David G. McBride,[1] M.A.; Michael J. Dietz,[1] M.S.; Matthew T. Vennemeyer,[1] B.A.; Sara A. Meadors,[2] M.A.; Robert A. Benfer, Jr.,[1] Ph.D.; and N. Louanna Furbee,[1] Ph.D.

Bootstrap Methods for Sex Determination from the Os Coxae Using the ID3 Algorithm

REFERENCE: McBride DG, Dietz MJ, Vennemeyer MT, Meadors SA, Benfer RA, Furbee NL. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. J Forensic Sci 2001;46(3):427–431.

ABSTRACT: This study presents a method for identifying small subsets of morphological attributes of the skeletal pelvis that have consistently high reliability in assigning the sex of unknown individuals. An inductive computer algorithm (ID3) was applied to a bootstrapped training set/test set design in which the model was developed from 70% of the sample and tested on the remaining 30%. Relative accuracy of sex classification was evaluated for seven subsets of 31 morphological features of the adult os coxae. Using 115 ossa coxarum selected from the Terry Collection, a selected suite of the three most consistently diagnostic attributes averaged 93.1% correct classification of individuals by sex over ten trials. Attribute suites developed collaboratively with three well known skeletal experts averaged 87.8, 91.3, and 89.6% correct. The full set of 31 attributes averaged 90.0% accuracy. We demonstrate a small set of three criteria, selected and ordered by ID3, that is more accurate than other combinations, and suggest that ID3 is a useful approach for developing identification systems.

KEYWORDS: forensic science, forensic anthropology, sex determination, os coxae, innominate, ID3, expert systems, bootstrap, resampling

Accurate determination of sex from morphological characteristics of the skeletal pelvis is partly dependent on the order and manner in which the traits selected for analysis are applied. It is clear from studies of the decision-making processes of experts in various fields that some factors are regarded as more informative than others (1). Such preferences are based on empirical and intuitive knowledge weighted by individual experience and consistently lead expert analysts to highly accurate conclusions. However, less experienced practitioners often have limited means by which to prioritize criteria or recognize a point of diminishing return, beyond which additional data are likely to add confusion instead of clarity. Phenice (2) addressed this issue by suggesting that individual sex could be estimated with approximately 95% accuracy using a small suite of three morphological pelvic traits and proposing subjective guidelines for the evaluative weight to be placed on each trait. More recently, Rogers and Saunders (3) exhaustively tested 17 morphological features individually and in various combinations to determine optimal strategies for estimation of sex from the os coxae. Selected pairs and suites of three traits were found to estimate sex more accurately (≤95% correct) than the entire suite of 17 applied collectively or any trait used individually.

We present an alternative approach for the identification of attribute subsets. Using an expert systems algorithm (ID3) to select and order the features with the greatest explanatory power, we identify a suite of three morphological criteria of the pelvis that correctly assigned individual sex with a mean accuracy of 93.1% over ten bootstrapped trials. This approach permits evaluation that is objective, testable, and repeatable by bootstrap methods, for qualitative criteria that do not lend themselves well to analysis by linear models (see Garson (4) for comparison of ID3 with multiple linear general hypothesis (MLGH) models and neural networks, using data designed for compatibility).

An expert system is a rule-based computer program designed to simulate the decision-making process of a human expert in a particular domain or area of expertise (1). Expert systems are often deployed where the limited endurance of human experts makes automation preferable, as in industrial processes that require continuous monitoring, or when cost favors the employment of nonexpert diagnosticians, as, for example, in telephone "help" lines for basic computer technical support. They may also supplement human experts in domains where the breadth and depth of available knowledge can exceed human capacity for timely recall, such as complex or unusual medical diagnoses.

Rule-based programming is one of the most common paradigms for developing expert systems. Commonly known as "rules-of-thumb," these heuristic models can be represented as decision trees composed of series of questions whose answers determine the next questions to be asked. The trees are used to develop decision rules for expert systems. Generally, decision rules are encoded as IF . . . THEN statements, which represent the heuristic, decision-making processes of human experts (1). Rules are developed from detailed interviews, discussions, and data validation with domain experts in a process of participant observation closely resembling ethnographic fieldwork. ID3 (Iterative Dichotomizer 3) is an algorithm developed by Quinlan (5–7) as a means by which to induce decision rules directly from data sets and thereby aid the development of rule-based expert systems.

Data sets appropriate for analysis with ID3 may be interval, ordinal, categorical, or logical in form, provided that they can be organized as *examples, attributes,* and *classes.* Each example may contain many *attributes,* which are treated as independent vari-

[1] Department of Anthropology, University of Missouri-Columbia, MO.
[2] Haskell Indian Nations University, Lawrence, KS.

* Originally presented in abridged form as: McBride DG, Vennemeyer MT, Dietz MJ, Meadors SA, and Benfer RA, Jr. A test of methods for sex determination for the os coxae using the ID3 algorithm. Poster presented at the Annual Meeting of the American Association of Physical Anthropologists, 28 April–1 May 1999, Columbus, OH.

Received 7 Jan. 2000; and in revised form 15 May 2000, 25 July 2000; accepted 25 July 2000.

ables, but it may contain only one *class,* serving as the dependent variable. In this paper, each os coxae used as an *example* has 31 *attributes,* with each attribute representing a single morphological characteristic. The known sex of each example constitutes its *class.*

The *root* of the decision tree is formed by selecting the attribute that partitions the largest number of examples into a single *class,* thus minimizing entropy. This procedure is reiterated until the remaining examples are classified into subsequent *branches* that each terminate in a single class (4,8,9). To do this, ID3 uses the standard information theory definition of entropy (10), in which

$$H = -\sum_i p_i \log_2(p_i)$$

where p_i is the proportion of *class* values in category i (see Quinlan (6), Schrodt (9), and Garson (4), for detailed discussion of the ID3 algorithm).

Mathematically, ID3 minimizes entropy in each iteration of its algorithm. It will use the fewest attributes necessary to reach a complete explanation of the given classes and will frequently not need to utilize all attributes in the data set (4). In terms of output, this can be viewed as maximization of the number of examples classified on each branch of the decision tree. Therefore, ID3 tends to construct short trees and will classify correctly all examples in the data set unless two or more examples have identical values for all attributes, but belong to different classes (9). In terms of sex classification from the os coxae, such a condition exists when two cases have identical scores for all attributes, yet one is female and the other is male. These ambiguous cases are assigned adjacently, as parallel branches on the same node of the tree.

ID3's prioritization of attributes may be viewed as problematic, since examples are removed from the evaluation process once they have been classified (4,9). However, adjacent classification of ambiguous examples is a potential advantage of ID3. Since a decision tree graphically represents the relations among examples and prioritizes attributes according to their explanatory power, ambiguous examples and their common attributes are easily identifiable. Additionally, ID3 treats missing data as classes. That is, a class is assigned for all possible values of attributes selected for the decision tree, even if no examples in the data set can be assigned to it. Since it is common for skeletal material to have missing or damaged parts due to differential preservation or recovery, postmortem damage, and/or other causes that are not randomly distributed, ID3 can be used to probe for causal relations among missing data, often with directly interpretable results (4,9).

Recent developments of ID3, such as C4.5 (11) and Bayesian classification systems, such as Autoclass (12), use more complex methods of accommodating missing and ambiguous (noisy) data, and do not require prior assignment of discrete classes, but these applications may not necessarily perform better than ID3.

Materials and Methods

The Terry Collection, housed at the Smithsonian Institution in Washington, D.C., presently contains 1728 specimens. The collection contains both sexes of individuals with African, European, and mixed ancestry, whose remains were originally obtained as medical school cadavers. The collection is well documented with respect to race, sex, and age at death. It is evenly distributed by race, except for white females, which are underrepresented in several age groups. Our sample consists of adult ossa coxarum originally selected for approximately equal distribution by age, exclusive of individuals aged over 90 years. The sample of 115 comprises 35 females and 80 males, of which 26 females and 38 males were documented as African-American (Black) and one male was recorded as Asiatic. The remaining nine females and 41 males appear in collection records as white. One of us (SAM) evaluated each of three attribute suites, totaling 31 morphological characteristics of the os coxae, as elicited from three experts in skeletal analysis (Table 1).

Expert A is a bioarchaeologist who occasionally works in forensic anthropology. Expert B is recognized as both a bioarchaeologist and forensic anthropologist, and Expert C is primarily known for work in forensic anthropology. Attribute suites were elicited using ethnosemantic methods (1) to identify the morphological traits upon which each expert is most likely to rely when he or she conducts a skeletal analysis. The suites comprise pelvic traits widely recognized as sexually diagnostic, but since each suite reflects an expert's personal experience in practice it is not necessarily the same suite she or he would freely recommend to others. We have therefore withheld the experts' identities.

Age at death, race, and sex were recorded for each selected specimen in a separate log prior to scoring, to minimize the examiner's awareness of life history data. Pelvic traits were scored in anatomical groups: general pelvic morphology was scored first, followed by features of the posterior and inferior pelvis and traits associated with the os pubis. The sequence is indicated in Table 1.

We noted that ID3 will correctly classify 100% of the examples in a data set unless two or more examples have identical attribute values and different classes. But this is not a useful representation of accuracy, since the decision tree is effectively "tested" on the data from which it is built. To test the predictive value of a rule, a training set/test set research design must be used.

A decision tree must be built using a data subset (training set), and its validity tested by applying it to a test set of new examples. The accuracy of the tree is measured by the number of new examples that cannot be fitted to the existing classification. By repeating this procedure many times, each with different training/test sets, it is possible to identify the relative accuracy and usefulness of the trees, as well as the attributes with the most consistent power of classification.

For the present study, ten training sets and ten separate test sets were created by bootstrapping the primary sample of 115 individuals (9,13). In a variation of a split-sample test (13), training sets of $n = 80$ examples (~70%) were drawn randomly with replacement. Test sets were comprised of examples not selected for the corresponding training sets. Test sets therefore ranged in size from 52 to 63 examples, with a mean of 57 and mode of 59. Due to the excess of males in our primary sample, a secondary data set of 30 males and 30 females ($n = 60$) was drawn without replacement from the original sample of 115, then resampled as described above, to permit a preliminary evaluation of the sensitivity of ID3 to bias effect due to the sex distribution in the primary sample. The secondary sample was drawn without replacement to ensure that no individuals could fall into both the training and test sets upon resampling.

Each of the seven attribute suites shown in Table 2 was subjected to 10 training set/test set trials with the following objectives: 1) all 31 attributes were combined to evaluate relative influences in sex determination; 2) the suites of traits preferred by each skeletal expert were tested separately to determine the effectiveness of each; 3) a suite of three attributes (A7, B7, and C4, shown in Table 1) was pruned from those ID3 consistently placed at or near the root of decision trees and evaluated for its effectiveness relative to other attribute suites; and 4) the three attributes recommended by Phenice (2), represented here by A4, B4, and C5, were tested for comparison with our selected suite of three attributes. The secondary sam-

TABLE 1—*Morphological attribute suites recommended by Experts A, B, and C.**

Attribute Suites	Scoring Order†	Attribute Description	Scoring Criteria 1	2	3	4
Expert A						
A1	20	Sub-pubic angle	Narrow	Broad		
A2	21	Sub-pubic concavity	Present	Absent	Indeterminate	
A3	16	Pubic element	Relatively broad and rectangular	Relatively narrow and triangular	Indeterminate	
A4‡	12	Ventral arc	Present or precursor	Absent	Indeterminate	
A5	25	Pubic strut length, relative to ischial strut length	Long	Short	Intermediate or Indeterminate	
A6	6	Sciatic notch	Broad	Narrow	Indeterminate	Broad but recurved
A7§	8	Preauricular sulcus	Absent	Shallow groove	Deep, pitted	Indeterminate
A8	30	Dorsal pitting of os pubis	Absent	Shallow	Deep, pitted	Indeterminate
A9	23	Medial aspect ridge of ischio-pubic ramus	Present	Absent	Ambiguous	Indeterminate
A10	4	Acetabulum	Relatively large	Relatively small	Indeterminate	
A11	3	Os coxae general appearance	Classic female	Non-classic female or possibly male		
A12	26	Shape of symphyseal face	Rectangular	Oval	Indeterminate	
Expert B						
B1	1	Overall size of os coxae	Large	Medium	Small	Indeterminate
B2	2	Overall robusticity of os coxae	Robust	Medium	Gracile	Indeterminate
B3	17	Pubic length (superior-inferior)	Long and narrow	Medium	Short and wide	Indeterminate
B4‡	19	Subpubic ramus	Concave	Straight	Convex	
B5	13	Ventral arc	Present	Indeterminate	Absent	
B6	24	Medial aspect ridge	Present	Indeterminate	Absent	
B7§	7	Sciatic notch	Wide	Medium	Narrow	Indeterminate
B8	9	Preauricular sulcus	Grooves of pregnancy	Present	Absent	Indeterminate
B9	29	Dorsal pitting of os pubis	Present	Trace	Absent	Indeterminate
B10	27	Pubic symphysis shape	Rectangle	Intermediate	D-shaped	
B11	31	Sacral articulation	Elevated	Medium	Flat	Indeterminate
B12	5	Ilium inclination	Flared	Medium	Vertical	Indeterminate
Expert C						
C1	15	Pubic bone wide relative to overall size	Observed	Not observed		
C2	10	There _____ a precursor ventral arc	Is	Is not		
C3	11	Ventral arc	Prominent	Absent	Indeterminate	
C4§	18	Subpubic concavity	Lateral recurve	No lateral recurve	Indeterminate	
C5‡	22	Medial aspect of ischio-pubic ramus	Distinct ridge	Broad, flat surface below symphyseal surface	Indeterminate pattern	
C6	28	Dorsal pitting of pubis	Circular, scooped-out appearance of pits	Absent	Indeterminate	
C7	14	Line of ventral arc	Parallels ventral border	Does not parallel ventral border		

* Attribute descriptions taken from expert's published descriptions or forms supplied by the expert. Therefore, descriptions of the same feature (e.g., A4, ventral arc and B5, ventral arc) differ slightly by expert.
† Order in which features were evaluated in laboratory.
‡ Attributes recommended by Phenice (2).
§ Attributes preferred by ID3.

TABLE 2—*Attribute suite error rates over 10 trials.*

Attribute Suite	Mean Error %	Minimum Error %	Maximum Error %
1. All 31 attributes	10.0	5.3	13.8
2. Expert A attribute suite	12.2	5.6	18.2
3. Expert B attribute suite	8.7	0.0	17.0
4. Expert C attribute suite	10.4	6.8	17.2
5. Three preferred attributes suite	6.9	3.4	13.0
6. Phenice attributes suite	10.8	5.3	15.9
7. Three preferred attributes with equal sex ratio data sample	13.5	6.3	27.6

ple of equal sex distribution was evaluated with our preferred suite of A7, B7, and C4 (Table 1).

Decision trees and examples not correctly classified were recorded for each trial. Each example that failed classification when tested was mapped to its source individual from the Terry Collection sample to identify consistently misclassified individuals, as well as patterns of misclassification resulting from age, sex, or attribute choice.

Results

Table 2 shows summary statistics for accuracy in classification over ten trials. Error rates represent the proportion of examples in the test sets that did not fit decision trees developed from the training sets. The Expert B attribute suite was more accurate on average (91.3%) than the full set of 31 attributes (90.0%) and the suites suggested by the other two experts (87.8 and 89.6%). The Phenice suite of ventral arc, subpubic ramus, and medial aspect of the ischio-pubic ramus, was less accurate than reported by Phenice (2), but still good (89.2%), and consistent with the results of Rogers and Saunders (3). Our selected attribute suite of preauricular sulcus, sciatic notch, and subpubic concavity (as defined by A7, B7, and C4, in Table 1) was the most accurate. Results were equal to or better than other suites in nine of the ten trials, for an average of 93.1% correct classification. It was also the most consistently accurate suite, with only one trial returning less than 90% accuracy (87.0%). Other suites tended to be more variable, returning either very good results having error rates well under 10%, or poor results with error rates of 12% or greater. All suites misclassified females more frequently than males. Misclassified females were evenly distributed through all age categories. Most misclassified males were in the older age categories (50–60 and 60–70 years). We observed no systematic misclassification by sex that was attributable to racial differences. Results from the secondary sample are discussed below.

Discussion

Our results demonstrate that ID3 is an effective tool for identifying subsets of morphological features of the os coxae with a high degree of accuracy (>90%) that is comparable to other methods (2,3,14). In fact, inspecting a small number of attributes may be more accurate than using all possible traits. The selected suite of preauricular sulcus, sciatic notch, and subpubic concavity should provide good results when scored as indicated in Table 1 and applied to estimate the sex of unknown individuals.

Systematic misclassification of females in trials based on the primary sample of 115 led us to draw a secondary sample that was evenly distributed by sex, as noted above, to permit a preliminary evaluation of the effect of sampling bias when using ID3. Our selected suite of A7, B7, and C4 misclassified an average of 13.5% of all cases in the secondary sample, compared to 6.9% for the suite when applied to the primary sample of 115. The difference was largely due to three poor trials of >20% error. However, error rates by sex were not biased.

These results suggest that ID3 has a potentially useful sensitivity to qualitative and quantitative changes in sample composition. The present study was not designed directly to address questions of bias in estimation of sex from skeletal remains as discussed by Buikstra and Mielke (15), Walker (16), or Weiss (17). It is therefore not clear whether the observed effects are due to sample, interobserver, or attribute bias. Success in scoring by SAM, by attending to the differences in the descriptive terminology used for scoring attributes (see, for example, A8, B9, and C6, in Table 1), may have a significant impact on bias in sex estimation from morphological characteristics.

We have demonstrated that the ID3 algorithm is useful in identifying potentially informative subsets of skeletal morphological attributes. Its advantage over using all available attributes is that it suggests relatively few attributes and presents an optimal order and application of the attributes. ID3 is capable of producing high accuracy in bootstrapped training set/test set trials, even with as few as three attributes. The ability to identify subsets of useful attributes can be particularly helpful in archaeological or forensic applications, where for example, missing or damaged skeletal elements may frustrate analysis by linear models or metric calculations, and selection of attribute subsets appropriate to the available material is desirable. Although the ID3 algorithm is not necessarily a substitute for human expert analysis, it can be applied effectively to any question in skeletal analysis where data can be organized into discrete classes that are composed of attributes.

Acknowledgments

We wish to thank the Smithsonian Institution, Washington, D.C. for access to the Terry Collection. The Weldon Springs Research Fund of the University of Missouri-Columbia provided support for the study with the expert consultants. We also wish to thank our anonymous reviewers for their time and valuable comments. We are indebted to the three unnamed experts in skeletal analysis.

References

1. Benfer RA, Jr., Brent, E E, Jr., and Furbee, L. Expert systems. Newbury Park: Sage, 1991.
2. Phenice TW. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 1969;30:297–302.
3. Rogers TL, Saunders SR. Accuracy of sex determination using morphological traits of the human pelvis. J Forensic Sci 1994;39:1047–56.
4. Garson GD. A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. Soc Sci Computer Rev 1991;9:399–434.
5. Quinlan JR. Induction over large data bases. Stanford: Computer Science Department, Stanford University, 1979.
6. Quinlan JR. Learning efficient classification procedures and their application to chess end games. In: Michalski RS, Carbonell JG, Mitchell RM, editors. Machine learning: an artificial intelligence approach. Palo Alto, CA: Tioga, 1983:463–82.
7. Quinlan JR. Induction of decision trees. Machine Learning 1986;1: 81–106.
8. Hart A. Knowledge acquisition for expert systems. New York: McGraw-Hill, 1986.

9. Schrodt PA. Predicting interstate conflict outcomes using a bootstrapped ID3 algorithm. Political Analysis 1990;2:31–56.
10. Shannon CE, Weaver W. The mathematical theory of communication. Urbana: University of Illinois Press, 1964.
11. Quinlan JR. C4.5: programs for machine learning. San Mateo: Morgan Kaufmann, 1993.
12. Stutz J, Cheeseman P. AutoClass—A Bayesian approach to classification. In: Skilling J, Sibisi S, editors. Maximum entropy and Bayesian methods. Dordrecht: Kluwer, 1996:117–26.
13. Diaconis P, Efron B. Computer-intensive methods in statistics. Sci Am 1983;248:116–30.
14. Sutherland LD, Suchey JM. Use of the ventral arc in pubic sex determination. J Forensic Sci 1991;36:501–11.
15. Buikstra JE, Mielke JH. Demography, diet, and health. In: Gilbert RI, Mielke JH, editors. The analysis of prehistoric diets. Orlando: Academic Press, 1985.
16. Walker PL. Problems of preservation and sexism in sexing: some lessons from historical collections for palaeodemographers. In: Saunders SR, editor. Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars' Press, 1995.
17. Weiss KM. On the systematic bias in skeletal sexing. Am J Phys Anthropol 1972;37:239–50.

Additional information and reprint requests:
David G. McBride
Department of Anthropology
107 Swallow Hall
University of Missouri-Columbia
Columbia, MO 65211
(573) 882-9403
e-mail: ldmcbride@socket.net