9-13-2010

Ancestry Estimation in a Web-based, Searchable Database of Orthodontic Case Files for Patient Care, Education, and Research

Philip J. Kroth

Heather J.H. Edgar

Edward F. Harris

Summers Kalishman

Shamsi Daneshvari

Recommended Citation

Kroth PJ, Edgar HJ, Harris EF, Kalishman S, Daneshvari S. Ancestry Estimation in a Web-based, Searchable Database of Orthodontic Case Files for Patient Care, Education, and Research. Abstract and poster presented at MedInfo 2010, the 13th World Congress on Medical and Health Informatics, September 13, 2010; Cape Town, South Africa.

Ancestry estimation in a web-based, searchable database of orthodontic case file for patient care, education, and research

Heather Edgar[1], Philip Kroth[1], Edward Harris[2], Summers Kalishman[1], Shamsi Daneshvari[1]
[1]University of New Mexico, [2]University of Tennessee

http://hsc.unm.edu/programs/ocfs

ABSTRACT

In 2005, the Maxwell Museum of Anthropology accepted a donation of orthodontic patient records from an orthodontist who has been practicing in the Albuquerque area since the early 1970's. This collection represents a diversity of patients not often encountered in orthodontic training in the United States. A virtual, de-identified, web-based version of a subset of the collection is now being developed. Users can search for cases with particular characteristics of interest (e.g., patient ancestry, extraction patterns, diagnoses, and cephalometric parameters), then review sequential intra-oral and X-ray images to observe treatment outcomes. An innovative feature of the database is that it records multiple ancestry estimations, made at multiple points in time by multiple raters, along with a list of ancestry indicators on which the estimations are made (e.g., skin color, hair form and color, facial shape, name, and locality). This paper describes how the database can be used to overcome the limited diversity in the patient populations available to most orthodontics trainees. When this project concludes, the database will contain approximately 400,000 digitized images from 5,650 individual cases.

INTRODUCTION

The U.S. has the most diverse population of any nation. However, orthodontic education often fails to reflect this diversity in the patient populations available to students during their training. Given that groups are not equally distributed across the landscape, within any given program, orthodontic students' training often cannot prepare them to accommodate the diversity they are likely to encounter in their future practices. In addition, "normality" of facial measurements in orthodontics is most often determined from comparison with cephalometric norms established from sentinel homogeneous populations.[1] In reality, many mean facial measurements and growth patterns correlate with social descriptors of race and ethnicity, and may be significantly different in populations other than those represented in the sentinel groups.[2-4] In order to remedy these shortcomings, we developed a freely available, web-based, searchable database of orthodontic cases representing a diverse range of American populations. The goal of this work is to improve orthodontic training experiences by increasing the variety of patients to which students are exposed, preparing them better for practice in the 21st century.

In 2005, the University of New Mexico's Maxwell Museum of Anthropology acquired the James Economides Orthodontic Collection. The collection was compiled from 1972 through 1999, and consists of dental casts, cephalometric radiographs, photos, and treatment records for approximately 5,650 orthodontic patients including records of approximately 600 sibling pairs and several multi-generational families.

Approximately 400,000 photos/images and 20,000 x-ray films are included in the collection. These images represent the facial, skeletal, and dental variation and treatment of the contemporary population in Albuquerque over the last 35 years, including people from a variety of ancestral backgrounds. The diversity and quality of this collection are its unique strengths, representing African, Asian, European, Hispanic, and Native American populations (Table 1). However, patient records included neither the patient's estimation of their own ancestry (self identification) nor Dr. Economides' estimation (community or physician identification).

Most orthodontists encounter patients with different orthodontic problem-ancestry combinations they have not encountered in training,[5] the database, for this web site was designed to allow users to search for specific orthodontic problems and then limit the results to designated ethnic and racial subgroups. This allows users to view variations in treatment outcomes across a specifiable range of facial forms.

Since neither self nor community identification of patients' race and ethnicity pre-existed in patient records, we estimated ancestry from the materials present in the records. In order to attempt to address the traditional weaknesses of provider coded ancestry determination, the system was uniquely designed to accommodate multiple ancestry estimations by numerous raters and times. There are a minimum of two raters' ancestry determinations per individual case stored in the database. This system is also innovative in that raters are required to record the indicators they used when estimating each patient's ancestry. This capability provides a much more rigorous and flexible method of identifying cases in the database using contemporary racial and ethnic indicators for any given time period. In order to develop this system, a set of terms describing possible ancestry had to be delimited, as well as a list of possible indicators.

MATERIALS AND METHODS

Ancestry Coding

There are numerous terminology systems available for describing the variation present in U.S. populations. Most of these terminologies list groups under the overall rubrics of "race" and "ethnicity." These categories are levels of socially ascribed folk taxonomies that often incorporate biological characteristics, such as skin color, as features used for group assignment.[6] Actual ancestry only overlaps with race and ethnicity to the extent that the biological characteristics used for group assignment are inherited.[7] Specificity of possible assignment varies among terminological systems. For example, the 2000 U.S. Census listed five overall racial categories and two ethnicities[8]; the CDC currently lists nine overall racial categories with hundreds of subcategories that subsume ethnic coding.[9, 10] Billion dollar decisions regarding the allocation of public funds for various healthcare, education, and other public programs are based on these classifications even though there are significant ambiguities in identifying and classifying populations according to race and ethnicity.

In order to determine what racial and ethnic categories should be included as variables in this database, while recognizing the ambiguity of racial and ethnic classification, we compared as many existing terminological schemes as possible to a set of three criteria: 1) familiarity of coding to raters and projected database users; 2) use of coding in medical research; and 3) applicability to the specific Economides Collection.

	%*	N*
Age at first treatment		
<18	70	3955
18+	30	1695
Sex		
Male	30	1695
Female	70	3955
Ancestry		
African American	2	113
Asian American	1	57
European American	76	4294
Hispanic American	17	960
Native American	4	226
Total:		5650

*Based on a sample of 100 records chosen at random, but stratified for patients from all decades of practice.

Table 1. Patient demographics.

Essential Indicators to be able to Determine Ancestry

1. Availability to raters in patient records, which include full facial photos, lateral view x-rays, and patient demographics, including age, sex, and address
2. Informative about patients' biology and/or ancestry
3. Visible to the general population (not requiring medical equipment to see). Prior to the finalization of the database format, discussions were held with potential raters and database users to limit the code set for ancestry indicators.

Statistical Methods

Rater agreement was determined for two different sets of data (n=1,900 sets of two observer observations), race/ethnicity categories and race/ethnicity indicator categories. Results presented here include agreement frequencies and Kappa coefficients calculated for a subset of race/ethnicity category data (n=1,076).

RESULTS

Ancestry Coding

A slightly modified version of the 2000 U.S. Census categories is being used to code ancestry in this database. The Census terminology was chosen because it is familiar to most Americans and is very commonly used in medical research. The modification made reflects the specific population of Albuquerque and Dr. Economides' patient sample. New Mexico is 42.1% Hispanic,[11] a code listed as an ethnicity in the Census' scheme. However, few people in New Mexico differentiate "Hispanic" as an ethnicity, and therefore are included as a different category with other codes that are listed in the Census as races. Raters have the codes "African American," "European American," "Asian American," "Hispanic American," "Native American," and "Native Hawaiian/Pacific Islander" available. Further, if they choose "Native American" a drop-down menu appears listing the 22 Native American tribes present in New Mexico. Often, tribal affiliation can be determined from a patient's address (ex. "Zuni Pueblo"). This data is coded in the database using the tribal affiliation codes as defined in the National Register's list of federally recognized Tribes.[12]

Indicators of Ancestry

1. Patient name
2. Patient address
3. Skin color
4. Facial shape
5. Fair form and color

Statistical Analyses

Figure 1 presents the frequencies of outcomes for two raters observations of race/ethnicity as frequencies for those cases when each observer selected only one race/ethnicity case. Hawaiians were removed from this analysis as there were only three total observations of this category. Overall agreement between two observers is 95%. Table 2 presents two sets of results, rater's choice of race/ethnicity indicators in cases where each observer chose only one category and where the two observers agreed on race/ethnicity, and the same for cases where observers disagreed. Table 3 presents Kappa statistics for inter-observer reliability for the test sample.

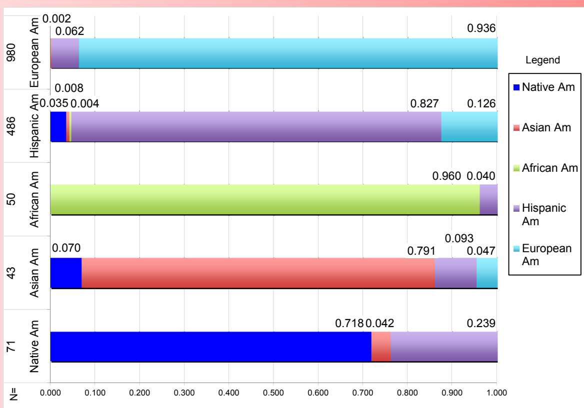

Figure 1. Frequency of race/ethnicity agreements and disagreements when two observers each chose only one category.

Indicators in cases of agreement

African Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.303			
Name	0.000	0.103	0.000		
Skin	0.000	0.028	0.110	0.324	
Face	0.000	0.014	0.103	0.014	0.324

Native Am

	Address	Hair	Name	Skin	Face
Address	0.009				
Hair	0.018	0.100			
Name	0.018	0.164	0.015		
Skin	0.158	0.038	0.067	0.129	
Face	0.023	0.026	0.070	0.032	0.132

Asian Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.186			
Name	0.000	0.050	0.179		
Skin	0.000	0.021	0.050	0.193	
Face	0.000	0.050	0.014	0.029	0.229

European Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.144			
Name	0.000	0.112	0.036		
Skin	0.000	0.055	0.113	0.200	
Face	0.000	0.032	0.107	0.028	0.174

Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.117			
Name	0.000	0.113	0.115		
Skin	0.000	0.068	0.075	0.178	
Face	0.000	0.061	0.072	0.020	0.180

Indicators in cases of disagreement

African Am - Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.222			
Name	0.000	0.111	0.000		
Skin	0.000	0.000	0.111	0.222	
Face	0.000	0.000	0.111	0.000	0.222

Native Am - Asian Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.125			
Name	0.000	0.125	0.000		
Skin	0.000	0.063	0.125	0.188	
Face	0.000	0.063	0.125	0.000	0.188

Asian Am- European Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.222			
Name	0.000	0.111	0.000		
Skin	0.000	0.000	0.111	0.222	
Face	0.000	0.000	0.111	0.000	0.222

Asian Am - Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.105			
Name	0.000	0.105	0.000		
Skin	0.000	0.053	0.105	0.211	
Face	0.000	0.105	0.105	0.053	0.158

European Am - Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.118			
Name	0.000	0.136	0.015		
Skin	0.000	0.054	0.142	0.175	
Face	0.000	0.039	0.139	0.021	0.160

Native American Am - Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.145			
Name	0.000	0.105	0.053		
Skin	0.000	0.053	0.026	0.197	
Face	0.000	0.053	0.079	0.092	0.197

African Am - Hispanic Am

	Address	Hair	Name	Skin	Face
Address	0.000				
Hair	0.000	0.125			
Name	0.000	0.063	0.000		
Skin	0.000	0.000	0.063	0.125	
Face	0.000	0.000	0.063	0.000	0.125

(highlighted) indicates greater than 20% of responses

Table 2. Frequencies of indicator choices for cases of agreement and disagreement.

DISCUSSION

	Observed Concordance	Expected Concordance	Kappa Coeffcient
African Am	25	29	0.83
Asian Am	20	29	0.627
European Am	670	647	0.928
Hispanic Am	266	318	0.82
Native Am	39	51	0.707

Table 3. Kappa statistics for agreement in a test sample of 1,076 paired observations.

Racial and ethnic classification is to some extent a moving target, in that how a person is classified today can be different from how they may have been classified 50 years ago, or could be classified 50 years from today.[13] In addition, it may be possible for the same rater to change classification decisions over time, as their exposure to persons of different groups also changes. Raters are affected by their own racial and ethnic backgrounds, upbringing, and biases, as is evidenced by the imperfect agreement among multiple classifiers as well as individual's self identification.[14]

Recognizing the problems and ambiguities with classifying patients by race and ethnicity, we realized that attempting to choose the perfect racial and ethnic coding system cannot solve this problem. Rather, we chose to design the database to capture multiple racial and ethnic estimations by numerous raters. At least two raters make independent estimations of ancestry for each patient in the database. Our experience with using two raters during the data entry and verification process showed a reasonable agreement at a given point in time (Table 3.) Concordance between the raters was significant for all ancestries except for Hawaiian. The database also records raters' self-identified ancestry, as well as the date each classification was performed. This design will allow future users of the database to create queries to accommodate the ambiguities and limitations of racial and ethnic classification of the database as these categories change over time. This also provides anthropologists with a very powerful tool for those interested in studying racial perceptions in contemporary human societies.

There are two potential applications for this effort. First, users of the database may opt to limit their studies to patients for whom all raters are in agreement. Second, independent estimates of ancestry, coupled with rater's recorded use of indicators of that ancestry, provide a new avenue of research into the nature of the taxonomy of race and ethnicity in the United States. Potential research questions to be addressed with this data include whether there is higher agreement for some estimated ancestries than others, and whether raters who agree on patients' ancestries are using the same indicators to make their estimate.

REFERENCES

[1] Riolo M, Moyers R, McNamara J, Jr., Hunter W. *An atlas of craniofacial growth: cephalometric standards from the University School Growth Study, the University of Michigan*. Ann Arbor: Center for Human Growth and Development, University of Michigan 1974.
[2] Howells W. *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference Among Recent Human Populations*. Cambridge: Harvard University 1973.
[3] Richardson ER. Evaluating Ricketts' Cephalometric Analysis as Diagnostic Aid in Black Females. International Association for Dental Research 37th Annual Meeting and Exhibition. Dallas, TX; 5 Apr 2008. http://iadr.confex.com/iadr/2008Dallas/techprogram/abstract_101538.htm.
[4] Richardson ER. Can Ricketts' Cephalometric Analysis be Applied to Black Males. International Association for Dental Research 86th General Session and Exhibition. Toronto, Canada; 4 Jul 2008. http://iadr.confex.com/iadr/2008Toronto/techprogram/abstract_108275.htm.
[5] Owens EG, Goodacre CJ, Loh PL, Hanke G, Okamura M, Jo KH, Munoz CA, Naylor WP. A multicenter interracial study of facial appearance. Part 1: A comparison of extraoral parameters. *Int J Prosthodont*. 2002 May-Jun;15(3):273-82.
[6] Berlin, Breedlove B, Breedlove DE, Raven PH. Folk taxonomies and biological classification. *Science*. 1966;154:273-5.
[7] Sauer N. Forensic Anthropology and the Concept of Race - If Races Dont Exist, Why Are Forensic Anthropologists So Good at Identifying Them. *Soc Sci Med*. 1991;34(2):107-11.
[8] Centers for Disease Control and Prevention. Reporting race and ethnicity data--National Electronic Telecommunications System for Surveillance, 1994-1997. *MMWR Morb Mortal Wkly Rep*. 1999 Apr 23;48(15):305-12.
[9] Centers for Disease Control and Prevention. Race and ethnicity code set 1.0. 2000 [cited 7 Oct 2009]; Available from: http://www.cdc.gov/nedss/datamodels/intro_race_ethnicity.pdf.
[10] Centers for Disease Control and Prevention. Race and ethnicity code set 1.0: Table 1 -- Concepts and Codes 2000 [cited 7 Oct 2009]; Available from: http://www.cdc.gov/nedss/DataModels/Race_Ethnicity_CodeSet.pdf.
[11] Bureau UC. 2000 [cited 6 Oct, 2009]; Available from: http://www.census.gov/prod/2001pubs/c2kbr01-3.pdf.
[12] Federal Register. Indian Entities Recognized and Eligible to Receive Services from the United States Bureau of Indian Affairs. 2008 [cited 7 Oct 2009]; Available from: http://www.loc.gov/catdir/cpso/biaind.pdf.
[13] Edgar HH. Biohistorical Approaches to "Race" in the United States: Biological Distances Among African Americans, European Americans, and Their Ancestors. *Am J Phys Anthropol*. 2009;39(1):58-67.
[14] Kressin NR, Chang BH, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. *Am J Public Health*. 2003 Oct;93(10):1734-9. PMC1448042./

This project was supported by Award Number 5G08 LM009381 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.