Document Title:	**Investigation of Subadult Dental Age-at-death Estimation using Transition Analysis and Machine Learning Methods**
Author(s):	**Nicholas P. Herrmann, Ph.D., Joseph T. Hefner, Ph.D., Jennifer E. Spence, Ph.D., Kelly Kamnikar, M.A.**
Document Number:	**306558**
Date Received:	**May 2023**
Award Number:	**2018-DU-BX-0182**

Investigation of Subadult Dental Age-at-death Estimation using Transition Analysis and Machine Learning Methods

Nicholas P. Herrmann, Ph.D.

Professor, Department of Anthropology, Texas State University
nph16@txstate.edu

Joseph T. Hefner, Ph.D.

Assistant Professor, Department of Anthropology, Michigan State University
hefnerj1@msu.edu

With contributions by:

Jennifer E. Spence, Ph.D.

Visiting Assistant Professor, Tulane University

Kelly Kamnikar, M.A.

Graduate Researcher, Department of Anthropology, Michigan State University

December 31, 2021

DUNS:

EIN:

Texas State University San Marcos, Texas 78666

January 1, 2019 – December 31, 2021

Final Report

Prepared by Jennifer Spence, Kelly Kamnikar, Joseph Hefner & Nicholas Herrmann

Table of Contents

Introduction

In January 2018, the National Institute of Justice (NIJ) funded a project entitled Investigation of Subadult Dental Age-at-Death Estimation Using Transition Analysis and Machine Learning Methods (Award No. 2018-DU-BX-0182). The major goals of this project were to develop a large reference sample of dental developmental data from contemporary populations, develop age estimation models within a transition analysis framework (TA), address common issues encountered in forensic casework that affect dental age estimates, like missing teeth, and develop and distribute an open-source analytical tool for practitioner use in casework and research. Overall, this project addresses the gap in best practice when estimating age in subadult skeletal remains by providing an accurate method with a contemporary sample, as well as the most appropriate statistical frameworks, specifically transition analysis (TA) and machine learning methods.

Project Purpose

Background

The purpose of this research is to address the gap in best practice when estimating age in subadult skeletal remains by providing an accurate method with a contemporary sample, as well as the most appropriate statistical frameworks, specifically transition analysis (TA) and machine learning methods.

Current methodology often relies on outdated samples and unknown sample sizes per age cohort, and visual comparison of dentition to pictures or drawings. Additionally, some methods require a complete suite of dentition to generate age estimates. Error rates are generalized into months and years and increase as age increases. Atlases and drawings are more general in nature, based on the overall development and eruption of dentition without capturing tooth-specific variation.

The goal of this research is to provide forensic anthropologists with an accurate age estimation method and tool based on a large, demographically diverse, modern subadult sample that accurately captures variation in the dental developmental process. Using dental development within a TA framework allows for statistical rigor to be associated with the method. Statistics included within the TA output serve as point age estimates along with associated estimated age-at-death ranges, which are unique to the individual. TA moves away from the stage-based age estimation methods, providing statistically-based, individualized results for each case.

Project Design

This project refines the age-at-transition values of development for three deciduous and ten permanent teeth in subadults creating the most appropriate numerical parameters from which to most accurately estimate age. This research improves best practice for estimating subadult age-at-

1

death in forensic casework via the application of TA to a complete or an incomplete suite of dentition.

The main objectives are: (1) to collect data for multiple populations to refine age at transition estimates by tooth, specific to various geographic groups in forensically-significant populations from which to derive reference data and identify patterns in dental development; (2) to **develop** age estimation models based on the data collected that conform to previous models of transition analysis; (3) to **investigate** exploratory analyses focused on the type and number of teeth need to accurately estimate age, which simulate common forensic scenarios where the teeth are missing due to taphonomic processes, and evaluate differentiation in age-at-transition by sex, and through various statistical approaches; and (4) to **develop and distribute** a freely available open-source computer application that can be easily accessed and used by practitioners when estimating age in unknown subadult remains.

First, this project incorporates recent reference material from contemporary populations to understand and refine the pattern of growth and development in modern subadult dentition. Second, this reference data informs the TA model used to generate age estimates for unknown individuals. The final estimation includes a maximum likelihood age estimate based on the tooth stage of each tooth present and associated confidence bands at the 90% and 95% confidence interval. Third, we calculate accuracy rates of specific teeth or different combinations of teeth in the age estimation model. This provides practitioners with specific guidelines on the number and type of teeth to use in age estimation when dentition is incomplete. It is anticipated that we will explore if certain teeth more heavily influence age estimations. Fourth, we examine other statistical approaches to aging, such as machine learning, to assess the efficacy of these systems relative to TA. Lastly, this project produces an easy to use, freely available, web-based application for practitioners to use when estimating age in subadults via dentition.

The computer program, Transition Analysis Dental Age (TADA), is in the last phase of this research is undergoing final adjustments before the release of the beta version. TADA is a Shiny-based web application using the Moorrees, Fanning, and Hunt (MFH) methodology scores within the transition analysis method. This program allows practitioners to enter dental developmental data from subadult cases and estimate age using the parameters previously calculated following transition analysis. The goal is to allow for the selection of ancestral groups (or all groups in the case of an unknown) and sex to ascertain a more accurate age estimation. TADA is in English and Spanish allowing for the wide use of the method and distribution of deliverables. TADA is described more in detail further in the report.

Samples

The first phase of research involved extensive data collection from radiographic data from individuals coming from different groups within the United States, Europe, and Africa. Reference groups represent samples from various ethnic and socioeconomic backgrounds, including African

2

American, European American, American Hispanic, British, and South African. Data collection focused on obtaining, at a minimum, several hundred individuals per age cohort from birth to 18 years old, as dental development reaches the end of its applicability.

This project incorporates recent reference material from contemporary populations with a minimum sample size of 150 individuals per age cohort (i.e., one-year increments) into an established database used in Kamnikar and colleagues' (2018) research (Table 1). The purpose of adding more reference data is to understand and refine the pattern of growth and development in modern subadult dentition.

Data collection proceeded through the establishment of new partnerships with institutions that have large subadult radiographic dental collections. These collaborations are listed below by institution. Table 1 shows the data source and number of images in each source.

| Sample | Collection | Sex | | | Reference |
		Males	Females	Unknown	
Barts & The London School of Medicine	Modern anonymized cases	527	520	0	Liversidge 2009
South Africa	Modern anonymized cases	492	560	221	Phillips & van Wyk Kotze 2009
Universite de Bordeaux	Modern anonymized cases	644	779	1	Heuzé & Cardoso 2008
Universidad Complutense de Madrid	Modern anonymized cases	772	999	14	Feijóo et al. 2012
University of San Antonio, Health Science Center	Modern anonymized cases	701	907	0	
Orthodontic Case File System, Maxwell Museum of Anthropology, University of New Mexico	Modern anonymized cases	1643	2130	0	Edgar 2013
New Mexico Decedent Image Database*	Modern anonymized forensic cases	1257	595	4	Edgar et al. 2020

* Provided access, downloaded and being processed for research

Barts and London School of Medicine

The sample from London, England comes from the Institute of Dentistry and the Barts and London School of Medicine and Dentistry. Pantographic radiographs were taken from patients ranging in age from 2.07 years old to 22.99 years old. Patients come from White and Bangladeshi

3

ethnic groups. Dr. Helen Liversidge provided previously coded dental data (Liversidge 2009).

South Africa

This collection was compiled from different sources of radiographic material. The first source comes from the archival records of the school of dentistry at the University of Western Cape. Images comes from mixed- ancestry individuals, including slaves, indigenous Khoisan, and European descendant children, and Xhosa children. The second source comes from two orthodontic offices in the state of Durban. The individuals in the sample range in age from 3 to 17 years old.

Universitie de Bordeaux

The collection from the Universitie de Bordeaux in Bordeaux, France comes from previously collected radiographs from Dr. Yann Heuzé. Dental development data were collected from radiographs of children ranging in age from 1-16 from orthodontist offices and private hospitals in southern France. Individuals in this sample come from different socioeconomic status; however, SES is not linked to a particular individual. The data excludes any pathological or anomalies in tooth numbers (Heuzé & Braga 2008).

Universidad Complutense de Madrid (UCMAD)

The UCMAD sample comes from the Dental School at the Universidad Complutense de Madrid in Spain. This sample comprises orthopantograms of males and females aged 2-16 years old (Feijóo et al. 2012).

University of Texas Health Science Center

The cases are from University of Texas Health Science Center in San Antonio and provided by Dr. Peter Gakunga. These represent orthopantomogram radiographs.

Orthodontics Case File System, Maxwell Museum, University of New Mexico

Data was collected from the Orthodontics Case File System, housed at Maxwell Museum of Anthropology at the University of New Mexico (UNM). This collection includes records from individuals who received orthodontic treatment from a private local orthodontist, Dr. James Economides, between 1972 and 1999. The collection is broadly representative of the population in New Mexico, primarily Albuquerque, during the years Dr. Economides practiced. This collection contains longitudinal data in the form of multiple radiographs for the same individual over a period of years. Some data (low-resolution radiographic images and limited demographic information) are available via the web-based system.[1] Information regarding an individual's age in days during treatment and address was obtained from the records themselves. Socioeconomic status for some individuals was calculated using census reports produced by the Federal Financial Institutions Examination Council.[2]

---

[1] https://hsc.unm.edu/programs/ocfs/
[2] https://www.ffiec.gov/census/default.aspx

New Mexico Decedent Image Database (NMDID)

The NMDID dataset includes whole body CT scans of New Mexicans who died between 2010 and 2017. This collection includes metadata associated with each individual (e.g., sex, ancestry, age at death, cause of death). These images are large and require 3D reconstructions to isolate the dentition as the region of interest for analysis; currently analysts are preparing the files from NMDID for dental development data collection. The NMDID source is unique because it contains potential dental data from very young individuals, usually not gleaned from orthodontic data sources.

Dental Development Data Collection Protocol

Prior to dental development data collection, project members participated in three organizational meetings to develop a standard operating procedure, address developing concerns or issues in terms of scoring radiographic images and maximize reliability across project members (analysts in particular). The first meeting (September 2019) included a tutorial presented by the project Forensic Odontologist, Dr. Fancher, and a discussion of methodology. The subsequent two meetings (October and November 2019) served as calibration sessions during which assessments of dental development were compared across analysts. Though initial reliability across analysts was quite high, these meetings served to ensure that data collection remains as consistent as possible throughout the project. A standard operating procedure (SOP) was created for all personnel working on the project which 1) described procedures for selecting, opening, enhancing, scoring, and closing images and 2) provided contact information in case scores needed to be corrected and/or questions arose. Please see Appendix A for the SOP.

Dental development data were collected from each radiograph present in all samples. Teeth were scored following the protocol developed by Moorrees et al. (1963a,b), the 17-stage scoring-system for permanent and deciduous teeth (Table 2).

Table 2. MFH scoring system. Adopted from Moorrees, Fanning, and Hunt (1963a,b).

Development Phase	Tooth Code
1	initial cusp formation
2	coalescence of cusps
3	cusp outline complete
4	crown ½ complete
5	crown ¾ complete
6	crown complete
7	initial root formation
8	initial cleft formation
9	root length ¼
10	root length ½
11	root length ¾
12	root length complete
13	apex ½ closed
14	apex closed
15	resorption ¼
16	resorption ½
17	resorption ¾

Dental development data collection began in October 2019. As of December 2021, analysts collected dental development data from four sources (the University of Bordeaux, the University of Madrid, the University of Texas Health Science Center, and the Orthodontics Case File System. Files from a fifth source, the New Mexico Decedent Image Database (NMDID), are being processed for data collection.

Table 3 shows the data source and number of images. Seventy-four percent of the available images have been scored and entered into the master database. An additional 1,857 images are anticipated from NMDID.

Table 3. Data source and number of images			
Data Source	Images (Total)	Images (Data Collected)	Percentage Completed
University of Bordeaux	1,424	1,424	100
University of Madrid	1,785	1,785	100
University of Texas Health Science Center	1,608	1,608	100
Economides Orthodontics Case File System	6,831	3,773	55
Data Collected as of December 2021	**11,648**	**8,590**	**74**
New Mexico Decedent Image Database	1,857	–	–
Anticipated Total Data Collected	**13,505**	–	–

A master Access database tool to store, organize, and reference compiled dental development and demographic data has been created and is in use. It currently contains dental and demographic data for 8,590 images from 6,671 individuals. Data are being added on a regular

6

basis. It is estimated that data from around 13,500 images (Table 3) will ultimately be included in the master database.

Methods

Data Cleaning and Preparation

All data were combined into one Excel sheet. As the original Moorrees et al. (1963) method used mandibular teeth, we focused our analysis on developmental scores in mandibular quadrants. During data collection, analysts scored the clarity of the dental radiographs as 'poor', 'fair' and 'good' which corresponded to the ability to see and interpret dental developmental scores. In the recalculation of the parameters behind the TA method, we included images that were deemed to be 'fair' or 'good' quality for the mandibular quadrants. When 'poor' was assigned with either 'fair' or 'good' in the mandibular quadrants, only dental developmental scores from the 'fair' or 'good' quadrant were included in the final dataset. See SOP in Appendix A.

All scores were arranged following the data inputs for the code provided in Shackelford et al. (2012) that was used by Kamnikar et al. (2018) from which this project stemmed (Figure 1). The focus of the table recalculations are mandibular teeth, as that was the protocol in the original Moorrees et al. (1963) method, and the central and lateral maxillary incisors.

ID	dc	dm1	dm2	UI1	UI2	LI1	LI2	C	P3	P4	M1	M2	M3	Neander	Obs
Test	10	10	10	2	2	2	2	2	2	2	2	2	2	FALSE	1
EX1	NA	NA	14	6	4	NA	7	NA	NA	7	8	3	NA	FALSE	1

Figure 1: Example of data format for use in R.

Age Estimation Transitions

The primary goal of developing the reference database is to update Shackelford and colleagues' (2012) transition analysis modeling using a representative sample from modern populations to aid in forensic identification of subadult human remains. This model examines mean age-at-attainment values to calculate a maximum likelihood estimation for age-at-death, along with ranges associated with confidence intervals at the 90% and 95% level. To expand, we added data from the samples listed in Table 1, which include longitudinal data from the Orthodontics Case File System. Longitudinal data provides the timing of the individual variation in transition from one stage to the next, creating realistic patterns of development via the individual level of information.

While the Shackelford method focused on age estimation in hominin fossils, it was validated for use in forensic casework by Kamnikar and colleagues (2018). This study showed that a modern reference sample provided more narrow confidence intervals and close age estimations than the previous model, suggesting that calculated age-at-death ranges, while smaller, could

7

results in a more constrained comparison of antemortem and postmortem data in the identification process. This study also identified the tendency to underestimate age for individuals over the age of 5 years, a pattern that increases in error with an increase in age cohort. The larger sample per older cohort collected in this study, clarifies this issue.

New transitions were calculated from the data collected during this granting period. The recalculated MFH, MFH2, and MUS tables are used by the transition analysis code to calculate age at death. We compared the transition ages to the ages calculated by the test sample used in Kamnikar et al. (2018). Age estimation bias improved over much of early childhood, but we do see some over estimation in the later ages (~10-20) (Figure 2). In general, the revised transitions reduce the overall bias as compared to Shackelford et al.'s (2012) transitions (i.e., MFH's original transitions). Cases with ages from 10 to 20 years old will be re-examined to confirm developmental scores and transitions will be revised.
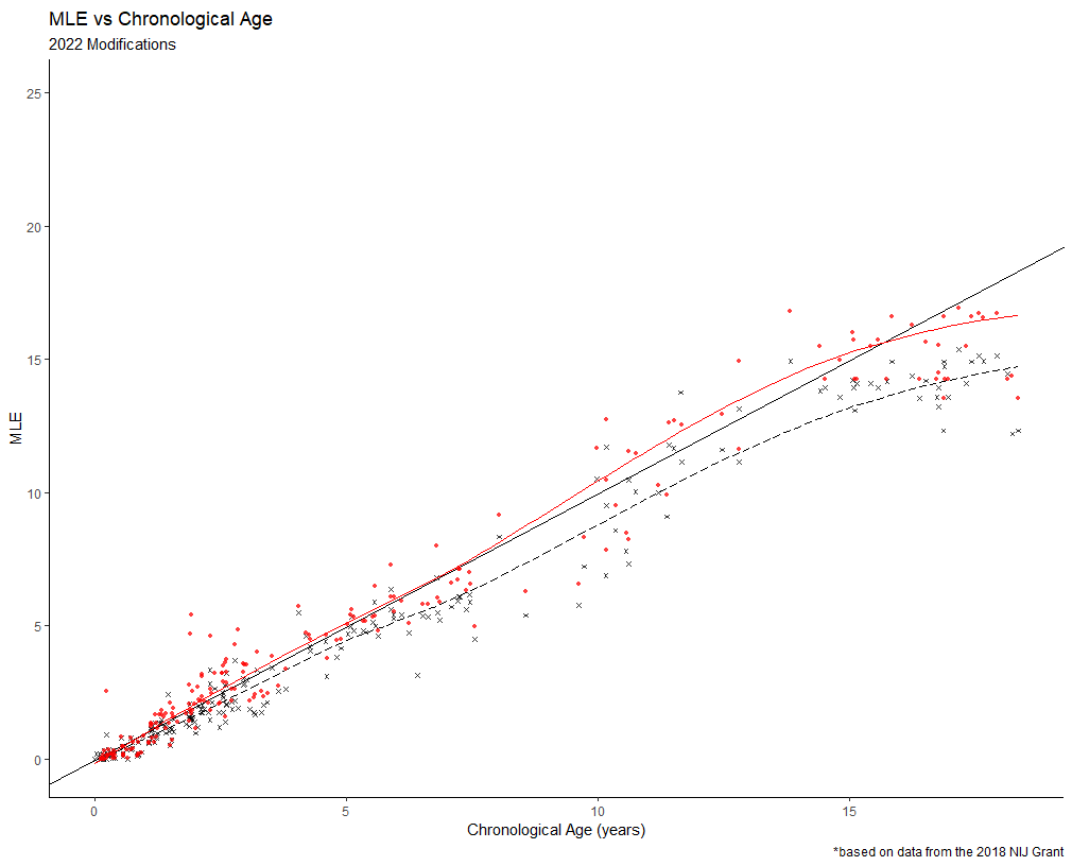
Figure 1. Comparison of original method and revised transitions for this present study.

As scoring nears completion, one major change for the estimation method will be in the tables used by R to calculate the age transitions. During analysis of the method with the new reference dataset, a train/test approach was followed to generate new model parameters. Here, 80% of the reference data were used to train the model, while 20% was used as a test sample for cross-validation. The data tables will be updated in the online tool once deemed appropriate and assess for error.

8

Machine Learning Methods

Additional age-estimation models were constructed with machine learning methods to compare with the accuracy of TA age estimation models. Machine learning algorithms depend on the speed and power of computers to permute models over many iterations to "tune" each for best fit. Many of these methods do not work directly with group parameters such as means and standard deviations; rather, they involve tuning thousands of random cutoff points in the sample using input features and output labels to establish a relationship among variables. Many of these models require complete datasets, so missing data can be a problem.

To circumvent missing data issues, we incorporated a multiple imputation by chained equation (mice) approach to impute missing data using the ***mice*** package (van Buren et al., 2011) in R (R Core Team 2022). As dental development is age dependent, the level and distribution of our missing data is not truly MCAR (missing completely at random). Rather, there is a nested structure. Deciduous teeth are missing more than adult dentition, and earlier developing teeth (e.g., *dc, dm1*) are missing more than other deciduous teeth. The 'mice' package permits data inspection, multiple imputations of missing values using several imputation approaches, imputation diagnostics, and pooling/majority voting of imputations for a final complete dataset.
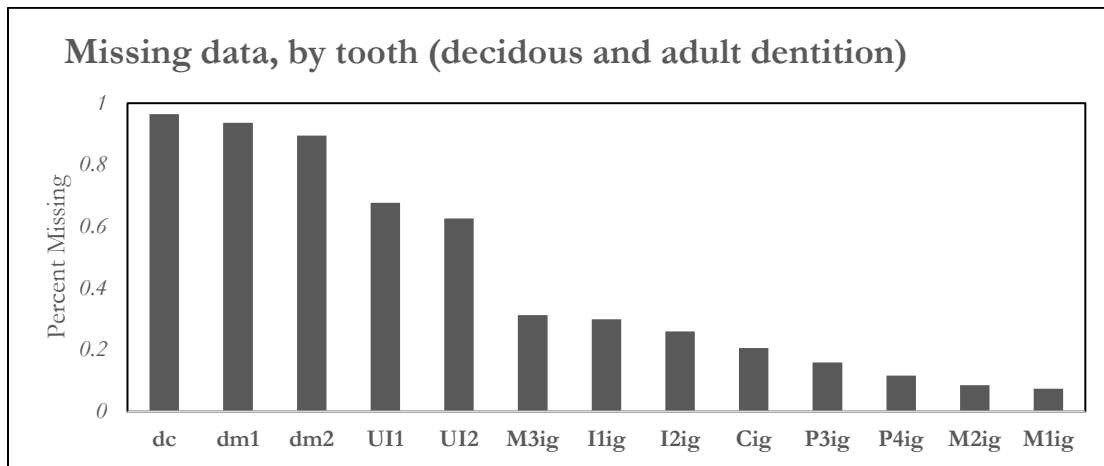
Figure 2. Distribution of missing data in the dataset used for machine learning methods.

For our purposes, we applied a built-in approach to deal with missing date that generates multiple imputations (replacement values) using all of the non-missing data, which van Buren and colleagues (2011) call *Fully Conditional Specification*. Each missing observation is imputed by a separate model built from available data. This is completed until all missing values are calculated. This process is then repeated *n* times and diagnostic tests are used to assess variable imputation. While the algorithm can impute mixes of continuous, binary, unordered categorical and ordered categorical data, our data are only ordered categorical, so we used the proportional odds model. Consistency between imputations was maintained by passive imputation and assessed using

9

diagnostic plots to verify the quality of the imputations. Figure 3 provides a density plot of all variables to inspect and compare the incomplete (blue) and imputed (red) data for a constrained solution. Ideally, these distributions would be nearly identical, and for some teeth there is a very good correspondence between the original and imputed values (c.f., UI1, I2ig). However, for others there is not very good correspondence, suggesting an amount of error may be introduced during the imputation process. There really is no ideal way to deal with missing data, particularly when the data are not "missing not completely at random" (MNCAR). Managing data expectations, accepting some level of bias in imputed values, and detailing the imputation approach can mitigate some of the issues associated with the imputation process. For our dataset, pairwise deletion would so diminish our sample that we would effectively have no data to work with. Subsetting the adult dentition does not improve the outcome. So, the imputed data are used to build and assess our machine learning models.
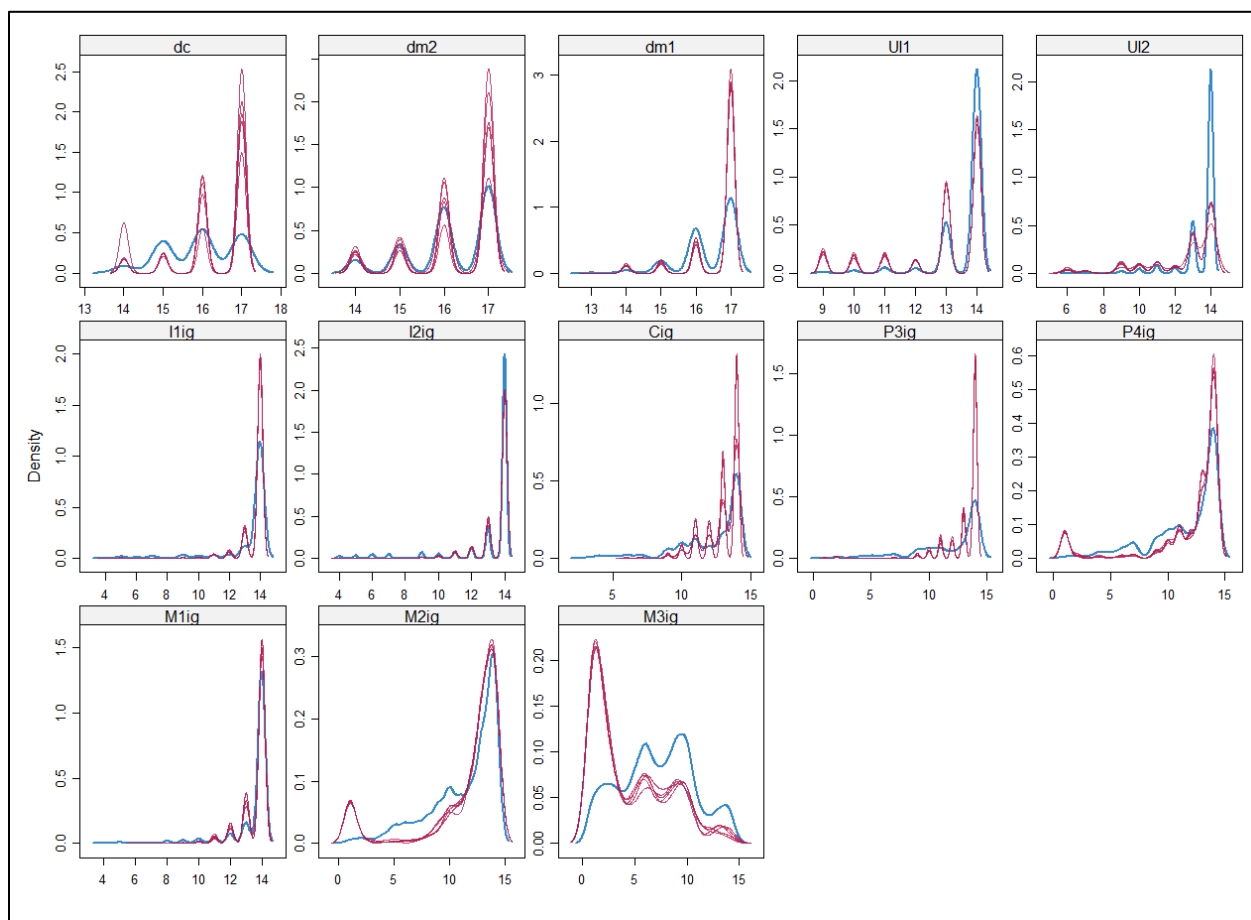
Figure 3. Density plots showing original, incomplete (blue) and imputed (red) distributions of scores for each tooth.

To assess age estimation models using dental development codes, we implemented several machine learning algorithms. These methods include artificial neural networks (aNNs), decision trees (DTs), generalized boosted regression, multivariate adaptive regression splines (MARS), random forest models (RFMs), and ridge/lasso regression. We used train/test cross-validation to ensure model accuracy and to avoid overfitting. The train/test approach splits the data into two

10

sets: a training set and a testing set. We elected to use 70% of the original data to train our model and 30% for testing. Each model is initially created using the training set; the test set is then used to generate new predictions using the initial model. Model assessment is assessed and, if necessary, the model is tuned and the process repeated. A robust ML algorithm needs substantial training data to properly learn how to interact with and predict variables and observations. We maximize training, test, and, if necessary, validation data to ensure the algorithm works as expected, including for future observations. Quantity alone is not enough, however, when it comes to the data used to construct our ML algorithms. The quality of those data is just as important. This has meant collecting as much real-world data from bite-wing and panoramic radiographs, etc. to capture many of the forms of input on which our future predictions might rely. This kind of real-world data is critical for ML methods, to capture forms mimicking how an application will receive user input in the future. This ensures our ML algorithms have the best chance of succeeding. Finally, we have included as much diversity (age, population, sex) in our ML sample as possible. This is essential to eliminate artificial intelligence bias, where an algorithm works better for a certain segment of the population than others because of training bias. With AI bias, the ML algorithm delivers results that may register as prejudice or impartiality against a one segment or group. To avoid this, we make every effort to train algorithms with artifacts comprising an equal and wide-ranging variety of inputs.

Each machine learning method used the same imputed dataset, comprising 7581 individuals. All analyses were conducted in R (R Core Team 2022) using a variety of packages (Table 4). Figure 4 provides a scatterplot of the actual and predicted age (and pseudo R-square value) for each of the machine learning approaches we assessed. These data are presented along with our results from the MLE estimates to provide a measure of how well each performs in the estimation of age using dental codes. Prior to any analyses, numeric features were scaled to avoid adverse influence on the modeling process. Centering and scaling were achieved using the '*preProcess*' function from the *caret* package. Next, full models are built for each of the ML methods using the training data. These models are then used to predict new observations (the test data) and model performance assessed. We evaluated model performance using two metrics: R-squared value and Root Mean Squared Error (RMSE). Ideally, lower RMSE and higher R-squared values indicate a good model fit to the data.

All the ML models performed well, with R-squared values ranging from 0.8876 (Gradient Boosted Regression) to 0.9021 (Random Forest Models). However, none of these performed as well as the MLE model with transition analysis, which had an R-square of 0.9644. These results suggest promise with ML models, however. Future efforts will center on additional tuning of these algorithms, more robust (and potentially more realistic) imputation methods better suited to MNCAR data, and more refined approaches that account for age/sex/population interaction. The R code for the ML analysis can be found at https://github.com/hefnerj1/MLDentalAge.

Table 4. Library (and purpose) for code used in machine learning models.

Library	Purpose	Reference
caret	Data processing	Kuhn (2021)
descr	Descriptive statistics	Enzmann et al (2021)
dplyr	Data manipulation	Wickham et al. (2021)
earth	Multivariate Adaptive Regression Splines	Milborrow (2021)
factoextra	Data mining	Kassambara and. Mundt (2020)
FactoMineR	Exploratory data analysis	Kassambara and. Mundt (2020)
gbm	Gradient Boosted Machines	Greenwell, Boehmke, Cunningham (2020)
ggiraphExtra	Graphics	Moon (2020)
ggplot2	Graphics	Wickham et al. (2021)
ggrepel	Graphics	Slowikowski (2021)
ggthemes	Graphics	Arnold (2021)
mice	Data Imputation	van Buuren and Groothuis-Oudshoorn (2011)
nnet	Artificial Neural Networks	Venables and Ripley (2002)
party	Decision trees	Hothorn, Hornik and Zeileis (2006)
plyr	Data manipulation	Wickham (2011)
randomForest	Random Forest	Liaw and Wiener (2002)
RColorBrewer	Graphics	Neuwirth (2022)
reshape2	Data manipulation	Wickham (2007)
rpart	Classification and Regression Trees	Therneau and Atkinson (2022)
VIM	Data Imputation	Kowarik and Templ (2016)
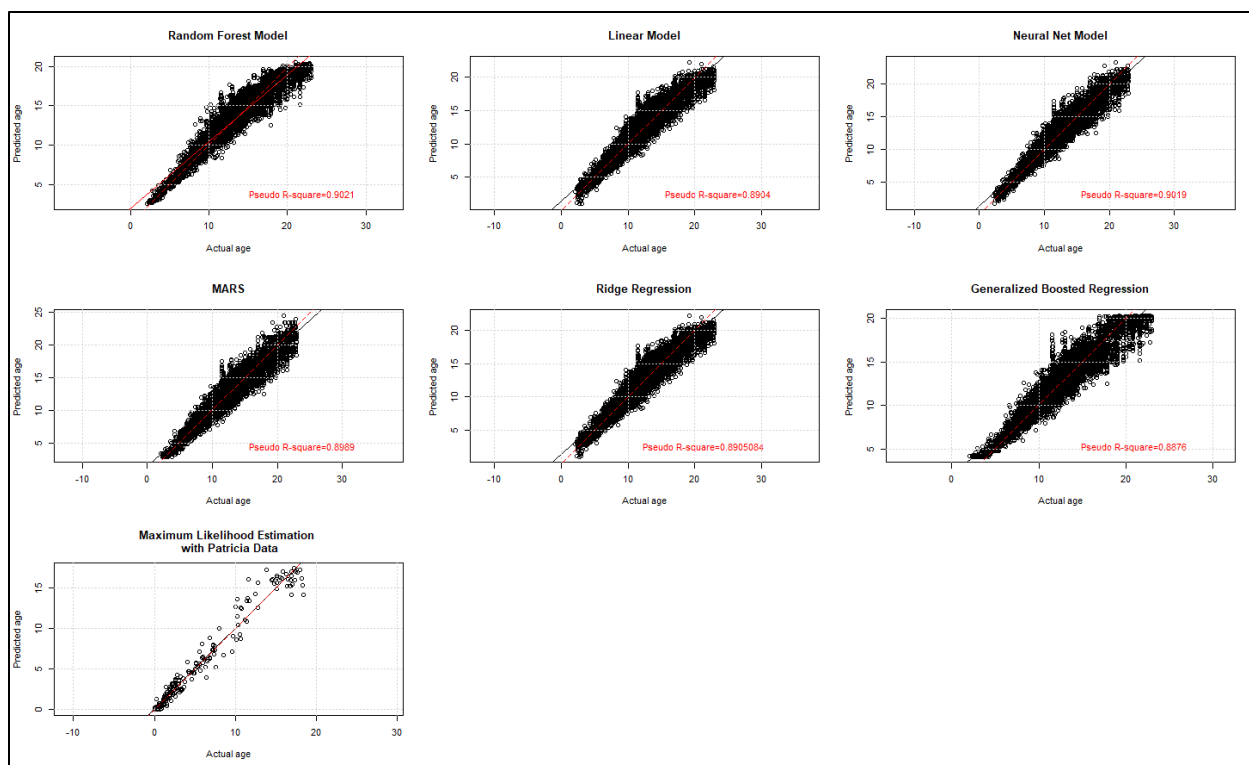
Figure 4. Scatterplot of actual (x-axis) and predicted (y-axis) age for each of the machine learning algorithms. Pseudo-R square values are included in the lower right quadrant for each.

Findings

During this research period, the majority of deliverables from this project were tangible, and include data collection, workshops and presentations, and the analytical tool (TADA). Additionally, we identified areas for future research and analysis.

Presentations, Publications, and Workshops

The PI and two graduate researchers involved with the project were invited to the annual meeting of the Asociacion Latinoamericana de Antropologia Forense in Santa Marta, Colombia (2018) and Puno, Peru (2019) to provide a hands-on workshop on transition analysis and age estimation for forensic practitioners in Mexico, Central, and South America. These workshops were mostly in Spanish and involved the use of basic coding and analytical skills using the statistical computing program R to run the analysis.

During the pandemic, the same team provided a virtual training for the Fundación de Antropología Forense de Guatemala on transition analysis and dental age estimation. This workshop directly stemmed from the ALAF workshop. Additionally, a senior graduate student researcher on the project presented TADA and the method to forensic odontologists and forensic anthropologists working on the Instituto Nacional de Ciencias Forenses in Guatemala City.

13

Additionally, aspects of this project were incorporated into several undergraduate and graduate student research projects. These research projects were presented at university-level conferences and national forensic meetings and are listed under **Dissemination of Research Findings**.

Analytical Tool

The Transition Analysis Dental Age Estimation tool (TADA) is a publicly available analytical tool for estimating age using dental development codes. TADA is currently in the late developmental stages and a Beta Version is expected to release in 2022. The program is currently in a testing phase by grant researchers, who collaboratively work and report issues via the GitHub platform.

TADA displays a user-friendly interface where practitioners can navigate through a clickable juvenile and adult dental arcade. The option to change dental arcades based on dentition present is user friendly for cases with mixed dentition. Within the arcade, practitioners choose the developmental code that best captures the stage exhibited by the selected tooth (indicated by highlighting the tooth in blue). Graphics displaying the developmental stages are situated to the right of the dental arcade. The program will automatically advance to next tooth for scoring and all scored teeth are highlighted in gray. Advanced user settings allow the users to choose between various tooth numbering systems, 2D or radiographic depictions of developmental stages, and the ability to disable automatic advancement. and the option to use the tool in English or Spanish. This allows for greater dissemination to an international audience.

After all present and observable teeth are scored, the user clicks "Review and Analyze" at the bottom of the interface which directs the user to the review page. Here, a list of all scored teeth and associated scores can be easily viewed before running the age analysis. After review, the user clicks "Calculate results" to run the age analysis. The left column on the interface provides numerical results, including the predicted age in years with the lower and upper boundaries of the age range estimate. A list of teeth and associated scores used in the analysis is also provided in the left column. In the right column, two graphical images of the overall age estimate and individual tooth estimates are provided for visual reference. Once dental developmental scores are entered, the application calculates age and produces two resulting graphics. The first graphic shows a TA generated age density plot and provides a maximum likelihood age estimate (point estimate) and confidence interval data (range estimates) based on teeth scored. Statistical measures of variance will be included for each tooth, as well as confidence intervals at 90% and 95%. The second graphic is a density plot of the TA estimate for each tooth scored. The purpose of this graphic is to show the transition likelihoods for each tooth, which, in theory, should align with all teeth scored. This graphic serves as a check for the user if any tooth densities are further ahead or behind others. Even though the statistics are robust, the Shiny application will provide an easy-to-use platform for all practitioners regardless of their background in statisticalanalysis like the computer program ADBOU for adult aging (Milner & Boldsen 2012).

14

The dissemination of TADA allows for the application of this method in an exponentially more user-friendly environment. Prior to the tool, we distributed R code for the developed models, finding that it can be intimidating and not appealing to practitioners. Therefore, this phase of the project is essential for demonstrating the utility of the method and the likelihood that it will be used in forensic investigations.

Future ideas for TADA include a warning message for scores that are not reasonable or expected for teeth, similar to the warning codes in FORDISC 3.0 and 3Skull. Additionally, a Read Me file will be written in English and Spanish to facilitate ease and appropriate use of TADA.

Implications for Criminal Justice Policy and Practice in the United States

Given that subadults represent a large portion of the population of missing and unidentified persons in the United States, it is important that we continue efforts to improve identification methods for this group. Because it is not yet possible to reliably determine sex in subadult skeletal remains (outside of DNA analysis) and very little research has been conducted on determining ancestry for these individuals, age determination is currently, the most reliable method of narrowing down candidate lists for identification.

Therefore, it is crucial that we aim to produce the most accurate age-at-death estimations possible. This project aims to generate more refined coverages in age, which allows for more accurate reported age ranges to further refine candidate lists. This method will meet the *Daubert* standards for expert witness testimony and the support for the methods employed in their practice. Further, in consultation with a board-certified forensic odontologist, the method will conform to the American Board of Forensic Odontology best practices and standards. The analytical tool offers a freely available computer software program to estimate age-at-death using dental development that can be incorporated into forensic casework, while conforming to best practice and international standards.

Future Directions

While data was collected from many different samples over the period of the grant, we are working collaboratively with institutions in India, Cyprus, and Guatemala for additional data collection. Data from these areas is important as they represent diverse sources of reference data for examining age estimates and transitions.

Further testing will include patterning of dental development data across populations to information the most appropriate model structures. We will do this first be examining the best performing combination of observed teeth, as well as the minimum number of teeth required to obtain an accurate determination. Next, we will examine the relationship between dental development score patterns and demographic parameters (i.e., sex and ancestry) to determine if significant differences in dental development occur between groups.

15

With the addition of individuals from the NMDID sample, we aim to look at socioeconomic status (SES) and its impact on dental development and age. Research suggests that dental development proceeds according to a strict growth schedule; however, with census level data available with this sample, addressing the impact of SES on individual transition and the overall estimate is worthwhile.

Finally, we noted discrepancies in scoring dental development for specific teeth. We think it is essential to develop standards to reduce error in coding teeth for dental development by novice practitioners or those unfamiliar with using the Moorrees et al (1963a,b) method. This is especially true for deciduous resorbing teeth and distinguishing between certain molars, like dm2 and M1, or M2 and M3. Efforts will target specific areas of difficulty to ensure all practitioners can identify and code each tooth with certainty.

Dissemination of Research Findings

Project Website

A website for the project is available at https://tada.wp.txstate.edu/. The website includes a landing page describing the project goals and source of funding, a team page with links to contact information and profiles for each project member, and pages for R code and references. The R code page comprises the full R code needed to run the age analyses with instructions, as well as downloadable tables needed to perform the analyses. The reference page provides project deliverables as well as other useful and necessary literature for users to understand the project background.

When the analytical tool is complete, there will be a GitHub link where practitioners can download the program directly along with a readme file. We chose GitHub for sharing as it allows for users to report issues with the tool in an open communication with the developer and team members.

Presentations

2022

Sauter CR, Kamnikar KR, Hefner JT, Herrmann NP, Spence JE. The impact of sex and population affinity on the dental development of subadult individuals. Poster accepted for presentation at the annual meeting of the American Academy of Forensic Sciences in Seattle, WA.

2021

Sauter CR. The impact of sex and ancestry on dental development. Mentors: Joseph Hefner and Kelly Kamnikar. University Undergraduate Research & Arts Forum, Michigan State University.

16

2020

Kamnikar KR, Plemons AM, Herrmann NP, Spence JE, Heuzé Y, Hefner JT. Comparison of two dental age estimation methods within transition analysis. Poster presented at the annual meeting of the American Academy of Forensic Sciences in Anaheim, CA.

Workshops

2021

Kamnikar KR. Six-day intensive training course on craniometric and macromorphoscopic data collection using ThreeSkull and FORDISC; transition analysis in age estimation using TA3 and TADA. Workshop presented in Spanish for the anthropologists in the Subárea de Antropología. *Instituto Nacional de Ciencias Forenses (INACIF)*, Guatemala City, Guatemala, June 2-9.

Kamnikar KR, Robledo I, Herrmann NP. Análisis transicional: aplicaciones para estimar la edad en restos humanos esquelitizados. Workshop presented in Spanish for the *Fundación de Antropología Forense de Guatemala (FAFG)*, Guatemala City, Guatemala (virtual), May 14.

2019

Kamnikar KR, Herrmann NP, Robledo I. Análisis transicional para la estimación de la edad biológica. Workshop presented in Spanish at the XV Congress of the *Asociación Latinoamericana de Antropología Forense* in Puno, Peru. October 21-25.

2018

Kamnikar KR, Herrmann NP, Plemons AM. Análisis transicional para la estimación de la edad biológica. Workshop presented in English and Spanish at the XIV Congress of the *Asociación Latinoamericana de Antropología Forense* in Santa Marta, Colombia. October 22-26.

17

Participants

Name: Nicholas P. Herrmann, PhD
Project Role: PI
Nearest Person Month Worked:
Contribution to the Project: Oversees project
International Collaboration: N/A

Name: Joseph T. Hefner, PhD
Project Role: Co-PI
Nearest Person Month Worked:
Contribution to the Project: Analytical tool development
International Collaboration: N/A

Name: Jennifer E. Spence, PhD
Project Role: Post-Doctoral ResearcherNearest Person Month Worked: 6
Contribution to the Project: Coordinate Graduate Research
International Collaboration: N/A

Name: James Fancher, DDS, PhD
Project Role: Senior ResearcherNearest Person Month Worked: 2
Contribution to the Project: Direct Quality Control Assessments
International Collaboration: N/A

Name: Kelly Kamnikar
Project Role: Graduate ResearcherNearest Person Month Worked:6
Contribution to the Project: Data collection, data processing
Funding Support: Graduate research assistant
International Collaboration: N/A

Name: Petra Banks
Project Role: Graduate ResearcherNearest Person Month Worked:5
Contribution to the Project: Data collection, data processing
Funding Support: Graduate research assistant
International Collaboration: N/A

Name: Amber Plemons
Project Role: Graduate ResearcherNearest Person Month Worked:1
Contribution to the Project: Data collection, data processing
Funding Support: Graduate research assistant
International Collaboration: N/A

Name: Ivanna Robledo
Project Role: Graduate ResearcherNearest Person Month Worked:5
Contribution to the Project: Data collection, data processing
Funding Support: Graduate research assistant
International Collaboration: N/A

Bibliography

Demirjian AH, Goldstein H, Tanner M. A new system of dental age assessment. *Human Biology* 1973;45:211.

Edgar H, Daneshvari BS, Moes E, Adolphi N, Bridges P, Nolte K. New Mexico decedent image database. Office of the Medical Investigator, University of New Mexico: Albuquerque, NM. 2020.

Feijóo G, Barbería El, De Nova J, Prieto JL. Dental age estimation in Spanish children. *Forensic Science International* 2012;223(1-3):371.e1-371.e5.

Heuzé Y., Braga, J. Application of Non-adult Bayesian Dental Age Assessment Methods to Skeletal Remains: the Spitalfields Collection. *Journal of Archaeological Sciences* 2008:35(2):368-375.

Heuzé Y, Cardoso HFV. Testing the quality of nonadult Bayesian dental age assessment methods to juvenile skeletal remains: The Lisbon Collection children and secular trend effects. *American Journal of Physical Anthropology* 2008;135:275-283.

Kamnikar KR, Herrmann NP, Plemons AM. New approaches to juvenile age estimation in forensics: Application of transition analysis via the Shackelford et al. Method to a diverse modern subadult sample. *Human Biology* 2018:90(1):11-30.

Liversidge HM. Permanent tooth formation as a method of estimating age. In: Koppe T, Meyer G, Alt KW, et al., (eds). *Comparative Dental Morphology*. Basel: Karger; 2009: 153-157.

Milner GR, Boldsen JL. Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology* 2012;148:98-110.

Moorrees CFA, Fanning EA, Hunt EE. Age variation of formation stages for ten permanentteeth. *Journal of Dental Research* 1963a;42:1490-1502.

Moorrees CFA, Fanning EA, Hunt EE. Formation and resorption of three deciduous teeth in children. *American Journal of Physical Anthropology* 1963b;21:99-108.

Phillips VM, van Wyk Kotze TJ. Dental age related tables for children of various ethnic groups in South Africa. *Journal of Forensic Odontostomatology* 2009;27:29-44.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011:45(3), 1-67. https://www.jstatsoft.org/v45/i03/

Appendix A: Data Collection SOP

TEXAS ★ STATE
UNIVERSITY ®
The rising STAR of Texas

Project Title	**Investigation of subadult dental age-at-death estimation using transition analysis and machine learning methods ("NIJ Dental")**
Project Personnel	
PI	Dr. Nicholas P. Herrmann 🐾
Co-PI	Dr. Joseph T. Hefner 🦌
Forensic Odontologist	Dr. J.P. Fancher 🐾
Postdoctoral Research Associate	Dr. Jennifer Spence 🐾
Graduate Research Assistants	Petra Banks 🐾
	Kelly Kamnikar 🦌
	Amber Plemons 🦌
Undergraduate Research Assistants	Olivia Green 🐾
	Carolynn Sauter 🦌
	Adrienne Stainton 🐾
Date	October 14, 2019
Revision Date(s)	January 23, 2020
	🐾 Texas State University 🦌 Michigan State University

STANDARD OPERATING PROCEDURE

A. Purpose

This SOP 1) describes procedures for selecting, opening, enhancing, scoring, and closing images and 2) contains contact information in case scores need to be corrected and/or questions arise. The Moorrees, Fanning, and Hunt (MFH) staging worksheets are available under NIJDental/Reference Tools for Scoring and NIJ/SCORING FILES/Reference Tools for Scoring.

B. Scope

This SOP is intended for all personnel working on the "NIJ Dental" project.

C. Procedures

1. Selecting Image for Scoring
   a. Find the folder with your name under NIJDental/SCORING FILES.
   b. Within your personal folder, select the folder labeled "Images TO BE SCORED."
   c. In this folder you should see multiple images that need to be scored. If this folder is empty, contact Jennifer (jes322@txstate.edu) so she can add images for you to score.

2. Opening Image for Scoring
   a. Use Photoshop or Windows Photo Viewer (or another program of your choice) to open the image you wish to score.

3. Enhancing Image for Scoring
   a. Depending on the quality of the image you're scoring, you may need to "enhance" the image to make it more legible. NOTE: Evaluate the image for clarity BEFORE you enhance it.
   b. DO
      1. Use the contrast and/or brightness controls to enhance image (Photoshop).
      2. Use the light and/or clarity controls to enhance image (Windows Photo Viewer).
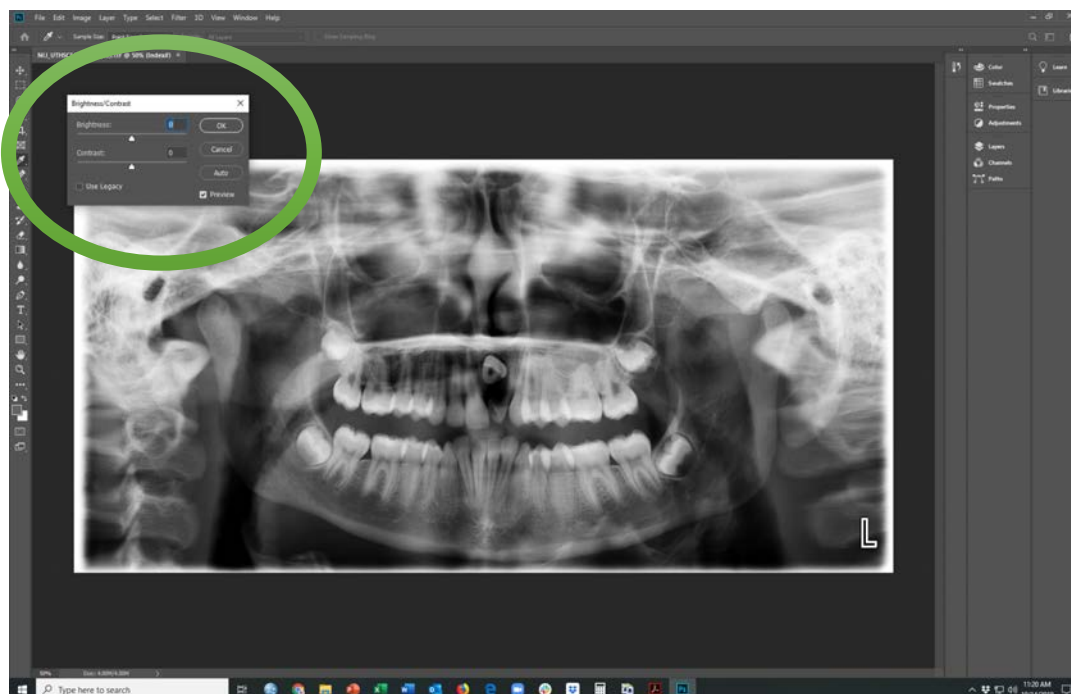
     c. DO NOT
         1. Use the sharpen tool to enhance image (Photoshop).
         2. Save image with any enhancements or modifications.
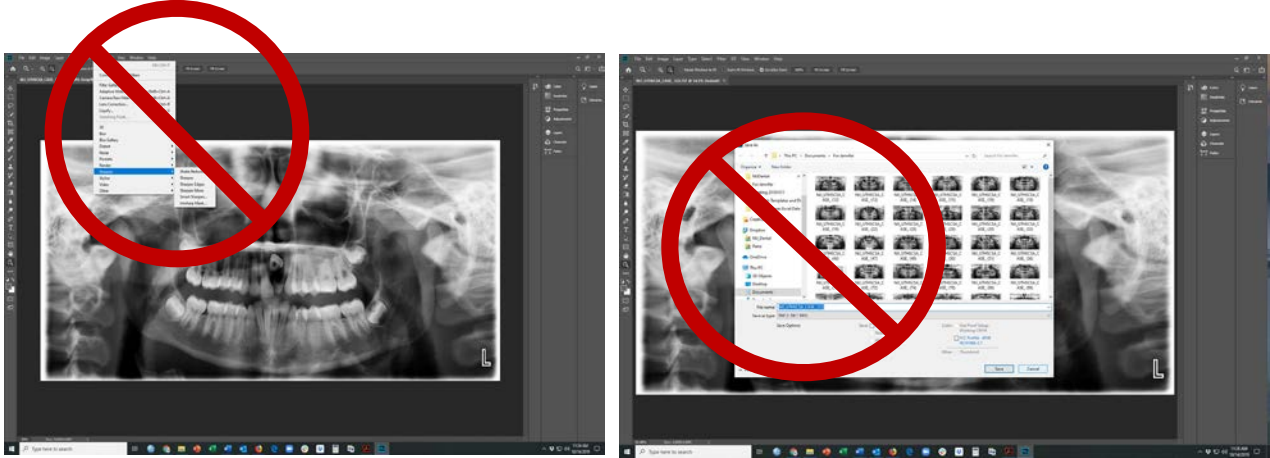
4. Scoring Images
     a. Open the Shiny app using Google Chrome (avoid Internet Explorer, Microsoft Edge, and Firefox; with these programs the Shiny interface will not be scaled appropriately).

https://macromorphoscopictraitanalysis.shinyapps.io/TA_Dental/

     b. There are three tabs in Shiny—Data Collection, Treatment, and Responses. Shiny will default to the Data Collection tab. Begin here.
     c. On the Data Collection tab in Shiny
         1. Enter the image number using all four digits (e.g., 0001).
         2. Enter your initials using three letters.
         3. Select the collection from the options. These codes should correspond to the first part of the file name of the image you're scoring.
         4. The current date (Date Entered) will appear automatically. You do not need to enter information in this box (unless you wish to change the date).
         5. If any of the teeth have a restoration (e.g., a root canal, crown, etc.) or orthodontic treatment (e.g., retainer, braces, etc.) check the box

titled "Restoration/Orthodontia Present." NOTE: If you check this box, you will need to fill out the relevant information (identify specifics) on the Treatment tab for each tooth affected.

6. Look carefully at the image you're scoring and rate each quadrant for clarity (poor, fair, or good). NOTE: Evaluate the image for clarity BEFORE you enhance it.

7. Use the guides above the empty fields in Shiny (which contain the Universal Numbering System numbers for each tooth and the anthropological label) to identify each tooth. ALWAYS use the UNS number when referring to a tooth. The anthropological notation is optional. NOTE: Charts with numbered teeth, images and descriptions for each scoring stage, and other reference materials are available under NIJDental/Reference Tools for Scoring and NIJ/SCORING FILES/Reference Tools for Scoring.

Shiny has three tab options. The app defaults to the Data Collection tab.

The UNS number and anthropological abbreviation are above each tooth. ALWAYS use the UNS number when referring to a tooth in Comments; the anthropological reference is optional.

Select the collection appropriate to the image you're scoring.

Enter the four-digit number of the image you're

Enter your initials using three letters.

Shiny will default to the current date. Change it only if necessary.

Check the box if any of the teeth in the image have restorations or orthodontic treatment. If you check this box, you will need to fill out information on the Treatment tab.

Evaluate each quadrant for clarity BEFORE you enhance the image.

4

8. Use the Guidelines for Scoring the Dentition (below) for each tooth.

GUIDELINES FOR SCORING THE DENTITION

INVENTORY	Begin by completing an **inventory of all 52 teeth** (permanent *and* deciduous). JP likes to go tooth by tooth, beginning with the maxillary right third permanent molar (#1). To orient yourself, you can look for the first permanent molar (M1) in each quadrant, but keep in mind that we might have images of individuals who are too young to have the M1 fully or even partially formed. Take your time! You will get faster as you gain more experience.
SCORING OPTIONS	For each tooth, there are four basic options for scoring. **1) The tooth is present and you are able to score it.** In this case, use the MFH staging worksheets as a guide to assign the tooth a numeric score. The following are the most likely values you will need for the images in our samples. 1–14 (no 8) — Developmental stages for single-rooted permanent teeth 1–14 — Developmental stages for multi-rooted permanent teeth 15–17 — Resorption stages for deciduous teeth **2) The tooth is present but you are unable to score it.** If you are sure a particular tooth is present but you are unable to ascertain its developmental score, enter 9999. Different raters have different levels of experience; it's okay if you are unsure. It's better to enter 9999 than a potentially incorrect score. **3) The tooth is not present.** If you are certain that a particular tooth is absent, leave the cell blank. Because of the age range in our samples, it is possible that many of the deciduous teeth and some of the permanent will be absent. **4) You cannot tell whether or not a tooth is present.** Occasionally the image quality, especially around the edges (e.g., M3s), will be such that you cannot actually determine whether a particular tooth is present or not. In this case, enter 8888. In addition, use 8888 for both premolars (in the same quadrant) for cases in which one premolar is present but you cannot determine which premolar it is.

SCORING OPTIONS (continued)	

COMMENTING ON TEETH	**Use the Universal Numbering System (1–32 and A–T) when you make comments about a tooth or teeth.** Including the anthropological designation is optional.
GENERAL SCORING	**Read the scoring descriptions for each stage.** Do not use the visual guides only. Refer to the MFH staging sheets frequently to make sure you're assigning the correct stage to each tooth.
	If formation appears to fall between two stages, **round up**.
	It is relatively common for **antimeres** (the same teeth on opposite sides of the mouth) to be at different stages. In addition, due to the nature of panoramic radiography, antimeres may appear as different sizes. **Score each individual tooth according to its own individual stage of formation.**
	For most teeth, **crown height** is about a third the length of the complete root, except for canines. Canine crowns are about one quarter the length of the complete root.
	Remember that **single-rooted teeth have no stage 8**.
PREMOLARS	When you encounter an image with only one premolar in a quadrant and **you cannot tell which premolar is present, enter 8888 for both premolars** and then **make a note in the Comments box with the score of the remaining premolar**, using the Universal Numbering System to specify each present premolar.
	If you can see that a premolar is **double rooted or has two canals**, make a note in the Comments box.

In this image (UCMAD_0625), you can see that each maxillary quadrant has only one premolar. Because it's not possible to tell which premolar (P3 or P4) is present, you should enter 8888 for each premolar and then make a note in the Comments box with the score for each premolar that is present, the premolar using the Universal Numbering System designation.

PREMOLARS (continued)

Sometimes **you may be able to determine which premolar in a quadrant is present**; if this is the case, make a note in the Comments box, leave the cell blank for the appropriate tooth, and score the present premolar if possible or enter 9999.

In this image (UCMAD_0625), it's clear that the lower P4s are absent, so you can score P3 and leave the P4 cell blank.

MOLARS	If two roots of the same tooth appear to have different levels of formation, **score the most developed root.**
	If **only one molar root is observable**, enter 9999 (i.e., do not score the tooth).
	It may be difficult to see the **interradicular area** (between the roots) in maxillary molars even though the roots appear to be relatively advanced.
DECIDUOUS TEETH	For **deciduous teeth**, it may be difficult to tell if roots are forming or resorbing. Look at the surrounding teeth to give you a clue. If you are unsure but you know the tooth is present, enter 9999.
	If a deciduous tooth is present but the **permanent tooth succeeding the deciduous tooth is absent**, do not score the deciduous tooth (enter 9999 instead).
PATHOLOGY	If the image appears to indicate a **pathology** (for example, trauma or a cleft palate) but the **tooth root and crown do not appear to be affected by the pathology, go ahead and score the tooth**. If you're unsure, enter 9999.
	If the tooth has apparent **major disease** (like caries), enter 9999 as the tooth may not be forming normally.
	It may be difficult to score a tooth if there is a **periapical pathology** (a lesion or inflammation around the apex of the root), especially if the pathology is associated with advanced caries in a developing tooth. Use 9999 if you are unsure.
RESTORATIONS	When a tooth has a **large restoration** (like a crown) or **major pulpal treatment** (root canal or treatment affecting more than half of the pulp), **do not score the tooth** and check the box indicating "Restoration/Orthodontia Present" on the Data Collection page and fill out the relevant information on the Treatment tab ("R" for restoration, "O" for orthodontic treatment, or "B" for both) for each tooth affected. Use the Comments box on either the Data Collection or Treatment page to enter details if you like.
	If the tooth has a **small restoration** that appears to affect one or several small surfaces only (the occlusal or buccal pits, for example) or a **minor pulpal treatment** (affecting less than half the pulp) and the tooth appears to be developing normally, go ahead and **score it (if you can)** or enter 9999 (present but not scoreable). Check the box indicating "Restoration/Orthodontia Present" on the Data Collection page and fill out the relevant information on the Treatment tab ("R" for restoration, "O" for orthodontic treatment, or "B" for both) for each tooth affected. Use the Comments box on either the Data Collection or Treatment page to enter details if you like.
	Keep in mind that restorations using **different materials may radiopaque differently**.

When any of the teeth in the dentition has a form of **orthodontic treatment** (braces, retainers, etc.), check the box indicating "Restoration/Orthodontia Present" on the Data Collection page and fill out the relevant information on the Treatment tab ("R" for restoration, "O" for orthodontic treatment, or "B" for both) for **each tooth that has a band, bracket, and/or or is spanned by an orthodontic apparatus, even if the tooth is not serving as an anchor**.

ORTHODONTIC TREATMENT

In the image above (BORDX_0685), on the Treatment page, you would enter the letter "O" for teeth 3, 6, 7, 8, 9, 10, 11, and 14. Because either the third or fourth premolar is missing in each maxillary quadrant, you cannot tell whether 4 or 5 and 12 or 13 is present, so you would leave the Treatment cell

|	
ORTHODONTIC TREATMENT (continued)	blank but make a note in the Comments box. And, even though only teeth 3, 7, 8, 9, 10, and 14 are affected by bands or brackets, the other teeth should be noted to have orthodontic treatment (an "O") because they are spanned by the apparatus. In the Comments box, make a note about which teeth are spanned, have brackets, and/or have bands. A clue to whether an apparatus is attached to the teeth or not (versus being removable) is whether you see radiopacity caused by the resin used to attach the appliance to the tooth. The person in this image (BORDX_0694) has a removable dental appliance. For this individual, you would check the box indicating "Restoration/Orthodontia Present" on the Data Collection page and on the Treatment page enter "O" for teeth 3, 7, 8, 9, 10, and 14 since those are the only teeth you can tell for certain that are affected by the apparatus, and you would make a note in the Comments box about what you see.

5. Submitting Scores and Closing Images
   a. When you have entered scores for all present teeth, left blank fields for absent teeth, and are sure about your decisions, click on the Submit button on either the Data Collection or Treatment page.
   b. Close the image file without saving changes.

c. Move the image file from the folder labeled "Images TO BE SCORED" to the folder labeled "SCORED Images" under your personal folder in NIJDental/SCORING FILES.

d. Enter the information about the scored image into the Excel file titled "Log of Cases Scored" under your personal folder in NIJDental/SCORING FILES.

6. Correcting Scores

a. Though you can check what your scores after you've hit the Submit button, it is not possible for you to change scores once you have submitted them in Shiny. If you submit your scores and realize you made a mistake (omitted teeth, entered incorrect values, etc.), or if you submitted a score before you were ready, make a note in the Comments box and send an email to Jennifer (jes322@txstate.edu) with the collection name and image number. She will either delete the entry or open it up for you to revise.

Though you can't change your scores once you've submitted them, you can check what you entered on the Response page.

D. Contact

If you have any questions or concerns, please contact Jennifer at jes322@txstate.edu (preferred) or (505) 507-3075.