
mm changed	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.
-3	0.103	0.056	**0.897**	0.138	**0.769**	0.261	0.231	0.173	**0.939**	0.526	0.061	0.245
-2	0.188	0.099	**0.812**	0.175	**0.687**	0.258	0.313	0.198	**0.860**	0.456	0.140	0.272
-1	0.320	0.152	**0.680**	0.200	**0.591**	0.239	0.409	0.211	**0.708**	0.340	0.292	0.259
0	0.489	0.206	**0.511**	0.209	0.489	0.206	**0.511**	0.209	0.489	0.206	**0.511**	0.209
+1	**0.660**	0.251	0.340	0.201	0.387	0.164	**0.613**	0.194	0.274	0.095	**0.726**	0.139
+2	**0.797**	0.281	0.203	0.176	0.294	0.119	**0.706**	0.167	0.130	0.031	**0.870**	0.073
+3	**0.889**	0.289	0.111	0.140	0.216	0.078	**0.784**	0.132	0.056	0.007	**0.944**	0.028

TABLE 3: FORDISC 2.0 Classification Results For UI-06-03 Using the Correct Measurements and the Manipulated Measurements in a 2-way Discriminant Analysis

	BNL				BPL				NLB			
	White Male		Black Male		White Male		Black Male		White Male		Black Male	
mm changed	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.
-3	0.064	0.018	**0.936**	0.065	**0.665**	0.023	0.335	0.016	**0.902**	0.063	0.098	0.022
-2	0.122	0.032	**0.878**	0.079	**0.567**	0.036	0.433	0.032	**0.785**	0.078	0.215	0.044
-1	0.219	0.049	**0.781**	0.085	0.463	0.051	**0.537**	0.054	**0.591**	0.079	0.409	0.067
0	0.363	0.065	**0.637**	0.083	0.363	0.065	**0.637**	0.083	0.363	0.065	**0.637**	0.083
+1	**0.536**	0.077	0.464	0.072	0.274	0.076	**0.726**	0.113	0.184	0.043	**0.816**	0.082
+2	**0.701**	0.082	0.299	0.057	0.196	0.081	**0.801**	0.142	0.082	0.022	**0.918**	0.067
+3	**0.827**	0.079	0.173	0.039	0.141	0.079	**0.859**	0.163	0.034	0.008	**0.966**	0.043

TABLE 4: FORDISC 2.0 Classification Results For UI-10-96 and UI-06-03 Manipulating the Group of Measurements (BNL,BPL,NLB) Simultaneously in a 2-way Discriminant Analysis.

	UI-10-96				UI-06-03			
	White Male		Black Male		White Male		Black Male	
mm changed	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.	Post. Prob.	Typ. Prob.
-3	**0.866**	0.356	0.134	0.196	**0.794**	0.009	0.206	0.005
-2	**0.773**	0.369	0.227	0.254	**0.671**	0.026	0.329	0.019
-1	**0.644**	0.312	0.356	0.259	**0.519**	0.049	0.481	0.048
0	0.489	0.206	**0.511**	0.209	0.363	0.065	**0.637**	0.083
1	0.336	0.097	**0.664**	0.128	0.232	0.063	**0.768**	0.104
2	0.211	0.030	**0.789**	0.054	0.138	0.045	**0.862**	0.099
3	0.124	0.005	**0.876**	0.015	0.078	0.022	**0.922**	0.070

UI-07-01 and UI-31-01 showed no changes in their classification after the variables were manipulated. The posterior probabilities remained at or near the 1.000 level.

Discussion

The specimens that were initially strongly classified (UI07-01 and UI-31-01) by FORDISC 2.0 showed no changes in classification after the variables were altered. This illustrates that individuals with a strong initial classification are not affected by any small subsequent alterations of the values of the variables.

The specimens that were initially weakly classified (UI-10-96 and UI-06-03) by FORDISC 2.0 showed changes in classification after the variables were altered by as little as 1 mm. As the value of the variable was increased or decreased the classification became stronger in that direction. This occurred when variables were altered independent of one another as well as when the variables were altered as a group. The change in classification and strength of the probabilities of the specimens are clearly seen in Tables 2, 3, and 4. UI-10-96 (Table 2) was initially misclassified as a "black male". When BNL was increased by 1 mm the specimen was correctly classified as a "white male". As the variable was increased by 1 mm the specimen continued to be classified as a "white male". UI-06-03 (Table 3) was initially classified as a "black male". When BNL was increased by 1 mm the specimen was misclassified as a "white male". As the value of the variable was increased so did the strength of the misclassification. For both UI-10-96 and UI-06-03, when BPL and NLB were increased by 1mm increments, the specimen had an even stronger "black male" classification. Table 4 shows the classification of the specimens after the suite of measurements were altered. A decrease in the values of the variables shifted the classification toward "white males" and an increase in the values shifted the classification towards "black males". This is not surprising as these two variables are direct measurements of non-metric traits which correspond to prognathism (BPL) and nasal breadth (NLB). Individuals of African ancestry tend to have a prognathic face and a wide nasal aperture whereas individuals of European ancestry tend to have an orthognathic face and a narrow nasal aperture. **WHAT ABOUT BNL AND ANCESTRY?**

The landmarks which define BNL, BPL, and NLB are relatively easy to identify on a cranium. Other measurements used by FORDSIC 2.0 are defined by landmarks which can be more difficult to locate (e.g., mastoid height, biorbital width, and nasal height). If the three measurements used in this study can have such a large impact on the outcome of a FORDISC 2.0 analysis when they are altered by ±1 mm, then the other measurements which are more difficult to take may introduce more error into the outcome.

CONCLUSION

FORDISC 2.0 is a powerful statistical tool for determining ancestry. However, it is clearly demonstrated above that only an anthropologist trained in non-metric analysis and anthropometry should be using this program. Even the slightest error (1 mm) in the measurements may trigger a misclassification of a specimen by FORDISC 2.0. In addition to human error one must also consider the condition and calibration of the instruments used to take the measurements. Worn tips on spreading calipers may result in a measurement that is less than the actual value. Even trained anthropometrists must use caution when gathering metric data and entering it into FORDISC 2.0.