

# Sex Determination of Human Skeletal Populations Using Latent Profile Analysis

Nicholas V. Passalacqua,<sup>1\*</sup> Zhen Zhang,<sup>2</sup> and Steven J. Pierce<sup>2</sup>

<sup>1</sup>JPAC Central Identification Laboratory, Joint Base Pearl Harbor-Hickam, HI

<sup>2</sup>Center for Statistical Training & Consulting, Michigan State University, East Lansing, MI

**KEY WORDS** sex determination; latent profile analysis; paleodemography

**ABSTRACT** Accurately estimating biological sex from the human skeleton can be especially difficult for fragmentary or incomplete remains often encountered in bioarchaeological contexts. Where typical anatomically dimorphic skeletal regions are incomplete or absent, observers often take their best guess to classify biological sex. Latent profile analysis (LPA) is a mixture modeling technique which uses observed continuous data to estimate unobserved categorical group membership using posterior probabilities. In this study, sex is the latent variable (male and female are the two latent classes), and the indicator variables used here were eight standard linear measurements (long bone lengths, diaphyseal and articular breadths, and circumferences). Mplus (Muthén and Muthén: Mplus user's guide, 6th ed. Los Angeles: Muthén & Muthén, 2010) was used to obtain maximum likelihood estimates for latent class membership from a known sample of

individuals from the forensic data bank (FDB) (Jantz and Moore-Jansen: Database for forensic anthropology in the United States 1962–1991, Ann Arbor, MI: Interuniversity Consortium for Political and Social Research, 2000) ( $n = 1,831$ ), yielding 87% of correct classification for sex. Then, a simulation extracted 5,000 different random samples of 206 complete cases each from the FDB (these cases also had known sex). We then artificially imposed patterns of missing data similar to that observed in a poorly preserved bioarchaeological sample from Medieval Asturias, Spain ( $n = 206$ ), and ran LPA on each sample. This tested the efficacy of LPA under extreme conditions of poor preservation (missing data, 42%). The simulation yielded an average of 82% accuracy, indicating that LPA is robust to large amounts of missing data when analyzing incomplete skeletons. *Am J Phys Anthropol* 151:538–543, 2013. © 2013 Wiley Periodicals, Inc.

The determination of biological sex from adult human skeletal remains is often a simple task, routinely performed when reconstructing demographic parameters from bioarchaeological remains, or when generating a biological profile for a forensically significant decedent. Although reference samples exist for many modern populations as to facilitate statistical support for sex assessment of unknown forensically significant remains (i.e., Fordisc: Ousley and Jantz, 1997), statistical rigor is often absent from sex determinations for groups of individuals encountered in bioarchaeological scenarios. Further, the accurate determination of sex from the human skeleton can be especially difficult in cases of fragmentary or incomplete sets of remains, as are often encountered in bioarchaeological contexts. This is what Konigsberg and Frankenberg (2007) refer to as “paleodemography under duress”, or where the typical anatomically dimorphic regions of a skeleton are incomplete or absent, and observers are often forced to make best guess estimates to classify biological sex based on whatever information is present.

Based on such a bioarchaeological scenario, Konigsberg and Frankenberg (2007) previously proposed using femoral circumference as a means of determining biological sex from skeletal remains recovered from the Mississippian Averbuch site (40DV60). In this case, sex determinations were based only on femoral circumference and classification of male/female was generated using a mixture model. These mixture model estimates were then compared to sex determinations based on the Phenice criteria (1969) with correspondence of the two classifications occurring in ~81% of cases ( $n = 376$ ).

Although innovative, Konigsberg and Frankenberg's univariate approach relies on the presence of at least one femoral midshaft and thus cannot be applied to skeletons missing this element. Furthermore, a measurement from a single bone potentially carries less information than a collection of measurements from multiple bones, and hence it may be possible to determine sex more accurately via multivariate statistical methods that take advantage of multiple osteometric measures.

This project has several goals: 1) to expand on the work of Konigsberg and Frankenberg (2007) by testing a similar method on a known sample under optimal conditions; 2) to generate error rates based on this known sample that can be used to guide future bioarchaeological applications; and 3) to mimic preservation rates and patterns observed in a poorly preserved skeletal sample on the known sample data, and run a new sex determination analysis to compare error rates under optimal

Grant sponsor: Department of Anthropology, The Michigan State University.

\*Correspondence to: Nicholas V. Passalacqua, JPAC Central Identification Laboratory, 310 Worcester Ave., Bldg. 45, Joint Base Pearl Harbor-Hickam, HI 96853-5530, USA.  
E-mail: passala5@gmail.com

Received 26 January 2013; accepted 24 April 2013

DOI: 10.1002/ajpa.22295  
Published online 11 June 2013 in Wiley Online Library (wileyonlinelibrary.com).

versus suboptimal conditions. Further, we aimed to develop a multivariate method for sex determination that could use multiple osteological measurements to improve accuracy, but would still be viable when some of those measurements are missing owing to poor preservation or incomplete recovery of the skeleton.

To achieve these project goals, latent profile analysis (LPA) was employed (McCutcheon, 1987). Further, multiple LPA models were applied to known skeletal samples in both optimal and suboptimal conditions. These simulations allowed us to examine the utility of LPA for determining the sex of complete and incomplete skeletons under conditions where we can directly assess model accuracy and the model's sensitivity to missing data. By demonstrating high levels of accuracy under various controlled but realistic conditions, we can increase our confidence that applying LPA to bioarchaeological samples will yield trustworthy results and recommend its usage in future bioarchaeological investigations.

## MATERIALS

Two samples were used in this study, an undocumented sex sample from Medieval Asturias, Spain and a modern documented sex sample from the forensic data bank (FDB) (Jantz and Moore-Jansen, 2000; for more information, see also Ousley and Jantz, 1997). The undocumented bioarchaeological sample consists of  $n = 206$  adult individuals recovered from multiple Christian church cemeteries from Medieval Asturias, Spain. Epiphyseal closure was complete for all bones examined in the undocumented individuals included in this project. The bioarchaeological sample from Medieval Asturias was chosen as a comparative sample to the complete known database, because the individuals are poorly preserved and served as an excellent example of "paleodemography under duress" (Konigsberg and Frankenberg, 2007). All standard cranial and postcranial measurements were collected on the skeletal remains from the Asturian sample whenever possible by a single observer (Buikstra and Ubelaker, 1994). This sample provided information about realistic patterns of missing data that result from poor preservation or incomplete recovery of remains.

The modern, documented sample ( $n = 1,831$ ; 1,173 males and 658 females) was drawn from the FDB. These FDB individuals included all those with complete measurement data available (Ousley and Jantz, 1997; Jantz and Moore-Jansen, 2000). The FDB data were crucial to our simulation study because in addition to complete skeletal measurements, it also contained the known biological sex of each skeleton in the sample. Comparing the latent profile assigned to each skeleton to the actual biological sex associated with it in the FDB allowed us to determine how accurately the LPA was classifying sex.

To determine sex from human skeletal remains, eight standard linear measurements were used. These measurements were chosen as they were the most commonly available measurements based on the poorly preserved skeletons from Asturias, Spain. These measurements included: humeral maximum diameter at midshaft (HUMMXD\_L), humeral minimum diameter at midshaft (HUMMWD\_L), femoral maximum head diameter (FEMHDD\_L), femoral sagittal midshaft diameter (FEMMAP\_L), femoral transverse midshaft diameter (FEMMTV\_L), femoral midshaft circumference

(FEMCIR\_L), tibial maximum diameter at the nutrient foramen (TIBNFX\_L), and tibial transverse diameter at the nutrient foramen (TIBNFT\_L). For this study, only left side measurements were included from the documented individuals.

## METHODS

Discriminant function analysis and clustering analysis are both common multivariate methods capable of grouping cases on the basis of multivariate measurements. However, traditional clustering methods based on distance measures, such as  $K$ -means and hierarchical methods, cannot handle missing data well (Jackson, 1968) as the calculation of distance requires fully observed measurements. Furthermore, traditional clustering analysis is not based on an underlying statistical model, and hence it cannot provide the probability that a given skeleton falls into each of the groups identified in the cluster solution (McCutcheon, 1987; Baxter, 1994). Therefore, we turned to LPA, which is conceptually similar to clustering analysis, but is a model-based method wherein the profiles (i.e., groups) to be identified in the data are conceptualized as the levels of a latent (unobserved) categorical variable (McCutcheon, 1987). The observed measurements used to determine the latent profile for each case are called the indicator variables, and can be continuous, dichotomous, or ordinal variables, or a mixture of all three (McCutcheon, 1987).

LPA is a statistical method that uses a set of observed indicator variables to identify coherent subgroups from the sample population using posterior probabilities (Lazarsfeld and Henry, 1968; McCutcheon, 1987; Lubke and Muthén, 2005). LPA is specifically designed for the use of continuous variables to define a categorical latent variable; however, conceptually, LPA is very similar to latent class analysis and factor mixture models (Lubke and Muthén, 2005; Everitt et al., 2011). Here, the biological sex of an individual to be determined from skeletal remains may be conceptualized as a latent (unobserved) categorical variable. The values of that variable represent two distinct profiles of skeletons (i.e., male and female) and the indicators of the latent variable (sex) are any number of osteological measurements (here, the eight linear measurements listed above). The latent variable (sex) is classified based on the posterior probabilities generated by the model. LPA assumes normality of the indicators (measurements) and independence between indicators within classes.

## Simulation design

The documented FDB sample with no missing data (optimal data conditions) was first used to test the efficacy of LPA to determine sex under optimal conditions. First, a sample of  $n = 206$  individuals was randomly selected from the full FDB sample. LPA with two classes and eight indicators was then run on these individuals. Our preliminary analyses demonstrated that males tend to have larger values on each of the indicators than females. Therefore, we used an empirical rule to label each profile as either male or female. Whichever group had higher mean values on more than half of the eight indicator measurements was labeled male; the other profile was then labeled female. The accuracy of the latent profile sex determination was recorded by comparing the latent profile assigned to each skeleton with the actual sex recorded in the FDB. This process was then repeated

TABLE 1. Summary of missing data in each indicator variable, sorted by percent of missing values in decreasing order

	Missing		Valid <i>n</i>	Mean	Std. Deviation
	<i>n</i>	Percent			
TIBNFX_L	131	63.6	75	33.05	3.456
TIBNFT_L	130	63.1	76	22.91	2.552
HUMMWD_L	127	61.7	79	18.34	2.521
HUMMXD_L	126	61.2	80	22.31	2.276
FEMHDD_L	123	59.7	83	45.36	3.721
FEMCIR_L	118	57.3	88	88.35	9.776
FEMMTV_L	105	51.0	101	26.53	2.171
FEMMAP_L	101	49.0	105	29.17	3.071

5,000 times on different random samples from the full FDB to generate the baseline of accuracy for sex determination using LPA under optimal conditions.

The previous simulation process was then repeated using the FDB sample; however, this time, rates and patterns of missing data were imposed on each FDB subsample out of the 5,000 replicates using the rates and patterns of missing data from the undocumented bioarchaeological sample (suboptimal data conditions). This was accomplished by matching each skeleton in an FDB replicate sample with a randomly selected skeleton from the bioarchaeological sample, then imposing the bioarchaeological skeleton's pattern of missingness on the data from the FDB skeleton. In this way, real patterns of missing data were used rather than randomly censoring data entries, which should result in a more accurate reflection of method performance under suboptimal conditions, given that the mechanism triggering the missing data is unknown.

Two statistical summaries were used to describe the distributions for the LPA accuracy rates obtained via the simulation. First, median accuracy serves as a good estimate of the typical accuracy in each condition because the distributions are unimodal, but slightly skewed. Second, the 95% credible intervals describe the upper and lower bounds surrounding the middle 95% of accuracy rates associated with each condition; they highlight the ranges of accuracy rates one can expect when using the LPA method under each condition. Finally, the accuracy cost of missing data was quantified by subtracting the accuracy rate observed with complete data from the corresponding accuracy rate observed with incomplete data for each replicate sample (cost = incomplete – complete). Cost was also summarized with medians and 95% credible intervals.

### Software

The simulation procedure was implemented using two pieces of statistical software. The actual LPA models were estimated with Mplus 6.0 (Muthén and Muthén, 2010), whereas R (version 2.15.0) was used to do the sampling, summarize the results across replications, and generate graphs. Each replication was run by using the MplusAutomation (Hallquist, 2012) package for R to export the data to Mplus for analysis and then import the results to R. Our simulation code is available upon request.

### RESULTS

Before presenting results from the FDB sample, we first illustrate the scope and nature of the missing data

problem in the bioarchaeological sample from Medieval Asturias. Table 1 summarizes the extent of missing data in each indicator variable for that sample, considering each variable independently of the others. Figure 1 shows the frequency of each distinct pattern of missing data observed in this sample. These patterns of missing data were imposed on the FDB data by randomly matching FDB skeletons to bioarchaeological skeletons.

To demonstrate the normality of the indicator measurements and relationships between indicators in the FDB sample, we first created a scatter-plot matrix with histograms and Pearson's correlations for the full FDB sample (Fig. 2). The lower triangle shows the scatter plot for each pair of indicators and the upper triangle shows the Pearson's correlation coefficient between each pair, with values close to 1, indicating high (linear) correlation. The histograms in the diagonal panels illustrate that all of the indicators have unimodal, roughly symmetrical distributions, which supports the LPA normality assumption.

The first column in Figure 2 further demonstrates the sex differences for all indicators. *T*-tests show that males have higher means than females for all eight indicators (details omitted, but available upon request). This validates our empirical rule for determining which latent profile corresponds to each sex after partitioning each replicate sample via LPA.

LPA also assumes that the indicators are uncorrelated within each group of cases. Figure 3 shows the correlations between indicators after breaking the FDB sample down by sex. Although not completely independent, the correlations are substantially weaker after dividing the sample by sex. The upper triangle in Figure 3 shows correlations for males, whereas the lower triangle shows them for females. Most of the correlations are considerably smaller in Figure 3 than corresponding correlations in Figure 2, with only one exception: FEMMAP\_L and FEMCIR\_L remain strongly correlated ( $r = 0.77$  for male and  $r = 0.83$  for females). This does not severely undermine the within-class independence assumption, and hence we proceeded with the LPA simulations.

Figure 4 shows the key results of our simulations as histograms of the LPA classification accuracy rates. Under optimal conditions (i.e., with complete data on all eight indicators), LPA models attained a median accuracy of ~85% for sex determination and a 95% credible interval of [76 and 91%] accuracy (Fig. 4a). The accuracy of LPA under suboptimal conditions (i.e., missing data, ~42%) remained surprisingly high: Median accuracy for sex determination was ~83%, with a 95% credible interval of [72 and 89%] (Fig. 4b).

Figure 5 shows the cost of the missing data by showing the distribution for the difference in accuracy rates between the complete and the incomplete data sets. Accuracy decreased by a median of only ~2.2%, with a 95% credible interval of [-4.4 and 10%], even with ~42% of the data missing.

### DISCUSSION

LPA and latent class analysis have great potential for application in bioarchaeology owing to the undocumented nature of these collections and the ability for these latent methods to determine group membership using posterior probabilities. Previous research has used similar mixture models for species recognition in hominoid samples (Kramer and Konigsberg, 1999) and to



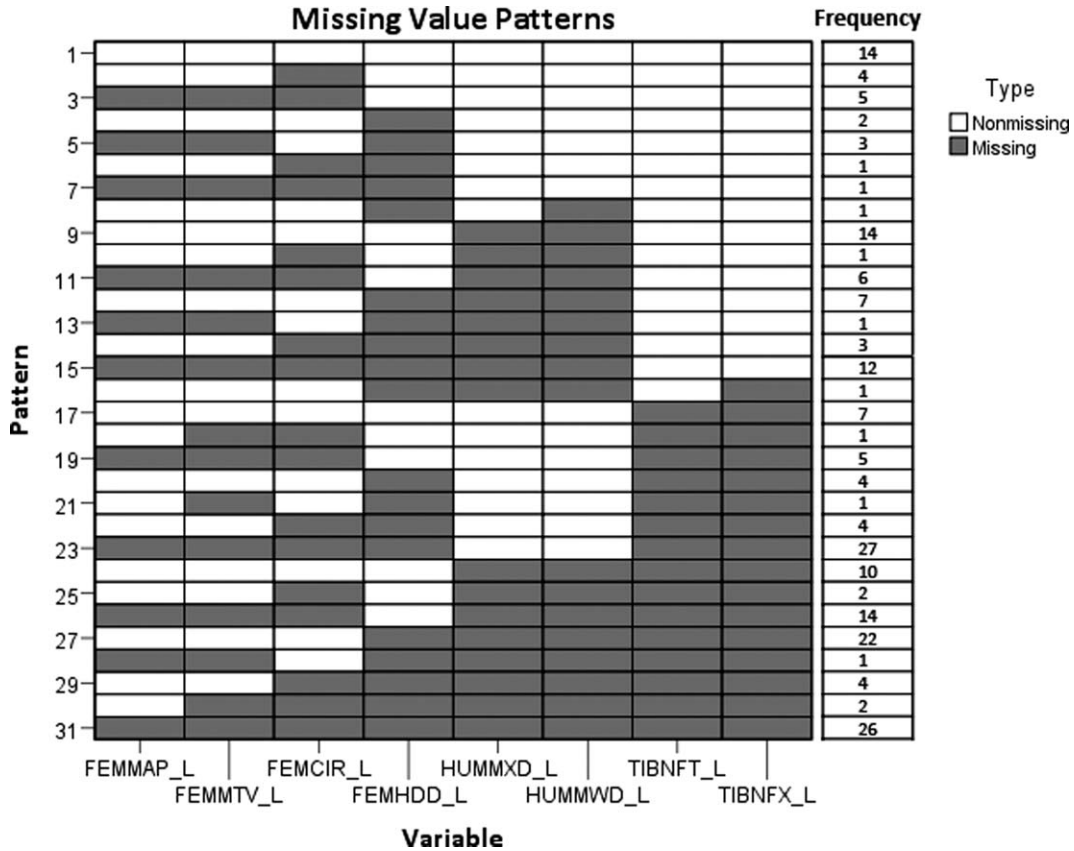


Fig. 1. Summary of missing value pattern observed on the medieval sample.

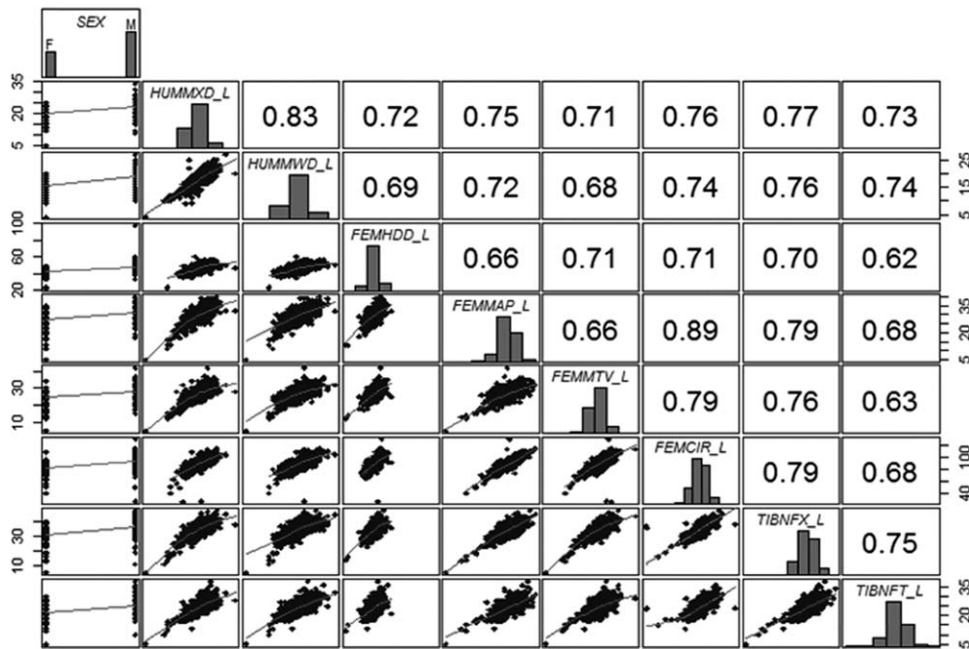


Fig. 2. Scatter-plot matrix showing scatter plots overlaid with locally weighted polynomial regression lines (lower triangle), histograms (diagonal panels), and Pearson's correlations (upper triangle) for eight osteological indicator measurements in the FDB data. The regression lines show the local trend/relationship between the pair of variables. They show that the linearity assumption of the Pearson's correlation coefficient is a reasonable for all of these relationships (none of the lines is dramatically curved).

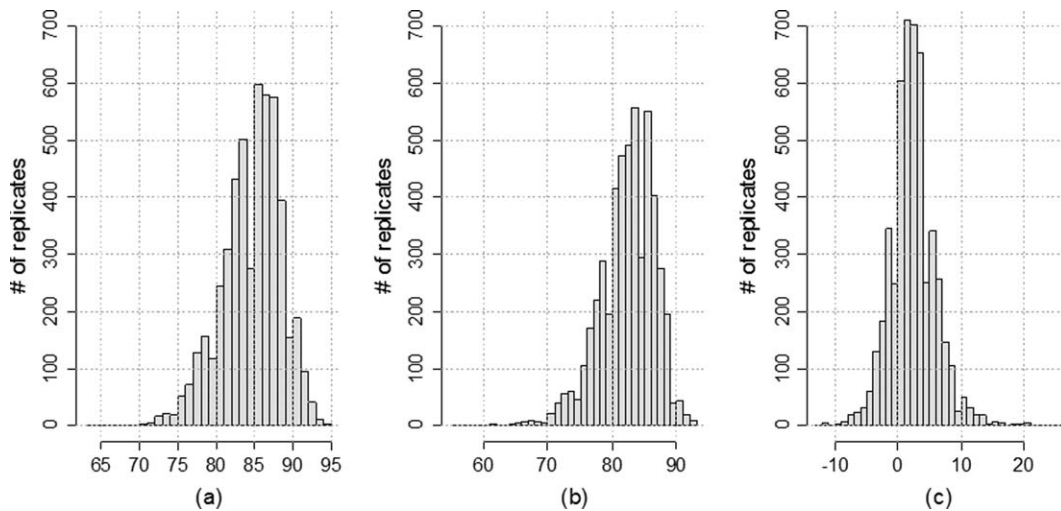
estimate sex from bioarchaeological remains (Konigsberg and Frankenberg, 2007); however, this is the first application of multivariate LPA to a known population to develop error rates based on different conditions.

The accuracy rates of the two models were significantly greater than the expected value of 50%. In comparison to the previous univariate study by Konigsberg and Frankenberg (2007), our multivariate model appears to be comparable, with a slight increase in accuracy, even when data are missing. The difference in accuracy between the full data set and the incomplete dataset, which had more than 40% of the data entries missing was very minor (~2.2%).

Finally, it should also be noted that although the measurements here were chosen based on the frequencies found from a poorly preserved skeletal sample from Medieval Asturias, Spain, the LPA method itself could be tailored to the unique context of any bioarchaeological data, with other measurements included or omitted as necessary.

HUMMXD_L	0.57	0.34	0.44	0.33	0.52	0.45	0.42
0.68	HUMMVD_L	0.28	0.39	0.31	0.47	0.42	0.46
0.43	0.38	FEMHDD_L	0.31	0.38	0.42	0.28	0.25
0.65	0.57	0.39	FEMMAP_L	0.27	0.77	0.53	0.38
0.57	0.46	0.50	0.53	FEMMTV_L	0.68	0.44	0.30
0.45	0.38	0.27	0.83	0.51	FEMCIR_L	0.61	0.47
0.63	0.59	0.48	0.73	0.66	0.48	TIBFX_L	0.49
0.61	0.58	0.31	0.57	0.43	0.34	0.64	TIBNFT_L

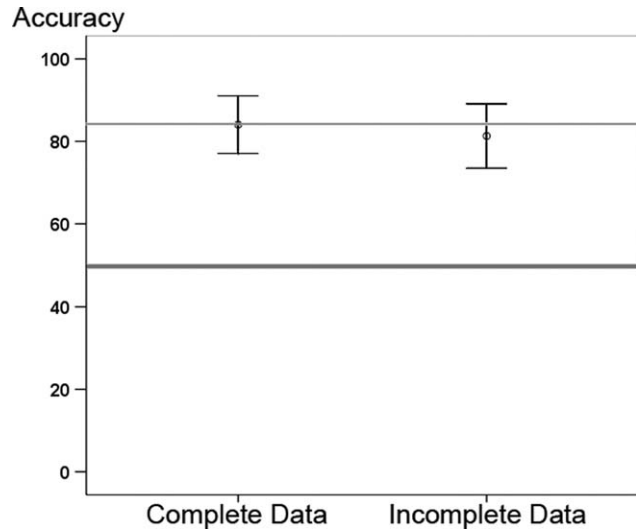
**Fig. 3.** Pearson's correlation coefficients for each pair of the eight indicator variables. Upper panel for male and lower panel for female with sex determined using LPA on FDB data.



**Fig. 4.** The histogram of (a) LPA classification accuracy rates given complete data, (b) LPA classification accuracy rates given incomplete data, and (c) the cost of missing data (i.e., difference in LPA classification accuracy rates between complete and incomplete data models). All panels are based on 5,000 replicates. Both accuracy rates and cost are expressed as percentages.

### CONCLUSIONS

LPA is a useful tool for estimating sex for populations of human skeletal remains. Further, LPA is robust enough to handle large amounts of missing data when analyzing incomplete skeletal populations with remarkably small loss of accuracy. In practical terms, the multivariate LPA method improves on Konigsberg and Frankenberg's (2007) univariate method by allowing sex determination in cases where femoral circumference data are not available because the presence of additional indicators permits LPA to still make good predictions.



**Fig. 5.** Accuracy of LPA sex determination for skeletons in the FDB under two conditions: complete data (optimal) and incomplete data (suboptimal). Accuracy rates are expressed as percentages. The dots are the median accuracy rates, and the whiskers indicate the upper and lower boundaries of the 95% credible intervals. The reference line at 50% is the expected accuracy under random guessing if one assumes a sex ratio of 1:1 between males and females.

Although most sex assessments in bioarchaeology are conducted on a case-by-case basis for each skeleton and often lack statistical support, LPA tells us the posterior probability that each individual belongs to each latent profile (i.e., sex) based on the entire sample population.

### ACKNOWLEDGMENTS

Thanks to the Museo Arqueológico de Asturias (located in Oviedo, Asturias, Spain) and staff for allowing access to the skeletal collections and as well as all their assistance during data collection. Special thanks to Ms. Belen García for her unwavering help and patience throughout data collection visits, Mr. Angel Villa Vladés for all his support and assistance, Dr. Rosario Suárez Vega for access to an additional skeletal collection, and Mr. Luis Cabo for his assistance.

### LITERATURE CITED

- Baxter MJ. 1994. Exploratory multivariate analysis in archaeology. Edinburgh: Edinburgh University Press.
- Buikstra JE, Ubelaker DH. 1994. Standards for data collection from human skeletal remains. Fayetteville: Arkansas Archeological Survey Research Series No. 44.
- Everitt, BS, Landau S, Leese M, Stahl, D. 2011. Cluster analysis. New York: John Wiley & Sons.
- Hallquist M. 2012. MplusAutomation: Automating Mplus Model Estimation and Interpretation. R package version 0.5-1.
- Jackson EC. 1968. Missing values in linear multiple discriminant analysis. *Biometrics* 24:835-844.
- Jantz RJ, Moore-Jansen PH. 2000. Database for forensic anthropology in the United States 1962-1991. ICPSR version 2000. Knoxville, TN: University of Tennessee, Dept. of Anthropology, 1998. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research.
- Konigsberg LW, Frankenberg SR. 2007. Paleodemography under duress: growth and sexual dimorphism of the minimum femoral circumference. *Am J Phys Anthropol Suppl* 47:146.
- Kramer A, Konigsberg LW. 1999. Recognizing species diversity among large-bodied hominoids: a simulation test using missing data finite mixture analysis. *J Hum Evol* 36:409-421.
- Lazarsfeld PF, Henry NW. 1968. Latent structure analysis. Boston: Houghton Mifflin.
- Lubke GH, Muthén B. 2005. Investigating population heterogeneity with factor mixture models. *Psychol Methods* 10: 21-39.
- McCutcheon AL 1987. Latent class analysis (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-064). Newbury Park, CA: Sage Publications Inc.
- Muthén LK, Muthén BO. 2010. Mplus user's guide, 6th ed. Los Angeles: Muthén & Muthén.
- Ousley SD, Jantz RL. 1997. The forensic data bank: documenting skeletal trends in the United States. In: Reichs KJ, editor. *Forensic osteology: advances in the identification of human remains*. Springfield, IL: Charles C. Thomas. p 441-458.
- Phenice TW. 1969. A newly developed visual method of sexing of the os pubis. *Am J Phys Anthropol* 30:297-302.