An Investigation and Critique of the DiGangi et al. (2009) Age-at-Death Estimation Method

Sara M. Getz, BS., Department of Applied Forensic Sciences, Mercyhurst College, Erie, PA

Introduction

Most methods of age-at-death estimation provide age ranges that are either too narrow or too wide to be of practical use, employ terminal age categories (such as 50+), and fail to provide prediction intervals using an explicit probability. To address these issues, the DiGangi et al. (2009) first rib age-at-death estimation method utilizes transition analysis and features of the first rib previously investigated in the Hamann-Todd collection by Kunos et al. (1999). The first rib was chosen because it is easily identifiable, likely to be preserved in forensic and archaeological contexts, and has been shown to exhibit remodeling into the eighth decade of life.

The method is described as a quick and accurate way to capture age-related information in any part of the lifespan. DiGangi et al. (2009) claimed that the categorical scoring method should reduce inter- and intraobserver error; however, the wide age intervals and large overlap between age stages (Figure 1) suggest that even if reliable scoring is possible, this method may have little practical value. The purpose of this study is to evaluate the accuracy, reliability, and practical value of this method.

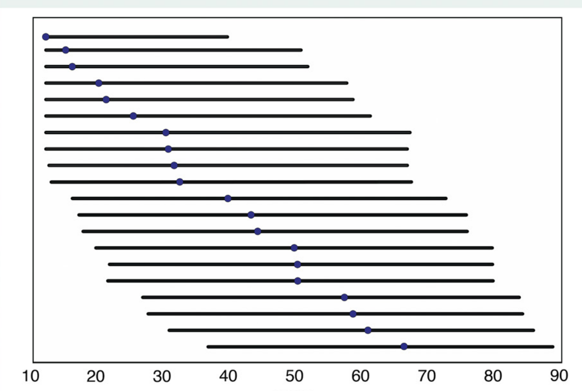

Figure 1. Graphical representation of the age intervals associated with the composite scores of the DiGangi et al. (2009) method found in Figure 8 of the origional publication

Materials and Methods

A sample was selected from all white males present in the Hamann-Todd Collection, resulting in a uniform age distribution in which all age groups were equally represented (Table 1). Individuals whose ribs were damaged, missing from the collection, or retained costal cartilage that obscured features to be scored, as well as those individuals exhibiting notable trauma or pathology, were excluded from the sample. Three graduate students with advanced osteological training, in addition to the author, were provided with copies of the original publication for review and reference. One week prior to the first data collection period, a meeting was held to discuss trait definitions and to practice applying the method.

According to the original publication, the application of this new method requires only that observers familiarize themselves with the descriptions of the traits to be scored and the example photos found in Appendix A of the article (Figure 2), score the features of the ribs as described, and refer to the table of appropriate age prediction intervals and best point estimates of age.

Table 1. Comparison of the samples used in the DiGangi et al. (2009) study and this study

	Original Study	This Study
Sample	Identified individuals from mass graves in former Yugoslavia	Hamann-Todd Collection (CMNH)
Sample Time Period	20th century	19th century
Ancestry	Bosnian white	American white
Sex	Male	Male
Sample Size	470	452
Observation Type	Independent observations	113 scored by 4 observers
Sample Range	12–90 years	21–88 years
Sample Mean	47.7 years	53.2 years

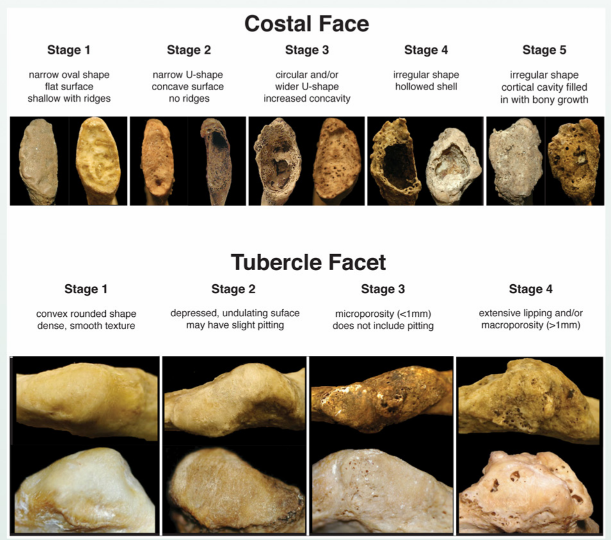

Figure 2. Costal face and tubercle facet stages with descriptions and images adapted from those found in Appendix A of the DiGangi et al. (2009) publication

Data were collected in two sessions. The four observers each scored the same sample of 113 individuals independently. The coding descriptions and features being scored were not discussed during the data collection period. One month after the original data collection, a subsample of up to 30 individuals from the total sample was re-coded by each observer to allow for the calculation of intraobserver agreement statistics.

Minitab 15 (2007) was used to check for data entry errors and visualize the data. The "irr" package in R.2.10.10 (2009) was used to assess the levels simple and extended inter- and intraobserver agreement for the costal face, tubercle facet, and combined scores. The extended agreement calculation employs a tolerance threshold to determine what constitutes agreement between scores; the tolerance is the distance that can exist between a pair of scores and the scores will still register as agreement.

Kunos et al. (1999) suggested that the features being coded are ordinal variants; however, this author maintains that without prior knowledge of the progression of age-related changes in bone, the stages described are not an inherently ordinal progression. To provide the most applicable and flexible statistics, the data were analyzed using tests for both nominal and ordinal data. Most notably, weighted and unweighted kappa were calculated for each observer and each observer pair for both the costal face and tubercle facet scores. Unweighted kappa treats all disagreements between observer scores as equal and is appropriate for nominal data. Weighted kappa weights differences between observer scores based on the magnitude of the disagreement and is appropriate for ordinal scales.

Results

The four observers correctly placed 84.3% of the sample individuals into an age interval that contained their true age. Chi-squared tests showed no significant differences between the frequencies of correct classification by each observer. This level of accuracy is lower than what would be expected given the use of the 95% probability density intervals. This method is highly accurate for individuals between 20 and 50 years of age and becomes increasingly less accurate from 55 years onward (Table 2). With the exception of four individuals between 20 and 35 years of age, all individuals incorrectly aged were above 55 years of age. The tendency of this method is to overestimate age in individuals under 40 years of age and to underestimate age in individuals over 40 years of age (Figure 3).

Table 2. Observed correct classification percentage (accuracy) experienced by each observer in each decade and in the pooled samples

		Observers			
Decade	**Pooled**	**1**	**2**	**3**	**4**
20s	93.3	100.0	93.3	93.3	86.7
30s	92.6	94.1	88.2	100.0	88.2
40s	100.0	100.0	100.0	100.0	100.0
50s	96.4	100.0	100.0	85.7	100.0
60s	73.7	68.4	78.9	73.7	73.7
70s	66.7	60.0	86.7	46.7	75.0
80s	61.5	69.2	69.2	53.8	53.8

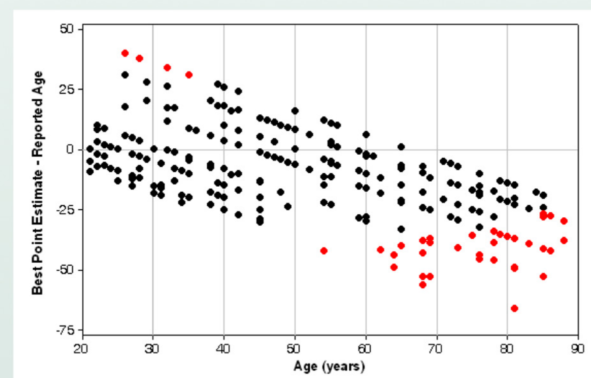

Figure 3. Distribution of the best point estimate of age minus the reported age from collection records for all individuals in the pooled sample (N= 452). Individuals whose reported age fell outside the age interval corresponding to their composite score are shown in red.

Interobserver Statistics

The percent of agreement between the scores of observers for the costal face was typically higher than for the scores of the tubercle facet (Tables 3 and 4). If two observers did not choose the same stage for a trait, the costal face was often scored only one stage apart; however, the tubercle facet was often scored two stages apart. The inability to reliably code the tubercle facet introduced much of the variability seen in the composite scores for each individual.

The interobserver differences were so large that in only 62.7% of cases did the scores of the four observers for an individual result in a composite stage (shown in Table 8 of DiGangi et al. [2009]) that were within six stages of each other. However, due to the large overlap in the age ranges provided for each unique combination of costal face and tubercle facet scores (see Table 1), multiple observers can produce more or less the same estimated age interval while having only minimal agreement in their scores for each rib feature (Figure 4).

The higher agreement between scores for the costal face was also seen in both forms of Cohen's Kappa statistic. The values obtained the weighted form of this statistic were higher than those obtained for the unweighted form for both features (see Table 4). Weighted Cohen's Kappa provided the highest value for interobserver agreement seen (Table 5).

Intraobserver Statistics

The patterns seen above in the differences in agreement for the costal face and tubercle facet (Table 6), as those as for weighted and unweighted Cohen's Kappa (Table 7), were also seen in the intraobserver recode sample data. Intraobserver agreement was higher for each feature than interobserver agreement; but it was still low, given the relatively simplistic nature of the scoring system.

Table 3. Simple and extended percent agreement by observer pairs by trait scored

	Costal Face				Tubercle Facet		
Observer Pair	**0**	**1**	**2**	**3**	**0**	**1**	**2**
1-2	63.7	94.7	98.2	100.0	51.3	72.6	99.1
1-3	48.7	75.2	93.8	100.0	48.6	75.2	97.3
1-4	58.4	74.3	99.1	100.0	58.4	74.3	99.1
2-3	62.8	93.0	97.3	99.1	45.1	74.3	95.6
2-4	71.7	93.8	96.5	99.1	59.3	80.5	98.2
3-4	65.5	90.3	95.6	97.3	39.8	67.3	93.8

Table 4. Comparison of unweighted and weighted forms of Cohen's Kappa by feature for the scores of each observer pair

	Costal Face		Tubercle Facet	
Observer Pair	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
1-2	0.50	0.74	0.29	0.45
1-3	0.51	0.69	0.23	0.37
1-4	0.59	0.72	0.33	0.49
2-3	0.47	0.72	0.26	0.45
2-4	0.61	0.73	0.41	0.56
3-4	0.51	0.62	0.13	0.26

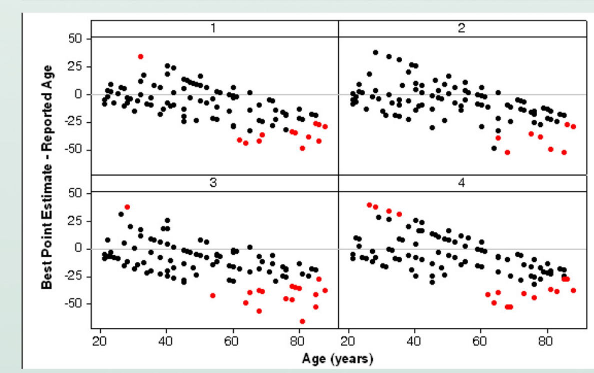

Figure 4. Distribution of the best point estimate of age minus the reported age for each individual in the sample of 113 individuals coded by four observers.

Table 5. Comparison of calculated interobserver agreement statistics

	Costal Face	Tubercle Facet
Kendall's Coefficient of Concordance W	0.66	0.52
Spearman's Rho Rank Correlations	0.66	0.47
Cohen's Kappa (weighted)	**0.74**	**0.56**
Cohen's Kappa (unweighted)	0.61	0.41
Fleiss' Kappa	0.53	0.27
Conger's Kappa	0.53	0.27
Light's Kappa	0.53	0.28
Iota (combined)	**0.50**	

Table 6. Simple and extended percent intraobserver agreement for each observer by feature scored

		Costal Face				Tubercle Facet		
Observer	**n**	**0**	**1**	**2**	**3**	**1**	**2**	**3**
1	27	77.8	96.3	100.0	----	77.8	85.2	100.0
2	30	66.7	93.3	100.0	----	60.0	93.3	100.0
3	27	67.9	85.7	96.4	100.0	46.4	82.1	96.4
4	28	63.0	100.0	----	----	44.4	85.2	100.0

Table 7. Comparison of unweighted and weighted forms of Cohen's Kappa by features for the scores of each observer and themselves

	Costal Face		Tubercle Facet	
Observer	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
1	0.69	0.86	0.58	0.67
2	0.55	0.79	0.42	0.77
3	0.56	0.64	0.26	0.55
4	0.44	0.84	0.29	0.59

Conclusions

The overall correct classification rate of the DiGangi et al (2009) method for the sample in this study (84.3%) was lower than expected using the published 95% intervals. Their method has high accuracy until approximately 55 years of age (because of very wide intervals), after which the accuracy declines with increasing age. Their method tends to overestimate the age of individuals under 40 years of age and underestimate the age of individuals over 40. Despite the apparent simplicity of the coding system provided, we found high interobserver differences, perhaps due to the use of composite scores based on multiple features. However, the high levels of interobserver error have relatively little impact on the overall accuracy of this method, due to the wide and overlapping age intervals provided for each composite score.

Given these results, it is likely that the DiGangi et al (2009) method does not capture age-related information in the most valuable manner. Although the authors state that their method can contribute to a multi-factorial approach to age-at-death estimation, the wide age ranges and systematic aging bias seen in this method will hinder rather than aid in most analyses. The use of transition analysis represents one of the most promising statistical approaches to age-at-death estimation; however, if the first rib contains information that is useful for age estimation, the DiGangi et al. (2009) method forces morphological variants of the first rib into supposedly age-progressive composite stages that seem to obscure this meaningful age-related variation.

Acknowledgements

Funding for this research was provided by the Faculty-Led Student Research Grant Program at Mercyhurst College and sponsored by Dr. Steven A. Symes. I would like to thank The Cleveland Museum of Natural History and Lyman Jellema for allowing access to the Hamann-Todd collection, Dr. Stephen D. Ousley for providing his statistical expertise and guidance, my graduate student cohort for their assistance in data collection and documentation, and David Pedler for his invaluable assistance with poster design and production.

References

DiGangi EA, Bethard JD, Kimmerle EH, and Konigsberg LW. 2009. A New Method for Estimating Age-at-Death from the First Rib. Am J of Phys Anthropol 138:164-176.

Kunos C, Simpson S, Russell K, Hershkovitz I. 1999. First rib metamorphosis: its possible utility for human age-at-death estimation. Am J Phys Anthropol 110:303–323.

Minitab 15 Statistical Software. 2007. [Computer software]. State College, PA: Minitab, Inc. www.minitab.com.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.Rproject.org.