Using Multivariate Adaptive Regression Splines to Estimate Subadult Age from Diaphyseal Dimensions

Kyra E. Stull, PhD[1]; Ericka N. L'Abbé, PhD[1]; Stephen D. Ousley, PhD[2]

[1] Department of Anatomy, Faculty of Health Sciences, University of Pretoria, South Africa
[2] Department of Anthropology, Mercyhurst University, Erie PA

Abbreviated title: Subadult Age Estimation from Diaphyses

Corresponding Author:
Kyra Stull
Department of Anatomy
University of Pretoria
Private Bag x 323
Arcadia 0007
South Africa
Email: kstullster@gmail.com
Tel: (864) 230-2301

ABSTRACT

Subadult age estimation is considered the most accurate parameter estimated in a subadult biological profile, even though the methods are deficient and the samples from which they are based are inappropriate. The current study addresses the problems that plague subadult age estimation and creates age estimation models from diaphyseal dimensions of modern children.

The sample included 1,310 males and females between the ages of birth and 12 years. Eighteen diaphyseal length and breadth measurements were obtained from Lodox Statscan radiographic images generated at two institutions in Cape Town, South Africa between 2007 and 2012. Univariate and multivariate age estimation models were created using multivariate adaptive regression splines (MARS). K-fold cross-validated 95% prediction intervals (PIs) were created for each model and the precision of each model was assessed.

The diaphyseal length models generated the narrowest PIs (two months to six years) for all univariate models. The majority of multivariate models had PIs that ranged from three months to five and six years. Mean bias approximated zero for each model, but most models lost precision after 10 years of age. While univariate diaphyseal length models are recommended for younger children, multivariate models are recommended for older children where the inclusion of more variables minimized the size of the prediction intervals. If diaphyseal lengths are not available, multivariate breadth models are recommended. The present study provides applicable age estimation formulae and explores the advantages and disadvantages of different subadult age estimation models using diaphyseal dimensions.

Keywords: age estimation; radiographs; diaphyseal lengths; diaphyseal breadths, juvenile

Age is often the sole parameter estimated by an anthropologist for a subadult biological profile. In contrast to adult age estimation, which is based on degenerative patterns, subadult age estimation is based on morphological and metric evaluation of indicators associated with growth and development. Dental development is considered the most accurate age indicator, but when dentition is unavailable, anthropologists often utilize long bone lengths (Ubelaker, 1987; Cardoso, 2007; Saunders, 2008; Franklin, 2010; Cardoso et al., 2013). However, nearly all age estimation methods based on diaphyseal lengths suffer from numerous problems that affect both the ability to estimate age and the accuracy of the age estimation. The errors in age estimation can be grouped into two categories, misapplication and the use of inappropriate samples.

Misapplication is a two-fold problem: (1) the failure to provide prediction intervals for an estimated age and (2) the difficulties in estimating age when diaphyses are present but the length measurements are unavailable. According to a 2009 survey of board certified forensic

anthropologists, the most common method to estimate age in subadult remains, when the dentition is not available, is to measure diaphyseal lengths and compare these to growth charts from longitudinal studies. Thus, one of the fundamental reasons for misapplication is the difference between research designs. Biological anthropologists use growth studies designed to assess diaphyseal lengths for known age and sex in normal children. Therefore, results were presented as percentiles, which documented normal variation in diaphyseal lengths. In contrast, biological anthropologists and bioarchaeologists estimate age given diaphyseal length and provide a prediction interval. One repercussion of different research designs is the lack of a prediction interval (PI), which fails to meet *Daubert* criteria. *Daubert* emphasizes scientifically tested methods with quantifiable findings (Dirkmaat et al., 2008; Christensen and Crowder, 2009; Ousley and Hollinger, 2009). Thus, methods are required to have a known error rate to be considered admissible scientific evidence (Melnick, 2005).

A PI provides bounds with an associated probability in which future observations should fall. For example, 95% prediction intervals should encompass the true age of the deceased 95% of the time; an error rate of 5% is associated with such an age estimate. Most of the original publications of longitudinal growth studies (i.e. Ghantus, 1951; Anderson et al.81, 1964; Maresh, 1970; Gindhart, 1973) allow for the derivation of an 80% PI by using the 10^{th} and 90^{th} percentiles. However, the 80% PI reflects the range of variation of diaphyseal lengths per age and not the variation in age per diaphyseal length; consequently, it cannot be used to estimate age.

While recent studies (i.e. Rissech et al., 2008, 2013; López-Costas et al., 2012; Cardoso et al., 2013) were designed to estimate age from long bone lengths, the authors only provided regression formulae with the standard error (SE) and R-squared values. Providing a SE implies

that the variation in diaphyseal dimensions is constant through all ages, even though variation in diaphyseal dimensions is known to increase with age. López-Costas et al. (2012) recommended the SEs should not be used until the remains are estimated to be approximately 8 years of age, thus the authors acknowledged the inability of the SE to capture variation in the younger individuals. The variation in diaphyseal lengths continues to increase after 8 years of age, so the SEs will only be valid at 8 years of age, not before or after.

The second component to misapplication is the difficulty of estimating age if long bone length measurements are unavailable. Anthropologists routinely suggest using more than one indicator to estimate age in adults, as using more indicators usually yields a more accurate estimate (McKern and Stewart, 1957; Boldsen et al., 2002; Buckberry and Chamberlain, 2002; Uhl and Nawrocki, 2010). However, longitudinal growth studies and recent anthropological studies have only provided univariate assessments with no attempt to evaluate a multivariate approach for subadult age estimation. Furthermore, anthropologists regularly analyze skeletal material that has been affected by taphonomic processes (e.g. carnivore activity) (Cunningham et al., 2011; Pokines, 2014), which affect the anthropologists' ability to estimate age from diaphyseal length measurements. An age method based solely on length measurements is problematic and speaks to the need for the inclusion of breadth measurements and a multivariate approach. Currently, only a few breadth measurements on the humerus, tibia and femur have been considered for age estimation (Rissech et al., 2008, 2013; López-Costas et al., 2012).

Additional errors are compounded in age estimations when the reference samples are inappropriate. Skeletal collections that are available for analysis have birth dates that extend from the 18th century to the mid 20th century. Formulae derived from the historical samples are applicable in a bioarchaeological setting, but the application of the proposed formulae to modern

populations ignores the well-documented worldwide secular increase in childhood stature (Meredith, 1976; Eveleth and Tanner, 1990; Bogin, 1999; Freedman DS et al., 2000; Loesch et al., 2000; Malina, 2004; Marques-Vidal et al., 2008; Hawley et al., 2009; Anholts, 2013). Furthermore, populations vary in stature and proportions, which drastically affects the accuracy of age estimates. Inaccuracies due to population differences have been previously presented by Stull et al. (2013a) and Hoffman (1979). The authors identified differences between 30 mm and 60 mm when the long bones of modern South Africans and Eskimos, respectively, were compared to middle class white North American children born in the mid 1900's (Hoffman, 1979; Stull et al., 2013a). To ensure the highest accuracy in age estimates, formulae should only be used on populations from which they were derived.

In the 2009 survey, Maresh (1970) was noted as the most frequently used "method" when anthropologists evaluate subadult material. However, three reference samples, with differences in temporality and/or demographic composition, have demonstrated the inaccuracies of the Maresh "method" (Stull et al., 2008, 2013a). An 80% age interval for the Maresh results was derived by choosing upper and lower limits based on the maximum length of the long bone falling between the provided 10^{th} and 90^{th} percentiles. The percentage of times the Maresh estimated age interval was correct was used to quantify the accuracy. Using the 80% Maresh interval resulted in a 57% accuracy for a modern sample of North American children between birth and one year (Stull et al., 2008); a 22% accuracy for a modern sample of South African children between birth and 12 years; and 41% accuracy for a historic sample of South African children between birth and 12 years (Stull et al., 2013a).

Overall, longitudinal studies do not capture as much variation in a population compared to cross-sectional studies. The growth patterns of children sampled in a longitudinal study will

seem more homogeneous than they actually are because of autocorrelation. When children are measured repeatedly, each child's measurements are correlated to other measurements close in time, which violates assumptions of independent errors (Bock and du Toit 2004). Data heaping is an additional consequence of longitudinal study designs and further reduces variation per age (Bock and du Toit, 2004). As a result, the variability in diaphyseal measurements is far greater among children randomly sampled from a population in a cross-sectional study than the same children repeatedly sampled in a longitudinal. Ultimately, cross-sectional data is better suited for estimating age while longitudinal data is better suited for evaluating growth (Eveleth and Tanner, 1990; Ousley, 2013).

The availability of appropriate modern skeletal samples is the most significant impediment in anthropological research of modern subadults. Suitable samples require a large number of known-age individuals to adequately capture the variation in the population and make meaningful statistical statements (Konigsberg and Holman, 1999). Although radiographic images routinely generated at hospitals and morgues can account for the paucity of modern skeletons in collections, conventional radiography generates distorted images that limit the collection of measurements. In contrast, a radiographic system that generates images with minimal distortion, such as Lodox Statscan, would be ideal for metric data collection.

The purpose of this study is to provide an age estimation technique that addresses the problems associated with previous studies. The current study incorporates a large, cross-sectional modern sample, a large number of measurements collected on all six long bones from Lodox Statscan radiographic images, and univariate and multivariate models that permit the creation of 95% PIs.

Table 1 – Samples sizes by age, sex and institution.

Age (years)	**Red Cross**		**Salt River**		**Combined**
	Females	*Males*	*Females*	*Males*	
0	6	9	23	22	60
1	22	30	10	13	75
2	27	68	8	17	120
3	40	64	12	10	126
4	39	53	11	5	108
5	33	66	4	5	108
6	39	68	1	2	110
7	42	68	1	4	115
8	46	60	0	3	109
9	41	57	0	4	102
10	34	64	1	2	101
11	43	62	0	3	108
12	21	43	2	2	68
Total	433	712	73	92	**1310**

MATERIALS AND METHODS

Lodox Statscan (www.lodox.com) radiographic images of 1,310 South African children aged between birth and 12 years comprised the sample (Table 1). Birth dates of the children were all after 1996. The images were generated at the Red Cross War Memorial Children's Hospital (Red Cross) and the Salt River Forensic Pathology Service (Salt River) in Cape Town, South Africa between 2007 and 2012. The Red Cross sample contributed the most to the total sample (87%), but the individuals were largely between the ages of four and 12 years. The Salt River sample contributed less overall, but contained more individuals of younger age. Overall, the sample had a generally equal number of individuals throughout all age ranges with the smallest sample sizes in the two youngest age intervals and the oldest age interval. Age intervals were based on chronological age, which was calculated from date of birth and date of death or date of imaging; individuals aged between 1.00 and 1.99 years formed the 1 year olds, aged between 2.00 and 2.99 years formed the 2 year olds, and so on.

More males (n=804) than females (n=506) are present in the sample because each institution reported more males than females being examined annually (Douglas et al., 2010; Groenewald et al., 2011). The Salt River sample had known ancestry, but the Red Cross sample did not. Because South Africans self-identify with factors such as religion, language, and region (Christopher, 2002; Treiman, 2007), these variables were used to estimate ancestry in the Red Cross sample. Of the 54% with known or estimated ancestry, black South Africans were the most prevalent group (60%), coloured South Africans were the next largest group (39%) and white South Africans were the least prevalent with only seven individuals. Coloured South Africans are a self-identified group unique to South Africa that is recognized as the most genetically admixed population in the world, containing numerous intra-and inter-continental genetic contributions (Patterson et al., 2010; Statistics South Africa, 2012; Petersen et al., 2013). The frequency of coloured South Africans is higher and the frequency of black and white South Africans is lower in the current sample compared to the frequencies for the entire country because of data collection in the Western Cape Province and difficulties in the estimation of ancestry from demographic variables (Statistics South Africa, 2012).

Eighteen measurements were collected from the six long bones from each image with custom software designed for Lodox (Diagnostic Viewing Software) (Table 2). Length and breadth measurement definitions were modified from fetal osteology or from adult postcranial standard measurements (Fazekas and Kósa, 1978; Moore-Jansen et al., 1994). Written definitions with images can be found in the supplemental material online (Table S1, Figures S1 – S8). Measurements were only recorded when the bone was in anatomical position and from the side of the body that would yield the least distortion, which were generally measurements in the scan

Table 2 – The measurements and associated abbreviations.

Humerus diaphyseal length	HDL	Radius midshaft breadth	RMSB
Humerus proximal breadth	HPB	Femur diaphyseal breadth	FDL
Humerus distal breadth	HDB	Femur distal breadth	FDB
Humerus midshaft breadth	HMSB	Femur midshaft breadth	FMSB
Ulna diaphyseal length	UDL	Tibia diaphyseal length	TDL
Ulna midshaft breadth	UMSB	Tibia proximal breadth	TPB
Radius diaphyseal length	RDL	Tibia distal breadth	TDB
Radius proximal breadth	RPB	Tibia midshaft breadth	TMSB
Radius distal breadth	RDB	Fibula diaphyseal length	FBDL

direction, or y-axis (see below) (Stull et al., 2013b). If both left and right-sided elements were in proper position then measurements were collected from the left-sided elements.

Lodox Statscan

The Lodox Statscan (Lodox Systems Pty [Ltd] Sandton, South Africa) is a full body, fast acquisition radiographic device that was initially designed in South Africa for the diamond mining industry (Knobel et al., 2006; Evangelopoulos et al., 2009; Douglas et al., 2010). The slot-scanning design offers the ability to generate images with minimal distortion because the X-ray source is projected through a collimated fan-beam onto a detector, which moves in synchrony over the patient. Although distortion still exists, it is minimal in comparison to distortion generated by conventional radiography machines. Percent agreement between measurements collected on the dry bone and the Lodox Statscan-generated images of the dry bone was reported at 97% at the +/- 2 mm level (Stull et al., 2013b). Furthermore, average differences between skeletal measurements collected from Lodox Statscan radiographic images and the original skeletal material were 0.5% for y-axis measurements (i.e. lengths) and 4% for x-axis measurements (i.e. breadths) (Stull, 2013). Differences were comparable to differences between measurements obtained on computed tomography images and dry bone, error rates noted in prospective growth studies that had controlled radiographic settings, and intra- and inter-observer error rates on dry bone (Green et al., 1946; Maresh, 1955; Gindhart, 1973; Hoffman, 1979; Stull et al., 2014).

Statistical analysis

Fifteen individuals were randomly selected from the entire sample to estimate the intra- and inter-observer error. The number of measurement comparisons was 152 for the intra-observer error and 149 for the inter-observer error. The intra- and inter-observer errors were

assessed with technical error of measurement (TEM), relative technical error of measurement (%TEM), and Bland-Altman plots (Bland and Altman, 1986; Ulijaszek and Kerr, 1999; Goto and Mascie-Taylor, 2007).

Scatterplots of age regressed on a measurement were plotted with loess (locally smoothed regression) lines to visualize relationships. Loess uses locally weighted polynomial regression to fit a smoothed curve through data without making assumptions about the form of the relationship. Based on the graphs, the relationship between age and postcraniometric variables was nonlinear and reflective of a biological growth curve. Multivariate adaptive regression splines (MARS) was used, with age as the response variable (y) and the measurement(s) as the predictor variable(s) ($x_{(i)}$), to adequately model the nonlinear relationships. Because subadult sex estimation is problematic, and the authors did not want to incorporate any compounded errors in the age estimates, the sexes were pooled for all models.

Univariate age-at-death MARS models were created for each measurement. Additionally, multivariate subsets of the data were created to demonstrate the potential utility and benefits of multivariate models. Multivariate subsets included bone subsets, similar measurement subsets, and an all measurement subset (Table 3). Two sets of variables existed for each multivariate subset, the original measurements and the principal component (PC) scores from the variance-covariance matrix. The principal component analysis (PCA) removed high inter-variable correlations, which affects the ability of MARS to recognize important variables. Because PCA was used to remove multicollinearity and not to reduce dimensionality, all PCs were retained regardless of the proportion of variance. Three models were created and compared for each variable or multivariate subset to find the fit with the smallest residual standard error; the models

Table 3 – Measurements included in each multivariate subset.

	n	Variables
Humerus	497	HDL, HPB, HDB, HMSB
Radius	281	RDL, RPB, RDB, RMSB
Ulna	400	UDL, UMSB
Femur	778	FDL, FDB, FMSB
Tibia	521	TDL, TPB, TDB, TMSB
Upper Limbs	96	HDL, HPB, HDB, RDL, RPB, RDB, RMSB, UDL, UMSB
Lower Limbs	458	FDL, FDB, FMSB, TDL, TPB, TDB, TMSB, FBDL
Proximal Elements	316	HDL, HPB, HDB, HMSB, FDL, FDB, FMSB
Distal Elements	100	RDL, RPB, RDB, RMSB, UDL, UMSB, TDL, TPB, TDB, TMSB, FBDL
All Measurement	157	HDL, HPB, HDB, HMSB, UDL, RDL, FDL, FDB, FMSB, TDL, TPB, TDB, TMSB, FBDL*

*Some variables were excluded to retain a large sample size

included age without a transformation, age with a square root (sqrt) transformation, and age with a cubed root (cbrt) transformation.

The 95% PIs of each observation for each measurement were based on the spread of the predictions made on the out-of-fold (see below) data for each univariate model. The residuals and chronological age were plotted with loess lines to determine the precision (i.e. bias) of each model. In order to test the accuracy of the PIs, k-fold cross-validation was used. The authors acknowledge that testing the derived 95% PIs on the sample that created the intervals produces overly optimistic results.

Multivariate adaptive regression splines (MARS). Multivariate adaptive regression splines is a flexible, nonparametric regression modeling technique that requires no assumptions about the relationships among the variables (Friedman, 1991; De Veaux and Ungar, 1994; Butte et al., 2010). Linear regression, truncated basis functions, and binary recursive partitioning are incorporated into MARS to approximate the underlying function and model relationships (Muñoz and Felicísimo, 2004). The ultimate goal of MARS is to identify the basis function, *f(X),* which is estimated by subdividing *X* into regions and obtaining estimates of *f(X)* for each region (Friedman, 1991; Sekulic and Kowalski, 1992). Piecewise linear basis functions are conducted for each predictor variable (*x)* and every possible value of *t* and take the form of a constant, a hinge function, or a combination of hinge functions. Hinge functions take the form:

$$(x-t)_+ = \begin{cases} x-t & if\ x > t, otherwise \\ 0, & \end{cases}$$

and

$$(t-x)_+ = \begin{cases} t-x & if\ x < t, otherwise \\ 0, & \end{cases}$$

The knot, *t,* forms a reflected pair for each predictor variable and represents a change in slope or

transition between polynomials (Friedman, 1991; Sekulic and Kowalski, 1992; Muñoz and Felicísimo, 2004; Hastie et al., 2009; Butte et al., 2010). The hinge function that results in a positive value is used. For example, if hinge functions for a model are (HDL – 222) and (222 – HDL), the change in slope is at 222 mm and the first hinge function is used if the diaphyseal length is greater than 222 mm and the second hinge function is used if the diaphyseal length is smaller than 222 mm.

Forward and backward iterative selection constructs a set of basis functions for the final MARS model (Friedman, 1991, 1993). The forward pass deliberately overfits the training data and the backward pruning pass removes excess basis functions that no longer contribute to the accuracy of the fit, which is based on the generalized criterion value (GCV). The GCV is an approximation of the prediction error ascertained by leave-one-out cross-validation (Friedman, 1991, 1993; Milborrow, 2013). The knot and variable that provides the best fit are retained in the model and the coefficients for each function are estimated by minimizing the residual sum of squares (RSS) (Friedman, 1991; Sekulic and Kowalski, 1992; De Veaux and Ungar, 1994; Hastie et al., 2009).

Multivariate adaptive regression splines models were built through k-fold cross-validation. K-fold cross-validation separates the data into K equal sized parts. The model is fit with the $K - 1$ parts and the prediction error of the fitted model is calculated when predicting the k^{th} part, or the out-of-fold data (Efron and Tibshirani, 1993; Kohavi, 1995; Hastie et al., 2009). The process is conducted for $k = 1, 2, \ldots K$ and the prediction error is averaged across all out-of-fold predictions. The current study uses $K = 10$ because it allows for a lower variance, but a possibly higher bias; however, this bias is alleviated if the training sample has at least 100 observations (Efron and Tibshirani, 1993). Each training set will be a different size because the

dataset specific to each variable is different; the largest univariate samples (i.e. HDL, FDL) have training sets with more than 100 observations though all of the univariate breadths and multivariate samples will not. The out-of-fold R-squared is averaged (cross-validated R^2) from the left-out subset, which is an estimate of the model performance on independent data (Efron and Tibshirani, 1993; Hastie et al., 2009; Milborrow, 2013). The generalized R-squared (GRSq) is based on the raw GCV and is a generalization of model performance. Adding terms (or hinge functions) generally always increases the R-squared statistic, but the GRSq may actually decrease because of the reduction of predictive powers (Milborrow, 2013). Additional information on MARS can be found in Friedman (1991) and Hastie et al. (2009). The earth package in R was used to create the MARS models (Milborrow, 2013; R Core Team, 2013).

K-fold cross-validated prediction intervals. K-fold cross-validated 95% PIs were created for each model to quantify prediction accuracy (Melnick, 2005; Dirkmaat et al., 2008). Gaussian PIs were inappropriate because MARS models are nonparametric. The primary difference between a k-fold cross-validated PI and a classic parametric PI is how the variance of the predicted values is estimated. With resampling, the average variance of predicted values of the out-of-fold samples is calculated from 100 iterations of the k-fold cross-validated (K = 10) training models. Overall, cross-validation supplies a realistic estimate of prediction error because the average error is estimated from a sample that is different from the sample that created the models (Efron and Tibshirani, 1993). Additional details on cross-validation can be found in Efron and Tibshirani (1993) and Hastie et al. (2009).

RESULTS

The mean intra-observer TEM and %TEM are small at 0.45 mm and 0.22%, respectively (Table 4). The mean inter-observer TEM and %TEM were slightly larger at 0.76 mm and 0.40%, respectively (Table 4). The measurements with the highest inter-observer %TEM were midshaft breadths of the ulna and radius whereas the measurements with the highest intra-observer %TEM were the ulna midshaft breadth and radius proximal breadth. Neither the intra-observer error nor the inter-observer error showed systematic bias on the Bland-Altman plots. Most intra-observer and inter-observer measurement differences were within 2 mm (Figure 1). However, the spread of differences was larger for the inter-observer error than the spread of differences for the intra-observer error.

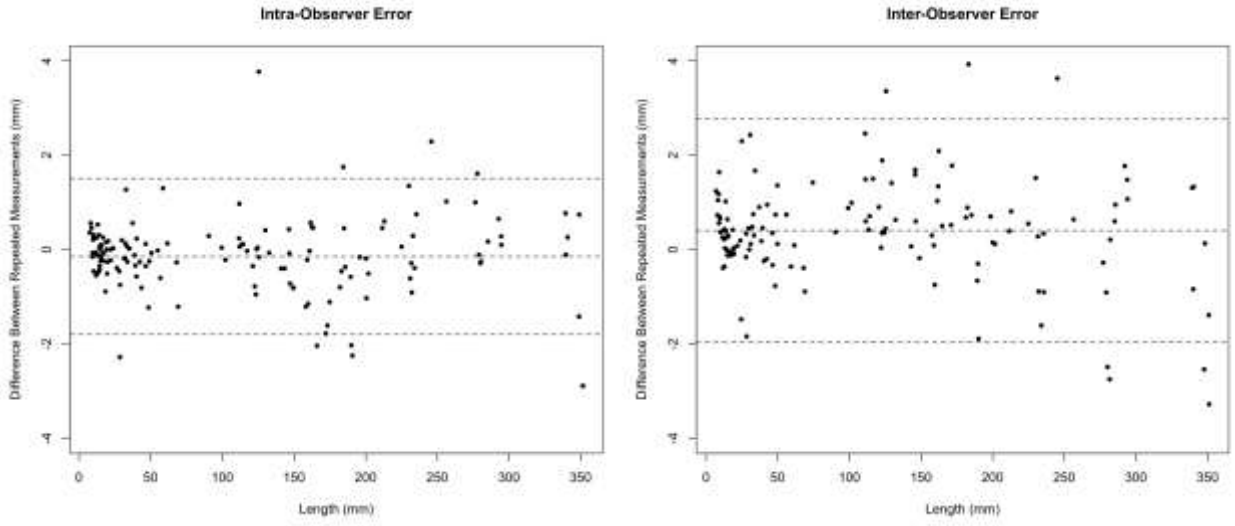

Fig. 1. Bland–Altman plots that illustrate the intraobserver error (left) and interobserver error (right).

Univariate models

Means, standard deviations and sample sizes for each measurement per age can be found in the supplemental material (Tables S2 – S7). The six diaphyseal lengths resulted with the smallest SEs and narrowest PIs, ranging from two months to five or six years (Table 5). Four of

Table 4 – TEM and %TEM for inter-observer error and intra-observer error.

	Inter-observer Error		Intra-observer Error	
	TEM	*%TEM*	*TEM*	*%TEM*
HDL	1.68	0.08	0.66	0.04
HPB	0.80	0.24	0.63	0.22
HDB	0.68	0.58	0.24	0.16
HMSB	0.30	0.22	0.29	0.20
UMXL	1.18	0.10	0.91	0.07
UMSB	0.66	1.92	0.30	0.84
RDL	1.15	0.06	0.35	0.02
RPB	0.23	0.49	0.26	0.75
RDB	0.02	0.04	0.25	0.44
RMSB	0.79	2.27	0.19	0.52
FDL	1.58	0.06	0.91	0.03
FDB	0.62	0.13	0.65	0.19
FMSB	0.15	0.09	0.16	0.12
TDL	0.87	0.06	0.70	0.04
TPB	0.40	0.16	0.29	0.09
TDB	0.81	0.49	0.19	0.21
TMSB	0.26	0.56	0.07	0.16
FBDL	0.48	0.02	0.71	0.03
Min	0.02	0.02	0.07	0.02
Max	1.70	2.27	0.91	0.84
Mean	0.76	0.40	0.45	0.22

Table 5 – The univariate models with the standard error (SE) and range of the k-fold cross-validated 95% prediction interval (PI). Variables are listed in decreasing order of SE.

	SE	**Size of 95% PI**
FDL	0.9	2 months – 5 years
TDL	0.95	6 months – 5 years
HDL	0.97	2 months – 6 years
FBDL	0.99	8 months – 5 years
UDL	1.01	3 months – 6 years
RDL	1.02	5 months – 6 years
TDB	1.35	2 years – 7 years
RPB	1.38	1 years – 8 years
TPB	1.39	2 years – 7 years
RDB	1.47	4 months – 9 years
FDB	1.51	2 years – 7 years
HDB	1.55	5 months – 10 years
TMSB	1.59	2 years – 10 years
HPB	1.6	1 years – 8 years
FMSB	1.62	1 years – 9 years
RMSB	1.76	5 months – 11 years
UMSB	1.76	9 months – 10 years
HMSB	2.18	3 years – 12 years

the five univariate midshaft breadths were the worst performing models, demonstrated by the largest SEs and widest PIs, ranging from five months to 12 years. The proximal and distal breadths clustered between the diaphyseal lengths and midshaft breadth measurements, with PIs ranging from four months to 10 years. Prediction intervals were narrower for the younger ages and wider for the older ages (Fig. 2). The associated diaphyseal measurements were removed if the lower 95% PIs included ages younger than birth. An example of the age estimation formula using the femur diaphyseal length is provided in Table 6. The remaining MARS formulae and k-fold cross-validated 95% PIs for the univariate diaphyseal length models can be found in the supplemental material (Tables S8 – S18).

Table 6 – MARS model for femur diaphyseal length with a cubed root transformation (cbrt) on age. The models is significant at p < 0.0001. The residual standard error (RSE) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cube root of age
(Intercept)	2.234***
h(FDL - 282.32)	0.006***
h(282.32 - FDL)	-0.008***
h(FDL - 167.91)	-0.002***
h(FDL - 342.03)	-0.002***
Observations	1117
cv R^2	0.95
Adjusted R^2	0.95
RSE	0.90
Note:	*p < 0.01; **p<0.001; ***p<0.0001

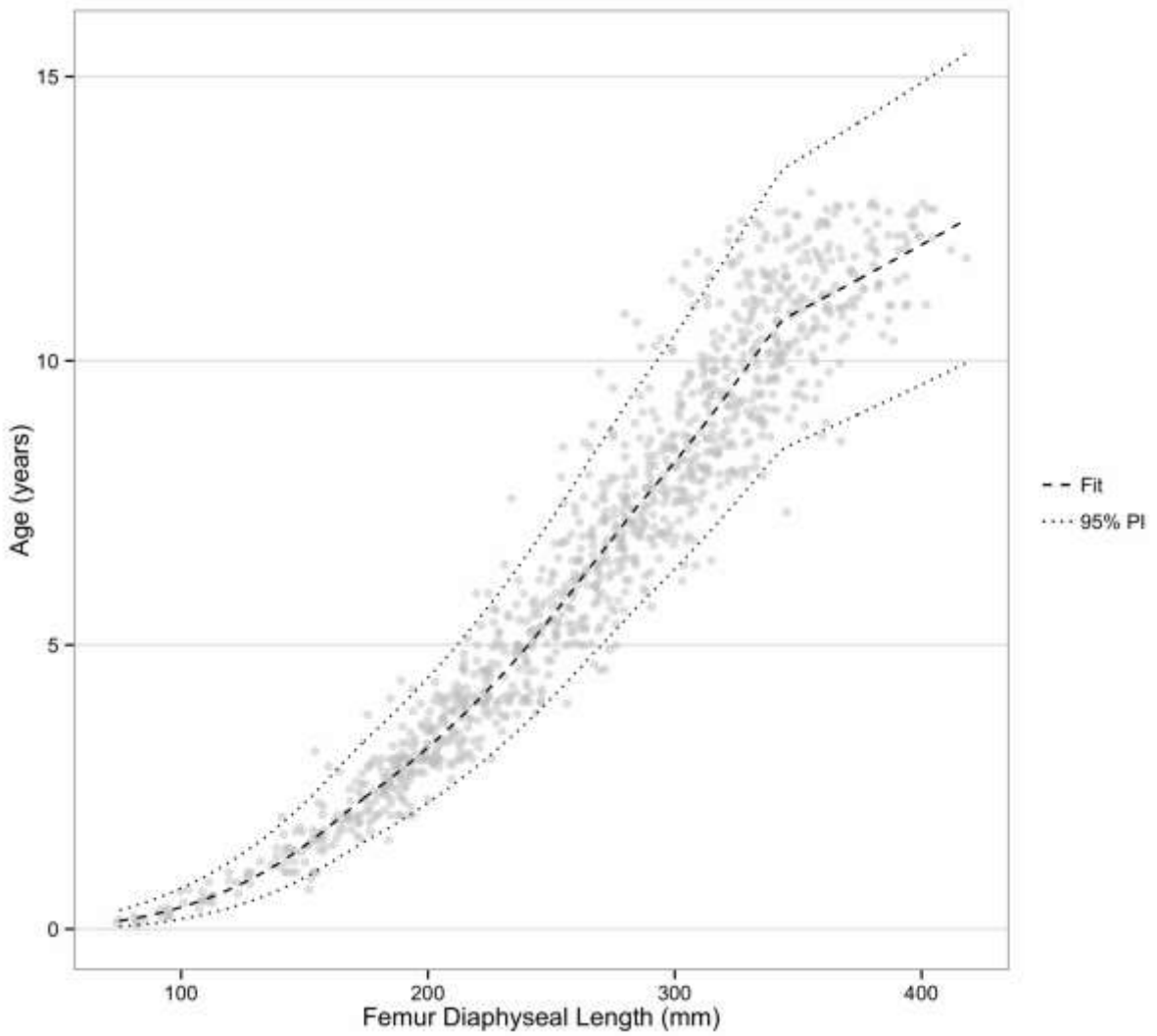

Fig. 2. The cross-validated 95% prediction intervals when age is regressed on femur diaphyseal length. Note how the prediction interval adjusts as age increases.

Although the bias was normally distributed, the mean bias always approximated zero, and the majority of individuals were aged within one year of their chronological age, an increase in error with an increase in age was apparent, showing heteroscedasticity in every univariate model. The bias ranged between -3 and +3 years, though most diaphyseal length models were unbiased until approximately 10 years of age (Fig. 3). Following 10 years of age, the degree of underestimation increased as age increased. Diaphyseal breadth models were more imprecise than diaphyseal length models, displaying a slight overestimation of age in the younger ages and

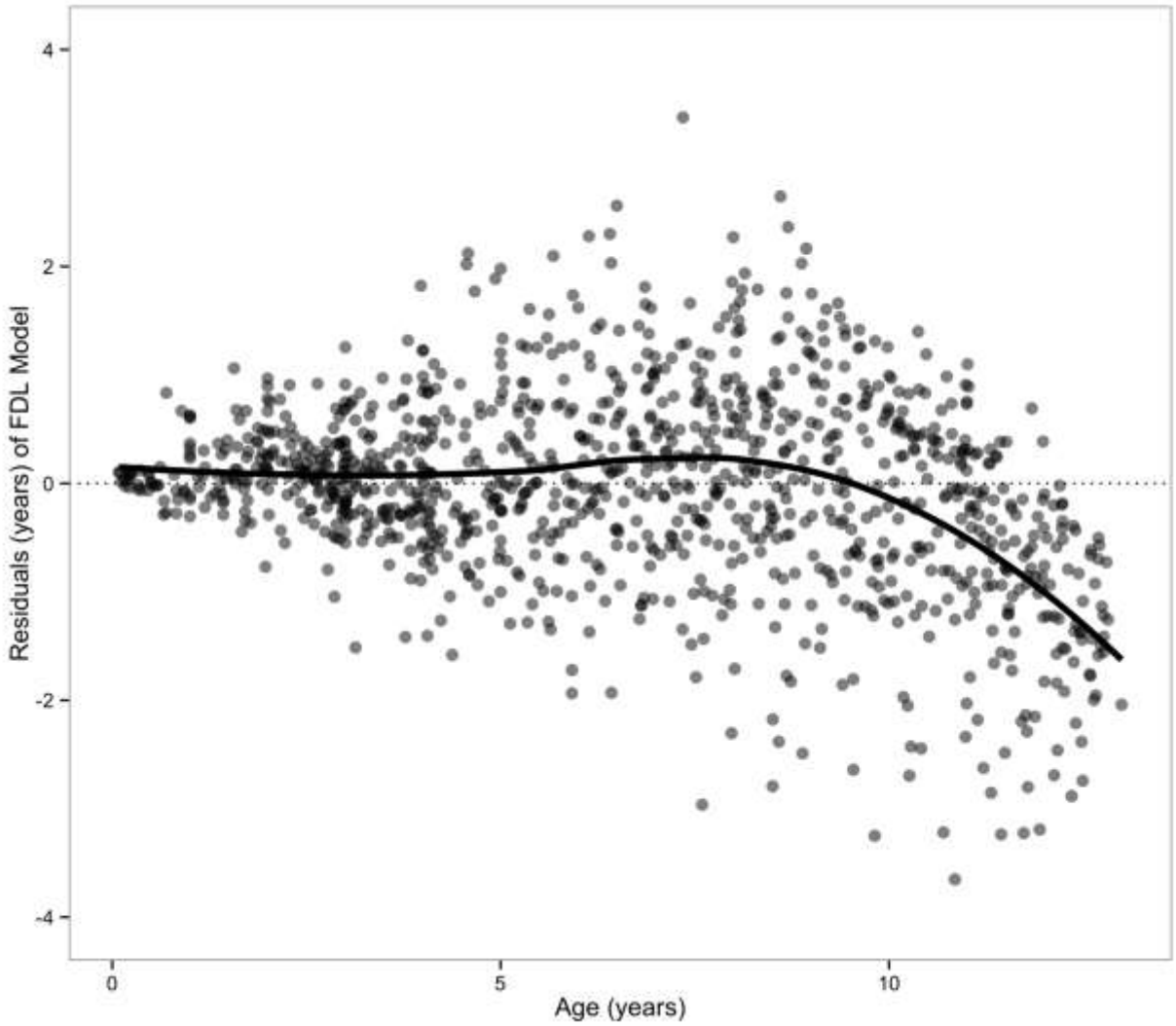

Fig. 3. The loess line depicts the loss of precision commencing around 10 years of age. All diaphyseal length models demonstrated the same trend when the residuals of the model were plotted with chronological age.

an underestimation of age in the older ages (Fig. 4). Even with the loss of precision in the older ages, the frequency of the chronological age falling within the PIs ranged from 94% to 100%.

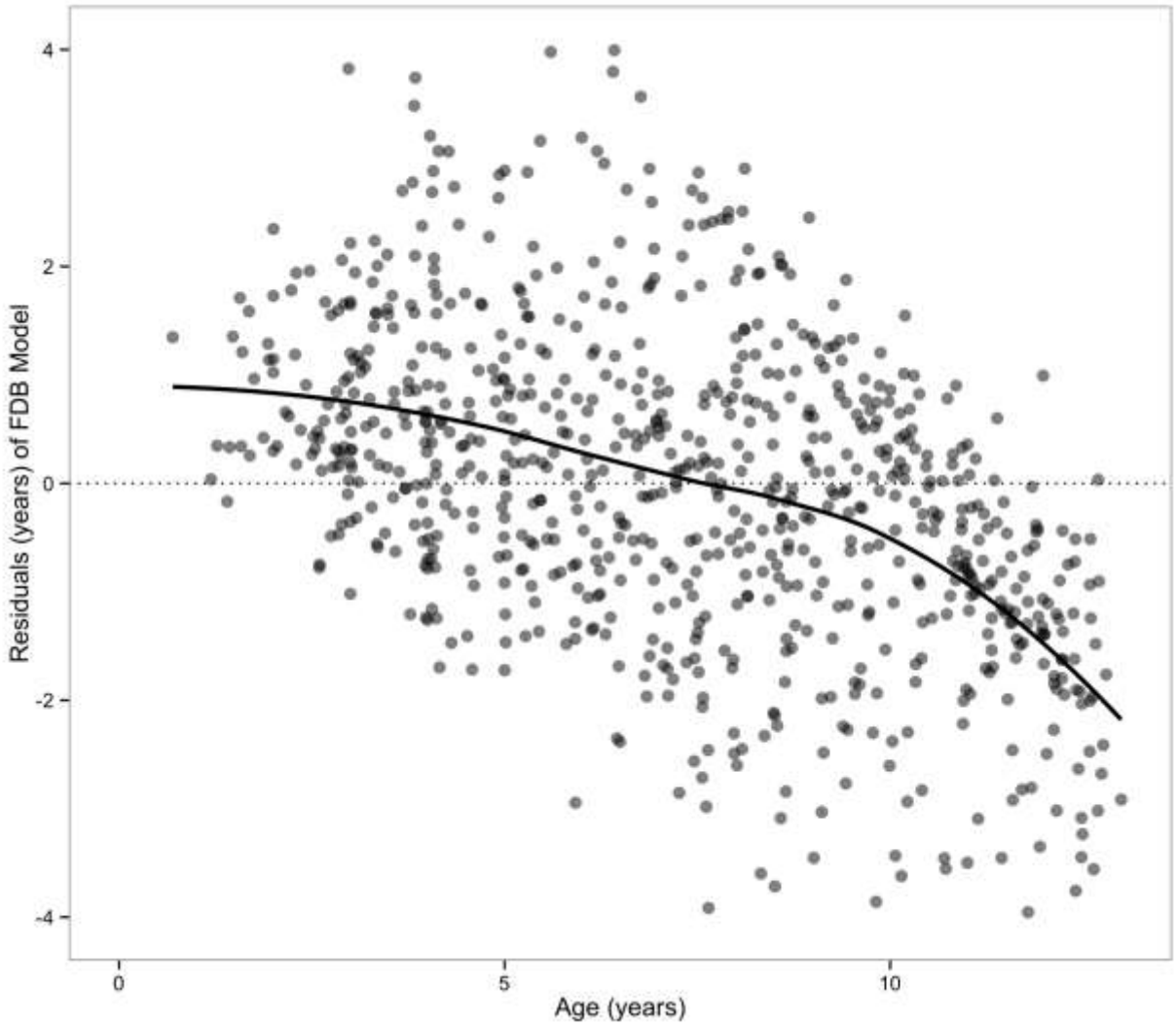

Fig. 4. The loess line depicts the trend to overestimate age in the younger ages and to underestimate age in the older ages. All diaphyseal breadth models demonstrated the same trend when the residuals of the model were plotted with chronological age.

Multivariate models

Model types, variables utilized in model creation, SE and the k-fold cross-validated 95% PIs for each multivariate model are shown in Table 7. Though the number of variables and their importance varied for each model, either the diaphyseal lengths or PC1 were retained and recognized as the most influential predictor variable(s) in every model. In the age models using the bone subsets, the diaphyseal length or PC1 were the only variable used for model creation for

Table 7 – The transformation required on age, stepwise selected variables, standard error (SE), and k-fold cross-validated 95% prediction interval (PI) for each multivariate model.

	Model	Variables	SE	Size of 95% PI
Radius	sqrt MARS	RDL, RMSB	0.98	4 months – 5 years
	cbrt MARS	PC1, PC2, PC4	0.97	3 months – 6 years
Tibia	cbrt MARS	TDL, TPB	0.98	1 year – 5 years
	cbrt MARS	PC1, PC2	0.97	1 year – 5 years
Upper Limbs	cbrt MARS	UDL, RMSB, RDB, HMSB	0.99	6 months – 7 years
	cbrt MARS	PC1, PC9	1.04	6 months – 6 years
Lower Limbs	cbrt MARS	FDL, TDB	0.86	1 year – 5 years
	cbrt MARS	PC1, PC2, PC5, PC6	0.85	1 year – 5 years
Proximal Elements	cbrt MARS	FDL, HDB, HPB, HDL	0.87	1 year – 5 years
	cbrt MARS	PC1, PC3	0.88	1 year – 6 years
Distal Elements	cbrt MARS	RMSB, TPB, TDB, FBDL, UDL, RDB, UMSB, TDL	0.78	1 year – 6 years
	cbrt MARS	PC1, PC4, PC6, PC7, PC9, PC10	0.79	1 year – 7 years
All measurement	cbrt MARS	FDL, HDB, TPB, RDL, UDL, TDL	0.80	1 year – 5 years
	MARS	PC1, PC2, PC3, PC5, PC8, PC10	0.77	1 year – 6 years

the humerus, ulna, and femur. The cross-validated 95% PIs for the majority of the multivariate models ranged between one year and five years.

Stepwise selected principal components generally had large contributions from the same measurements that were stepwise selected in the model using the original measurements. For example, six of the 14 measurements were stepwise selected in the all-measurement model (FDL, HDB, TPB, RDL, TDL, and UDL). Similarly, MARS stepwise selected six of the 14 predictors in the all-measurement model using PC scores. The stepwise selected variables, in the order of variable importance, were PC1, PC3, PC2, PC10, PC5 and PC8. The largest contributions to PC1-3 and PC5 were from diaphyseal lengths (FDL, TDL, UDL, HDL) whereas PC8 and PC10 had large contributions from diaphyseal breadths (TDB, HPB, FDB and TMSB) (Table 8).

The bias in the multivariate models ranged between -2 and +2 years. The mean bias of the multivariate models always approximated zero, though some models demonstrated a slight loss of precision as age increased (Fig. 5). Heteroscedasticity was apparent though not to the severity of the univariate models. Even though the models lose precision and may be slightly biased, 100% of the predicted ages fell within the 95% PIs.

Table 8 – The eigenvectors and proportion of variance for PCs 1 through 10 using the variance-covariance matrix for the all measurement model.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**
TDL	-0.460	-0.158	0.517	-0.110	-0.231	0.052	-0.614	-0.192	0.063	-0.028
TPB	-0.067	-0.087	-0.210	-0.457	-0.033	-0.117	-0.043	0.236	0.215	0.142
TMSB	-0.025	-0.023	-0.058	-0.157	-0.001	0.036	0.076	-0.003	-0.108	0.479
TDB	-0.051	-0.028	-0.097	-0.302	-0.057	0.035	-0.094	-0.042	-0.293	0.543
UDL	-0.274	-0.472	-0.293	0.187	0.314	0.104	-0.202	0.343	-0.505	-0.218
FDL	-0.540	0.700	-0.129	-0.062	0.439	0.028	-0.027	-0.030	-0.032	-0.028
FDB	-0.072	-0.091	-0.255	-0.561	-0.076	-0.281	-0.069	0.212	0.227	-0.353
FMSB	-0.025	-0.029	-0.045	-0.115	-0.044	0.054	-0.036	0.042	-0.190	0.274
RDL	-0.262	-0.412	-0.277	0.216	0.319	-0.194	0.056	-0.442	0.480	0.235
HMSB	-0.016	-0.058	-0.081	-0.114	-0.040	0.097	0.109	-0.256	-0.244	0.075
HDB	-0.047	-0.094	-0.141	-0.209	-0.021	0.901	0.114	-0.097	0.240	-0.139
HDL	-0.354	0.136	-0.447	0.356	-0.718	-0.007	0.033	0.095	0.044	0.042
HPB	-0.042	-0.027	-0.140	-0.247	-0.146	-0.157	0.182	-0.645	-0.398	-0.353
FBDL	-0.458	-0.206	0.430	-0.076	-0.058	-0.059	0.705	0.214	-0.031	0.002
Proportion of Variance	0.985	0.004	0.003	0.003	0.002	0.0005	0.0004	0.0004	0.0003	0.0002

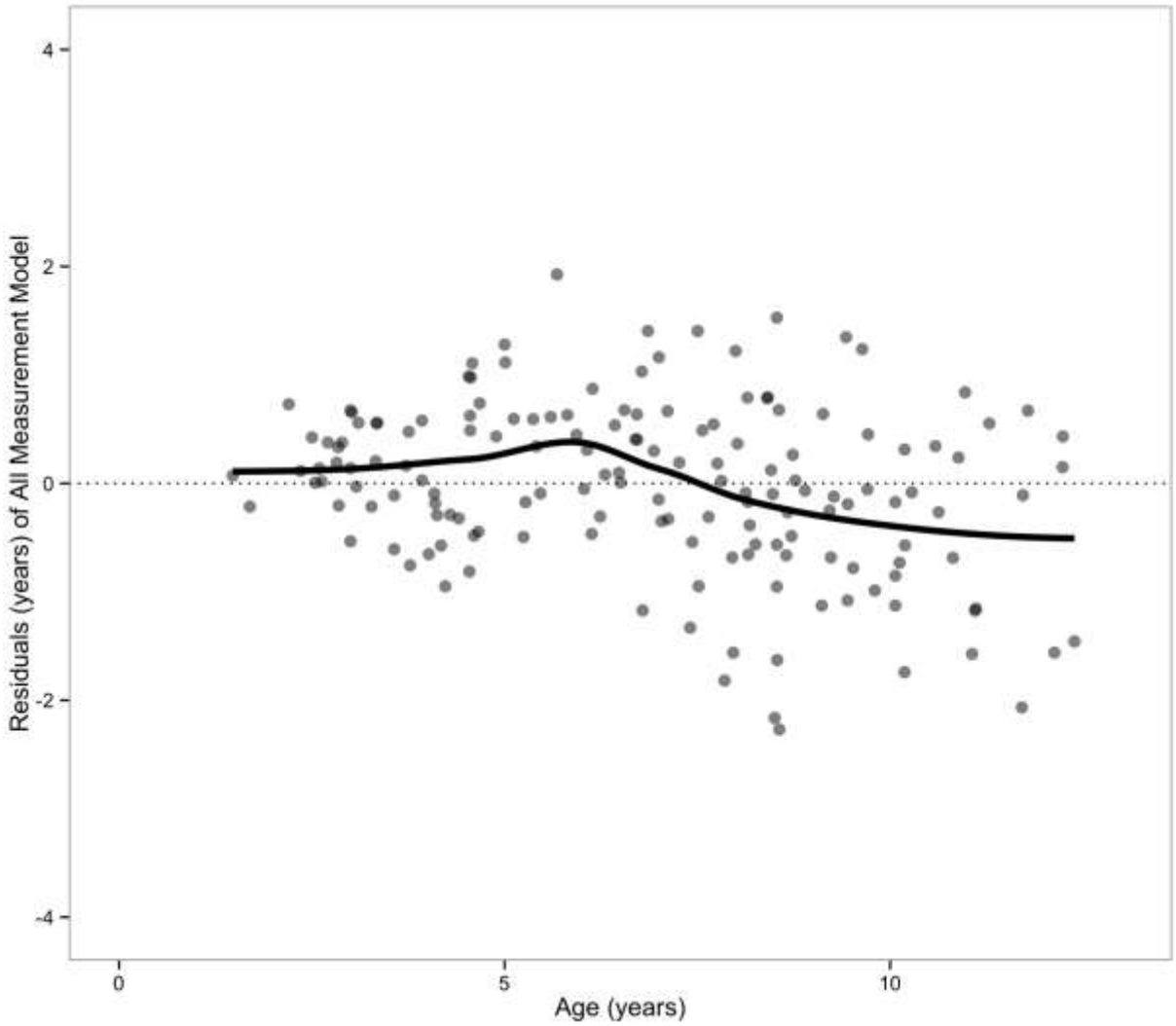

Fig. 5. The loess line depicts the high precision, and the spread of the residuals shows the small range of error for the all measurement model.

DISCUSSION

This study addressed the problems associated with subadult age estimation by integrating a large modern sample, a univariate and multivariate approach, and appropriate statistics that model nonlinear relationships and produce explicit PIs. The sample is the largest modern cross-sectional sample of subadult diaphyseal dimensions from one population ever to be collected. Data are representative of the current South African demographics and reflects modern variation in diaphyseal dimensions. This study also presented the largest number of univariate models as well as the first multivariate approach to age estimation using diaphyseal dimensions. K-fold

cross-validated 95% PIs were created, which provide an appropriate technique to estimate age in subadults. Thus, the presented method satisfies *Daubert* criteria as each model has a prediction interval with an explicit error rate that accounts for the variation in age.

The largest measurement errors in the current study were associated with the smallest measurements (UMSB and RMSB); however, the differences in repeated measurements for UMSB and RMSB were never greater than 1 mm. Comparisons of measurement error obtained from subadult skeletal data are limited; consequently, the only available comparative studies use the diaphyseal lengths. Overall, the current study had lower %TEM values than Cardoso et al. (2013). An evaluation of published literature demonstrates the intra- and inter-observer TEM and %TEM obtained in the current study are comparable to other studies, even though the specifics for each study vary (i.e. anthropometrics, craniometrics) (Utermohle et al., 1983; Ulijaszek and Kerr, 1999; Cardoso, 2005; WHO Multicentre Growth Reference Study Group, 2006; Sicotte et al., 2010).

Prediction intervals and model selection

The PIs produced from the MARS models compensate for the smaller range of variation in diaphyseal dimensions of the younger children and the larger variation in diaphyseal dimensions of the older children. For all measurements and multivariate subsets, the size of the PIs increased with age, especially during early adolescence. Although all children increase in size, the timing of the adolescent growth spurt is highly variable in boys and girls and among populations (Hauspie and Roelants, 2012). For example, the 95% PI for the smallest FDL (75 mm) is approximately a three-month interval, from 0.04 years to 0.33 years, and the 95% PI when FDL is 348 mm is approximately five years, from 8.5 years to 13.4 years. The lower bounds were adjusted in all models to only include ages after birth, which was the age range of

the collected sample. When the upper PI bound is greater than 12.99 years, age estimates should be augmented with additional age indicators, such as epiphyseal fusion or dental development.

Evaluation of the k-fold cross-validated 95% PIs revealed the multivariate models had smaller PIs for the older children than most of the univariate models and univariate models had smaller PIs in younger children than most of the multivariate models. While some of the univariate breadth models have 95% PIs as wide as ten years, the multivariate subsets offer 95% PIs of five or six years for comparable ages. To demonstrate the differences between multivariate and univariate models, the sample was separated into age intervals that follow the subdivisions of growth, such as less than 3 years, between 3 and 6.99 years, and between 7 and 12.99 years (Bogin, 1999). Pearson correlation coefficients were obtained for each diaphyseal length and age subdivision (Table 9). High correlations were noted for the diaphyseal lengths with age in the youngest age interval, showing multicollinearity and indicating that including multiple diaphyseal lengths will not improve predictions. The correlation coefficients decrease with age. A multitude of interacting factors affect growth and ultimately result in adults of different sizes and proportions. Older children are more variable in their proportions and the inclusion of more variables reduces the PI.

A breadths-only model was created because diaphyseal lengths were always retained in the multivariate models and the authors wanted to purposefully remove the variables that presented with the strongest relationships with age. Furthermore, a breadths-only model may be applicable when remains are damaged. The size of the k-fold cross-validated 95% PIs ranged between 2 and 7 years, which were narrower than all of the univariate breadth models and only slightly wider than the univariate diaphyseal length models and multivariate models at the oldest

Table 9 – Pearson correlation coefficients of the diaphyseal lengths per three growth periods. The <3 interval included individuals from birth to 2.99, the second interval included individuals from 3.00 to 6.99 years and the oldest interval included individuals from 7.00 to 12.99 years.

	Age (years)		
	<3 (n =128)	*3 – 6.99 (n =265)*	*7 – 12 (n =300)*
HDL	0.873	0.849	0.675
RDL	0.859	0.831	0.669
UDL	0.849	0.829	0.666
FDL	0.908	0.876	0.725
TDL	0.887	0.844	0.684
FBDL	0.892	0.856	0.679

ages. Thus, if diaphyseal lengths are not available, the application of a multivariate breadth model is preferred to a univariate breadth model.

Precision and accuracy

Bias results of all models indicate MARS loses precision in children older than 10 years of age. Though the increased imprecision with increased age persisted in the multivariate models, the bias was smaller than in the univariate models. However, even though the models lose precision, 95% – or more – of the observed values fell within the 95% PIs, indicating high accuracy. Because of the trend to underestimate age as age increased, the bias was plotted by sex to explore what other factors may be affecting the precision of the models. Based on the loess lines, males and females followed similar trends in bias for all univariate models. However, larger sex disparities were apparent in the univariate breadth models than in the univariate length models and the multivariate models (Fig. 6). The sex discrepancy in the bias of the diaphyseal breadth models is likely due to the greater number of significant differences between males and females in breadth than in length measurements (Stull, 2013). Sex differences in the multivariate models were less than univariate breadth models, which further supports the suggestion to apply multivariate breadth models rather than univariate breadth models.

Lack of sex differences in the bias of univariate diaphyseal length models contradicted the expectations of lost precision due to females entering the adolescent growth spurt earlier than males. The results indicate other biological factors are responsible for the increased bias with increased age. A recent study demonstrated statistically significant differences in height for the three largest South African populations (black, white and coloured) in a large and evenly distributed sample of children between six and 10 years of age (Anholts, 2013). Because ancestry is unevenly represented in the current sample, a Student's t-test would likely not reflect accurate

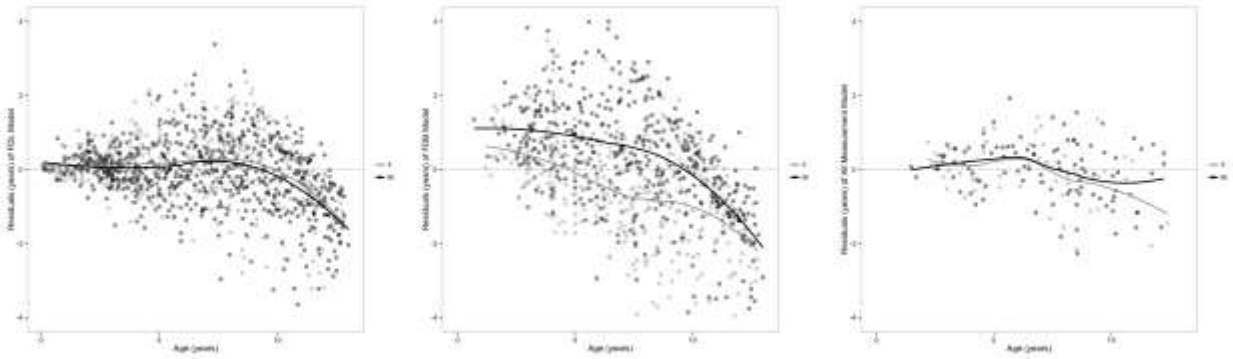

Fig. 6. The loess lines depict the local estimated bias of males and females for a univariate length (left) and breadth (middle) model and the multivariate all measurement model (right). The figures show the lack of sex differences in the univariate length and multivariate model and the pronounced sex differences for the univariate breadth model.

patterns in the population. Future research should assess population differences in diaphyseal dimensions.

Because current techniques are not available to estimate sex or ancestry, the presented models compensated for all the variation in the sample, which resulted in a larger distribution of estimated ages. If variation were eliminated, such that sex and/or ancestry was known, the 95% PIs would be narrower and the precision would increase. For example, a model based on femoral diaphyseal lengths of black South Africans resulted in a bias of -3 years and +2 years and the 95% PIs ranged from a few months to 4 years. Though the ancestry-specific model still lost precision with increased age, the degree was less and the 95% PIs were narrower than the original model, which pooled all populations. Not all variation can be removed because it exists on an individual level and as such, the 95% PIs will reflect the range of variation (Konigsberg and Holman, 1999).

CONCLUSIONS

Anthropologists have previously used diaphyseal dimensions to estimate subadult age without an appreciation of error in the estimates. Two large sources of error can be attributed to

misapplication and inappropriate samples. The current study showed that diaphyseal dimensions can be used to estimate subadult age using univariate and multivariate models with 95% PIs. The ability to create the largest modern subadult sample was feasible through the application of minimally distorted Lodox Statscan radiographic images. The populations that comprised the sample are reflective of the South African population, which allowed for a country-specific technique with immediate applicability.

Multivariate models result in the smallest PIs in older children while univariate models result in the smallest PIs in younger children. Because a specific age cannot be provided for when the univariate or multivariate model is preferred, the anthropologist should choose the model with the narrowest PIs. Furthermore, a multivariate breadth model should be used if diaphyseal lengths are unavailable. The multivariate subsets serve as an example of the potential success of multivariate models when estimating age, however, there are a vast number of tables required to present all combinations of variables. Thus, a computer software program is currently being developed from the South African subadult data that will allow for all possible measurement permutations and k-fold cross-validated 95% PIs. Consequently, the user will be able to provide age estimations specific to each analysis of subadult remains aged between birth and 12 years. In an effort to increase applicability, a sample of modern North American children is currently being collected.

The study documented the first application of MARS in anthropology and highlighted its valuable qualities, especially the flexibility to model the nonlinear relationship between chronological age and diaphyseal dimensions and to generate forward and backward stepwise selected, k-fold cross-validated models. Although MARS performed well through the majority of

the ages, the models lost precision after 10 years of age, reflecting greater variability at those ages.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Trevor Hastie and Stephen Milborrow for invaluable statistical advice regarding MARS and the creation of prediction intervals. The project would have never been completed without the permission to use the images, which was granted by Dr. Thomas Blake at Red Cross War Memorial Children's Hospital and Dr. Lorna Martin at the Forensic Pathology Service, Salt River. We are also grateful to Stef Steiner for his assistance with image acquisition. The medical records staff at Red Cross, and in particular Leah Muller, was of great help. Additional thanks to Michael Kenyhercz for the editorial comments, Jolandie Myburgh and Glenn Eanes for the assistance with data collection, and two anonymous reviewers whose comments strengthened the manuscript.

LITERATURE CITED

Anderson M, Messner M, Green W. 1964. Distribution Lengths of the Normal Femur and Tibia in Children from One to Eighteen Years of Age. J Bone Jt Surg 46A:1197–1202.

Anholts A. 2013. Secular trends in the height and weight of South African children aged 6 to 10 years. BSc Hons Thesis: University of Pretoria, South Africa.

Bland M, Altman D. 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet:307–310.

Bock R, du Toit S. 2004. Parameter estimation in the context of non-linear longitudinal growth models. In: Hauspie R, Cameron N, Molinari L, editors. Methods in Human Growth Research. Cambridge University Press. p 234–257.

Bogin B. 1999. Patterns of Human Growth. Cambridge: Cambridge University Press.

Boldsen J, Milner G, Konigsberg L, Wood J. 2002. Transition Analysis: a new method for estimating age from skeletons. In: Hoppa R, Vaupel J, editors. Paleodemography: age distributions from skeletal samples. Cambridge: Cambrige University Press. p 73–106.

Buckberry J, Chamberlain A. 2002. Age estimation from the Auricular Surface of the Ilium: A revised method. Am J Phys Anthropol 119:231–239.

Butte NF, Wong WW, Adolph AL, Puyau MR, Vohra FA, Zakeri IF. 2010. Validation of Cross-Sectional Time Series and Multivariate Adaptive Regression Splines Models for the Prediction of Energy Expenditure in Children and Adolescents Using Doubly Labeled Water. J Nutr 140:1516–1523.

Cardoso H. 2005. Patterns of Growth and Development of the Human Skeleton and Dentition in Relation to Environmental Quality: A Biocultural Analysis of a Sample of 20th Century Portuguese Subadult Documented Skeletons. PhD Dissertation, McMaster University.

Cardoso HFV, Abrantes J, Humphrey LT. 2013. Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period. Int J Legal Med. doi: 10.1007/s00414-013-0925-5.

Cardoso HFV. 2007. Environmental Effects on Skeletal Versus Dental Developement: Using a Documented Subadult Skeletal Sample to Test a Basic Assumption in Human Osteological Research. Am J Phys Anthropol 132:223–233.

Christensen AM, Crowder CM. 2009. Evidentiary Standards for Forensic Anthropology. Journal of Forensic Sciences 54:1211–1216.

Christopher A. 2002. "To Define the Indefinable": Population Classification and the Census in South Africa. Area 34:401–408.

Cunningham S, Kirkland S, Ross A. 2011. Bone Weathering of Juvenile-Sized Remains in the North Carolina Piedmont. In: Ross AH, Abel SM, editors. The Juvenile Skeleton in Forensic Abuse Investigations. New York: Springer. p 179–196.

Dirkmaat D, Cabo L, Symes S, Ousley S. 2008. New perspectives in forensic anthropology. Am J Phys Anthropol 137:33–52.

Douglas T, Pitcher R, van As A. 2010. Full-body digital radiographic imaging of the injured child. Continuing Medical Education 28:108–112.

Efron B, Tibshirani R. 1993. An introduction to the bootstrap. New York: Chapman & Hall.

Evangelopoulos D, Deyle S, Zimmermann H, Exadaktylos A. 2009. Personal experience with whole-body, low dosage, digital X-ray scanning (LODOX-Statscan) in trauma. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 17:41–45.

Eveleth P, Tanner J. 1990. Worldwide Variation in human growth. 2nd Edition. Cambridge: Cambridge University Press.

Fazekas S, Kósa F. 1978. Forensic Fetal Osteology. Budapest: Akademiai Kiado.

Franklin D. 2010. Forensic age estimation in human skeletal remains: Current concepts and future directions. Legal Medicine 12:1–7.

Freedman DS, Kettel Khan L, Serdula MK, Srinivasan SR, Berenson GS. 2000. Secular trends in height among children during 2 decades: The bogalusa heart study. Arch Pediatr Adolesc Med 154:155–161.

Friedman J. 1993. Fast MARS. Stanford, CA: Department of Statistics, Stanford University.

Friedman JH. 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19:1–67.

Ghantus M. 1951. Growth of the Shaft of the Human Radius and Ulna during the First Two Years of Life. American Journal of Roentgenology 65:784–786.

Gindhart PS. 1973. Growth standards for the tibia and radius in children aged one month through eighteen years. Am J Phys Anthropol 39:41–48.

Goto R, Mascie-Taylor N. 2007. Precision of Measurement as a Component of Human Variation. J Physio Anthropol 26:253–256.

Green W, Wyatt G, Anderson M. 1946. Orthoroentgenography as a method of measuring the bones of the lower extremities. J Bone Joint Surg Am 28:60–65.

Groenewald P, Bradshaw D, Msemburi W, Neetheling I, Matzopolous R, Naledi T, Daniels J, Dombo M. 2011. Western Cape Mortality Profile 2009. Cape Town: South African Medical Research Council.

Hastie T, Tibshirani R, J Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer-Verlag.

Hauspie R, Roelants M. 2012. Adolescent Growth. In: Cameron N, Bogin B, editors. Human Growth and Development. 2nd ed. New York: Elsevier. p 57–79.

Hawley N, Rousham E, Norris S, Pettifor J, Cameron N. 2009. Secular trends in skeletal maturity in South Africa: 1962-2001. Ann Hum Biol 36:584–594.

Hoffman JM. 1979. Age Estimation from Diaphyseal Lengths: Two Months to Twelve Years. J Forensic Sci 24:461–469.

Knobel G, Alexander G, Bowie G. 2006. Lodox Statscan proves to be invaluable in forensic medicine. South African Medical Journal 96:593–594.

Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. p 1137–1143. Available from: http://dl.acm.org/citation.cfm?id=1643031.1643047

Konigsberg L, Holman D. 1999. Estimation of age at death from dental emergence and implications for studies of prehistoric somatic growth. In: Hoppa R, Fitzgerald C, editors. Human Growth in the Past: Studies from Bones and Teeth. Cambridge: Cambrige University Press. p 264–289.

Lodox Systems Pty (Ltd), Sandton, South Africa. Available from: http://www.lodox.com

Loesch DZ, Stokes K, Huggins RM. 2000. Secular trend in body height and weight of Australian children and adolescents. Am J Phys Anthropol 111:545–556.

López-Costas O, Rissech C, Trancho G, Turbón D. 2012. Postnatal ontogenesis of the tibia. Implications for age and sex estimation. Forensic Sci Int 214:207.e1–11.

Malina R. 2004. Secular trends in growth, maturation, and physical performance: a review. Przeglad Antropologiczny-Anthropological Review 67:3–31.

Maresh M. 1955. Linear growth of long bones of extremities from infancy through adolesence. American Journal of Diseases of Children 89:725–742.

Maresh M. 1970. Measurements from Roentgenograms. In: McCammon R, editor. Human Growth and Development. Springfield, IL: CC. Thomas. p 157–200.

Marques-Vidal P, Madeleine G, Romain S, Gabriel A, Bovet P. 2008. Secular trends in height and weight among children and adolescents of the Seychelles, 1956–2006. BMC Public Health 8:166.

McKern T, Stewart T. 1957. Skeletal age changes in young American males: analysed from the standpoint of age identification. Natick, MA: Headquarters Quartermaster Research and Development Command.

Melnick RL. 2005. A Daubert motion: a legal strategy to exclude essential scientific evidence in toxic tort litigation. Am J Public Health 95 Suppl 1:S30–34.

Meredith H. 1976. Findings from Asia, Australia, Europe, and North America on Secular Change in Mean Height of Children, Youths, and Young Adults. Am J Phys Anthropol 44:315–326.

Milborrow S. 2013. earth: Multivariate Adaptive Regression Spline Models. Available from: http://CRAN.R-project.org/package=earth

Moore-Jansen P, Ousley S, Jantz R. 1994. Data Collection Procedures for Forensic Skeletal Material. Knoxville: Department of Anthropology, The University of Tennessee.

Muñoz J, Felicísimo ÁM. 2004. Comparison of statistical methods commonly used in predictive modelling. J Veg Sci 15:285–292.

Ousley S, Hollinger RE. 2009. A Forensic Analysis of Human Remains from a Historic Conflict in North Dakota. In: Steadman DW, editor. Hard Evidence: Case Studies in Forensic Anthropology. Upper Saddle River: Prentice Hall. p 91–101.

Ousley S. 2013. Final Report: A Radiographic Database for Estimating Biological Parameters in Modern Subadults. Erie, PA: Mercyhurst University.

Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM. 2010. Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19:411–419.

Petersen DC, Libiger O, Tindall EA, Hardie R-A, Hannick LI, Glashoff RH, Mukerji M, Fernandez P, Haacke W, Schork NJ, Hayes VM, Indian Genome Variation Consortium. 2013. Complex Patterns of Genomic Admixture within Southern Africa. PLoS Genet 9:e1003309.

Pokines JT. 2014. Faunal Dispersal, Reconcentration, and Gnawing Damage to Bone in Terrestrial Environments. In: Pokines JT, Symes SA, editors. Manual of Forensic Taphonomy. Boca Raton: CRC Press. p 201–248.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/

Rissech C, López-Costas O, Turbón D. 2013. Humeral development from neonatal period to skeletal maturity—application in age and sex assessment. International Journal of Legal Medicine 127:201–212.

Rissech C, Schaefer M, Malgosa A. 2008. Development of the femur--Implications for age and sex determination. Forensic Sci Int 180:1–9.

Saunders S. 2008. Juvenile skeletons and growth-related studies. In: Katzenberg M, Saunders S, editors. Biological Anthropology of the Human Skeleton. 2nd ed. Hoboken: John WIley & Sons. p 117–147.

Sekulic S, Kowalski BR. 1992. MARS: A tutorial. Journal of Chemometrics 6:199–216.

Sicotte M, Ledoux M, Zunzunegui M-V, Aboubacrine SA, Nguyen V-K. 2010. Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa. BMC Medical Research Methodology 10:102.

Statistics South Africa. 2012. Census 2011: Statistical Release. Pretoria: Statistics South Africa.

Stull K, Frazee K, Cabo L. 2008. Accuracy of Metric Infant Age Estimation Methods. Program of the 77th Annual Meeting of the American Association of Physical Anthropologists. Columbus, OH. pp: 202.

Stull K, L'Abbe E, Ousley S. 2013a. Testing the Accuracy of North American Growth Standards to Estimate Age-at-Death in Modern South African Subadults. Proceedings of the 65th Annual Meeting of the American Academy of Forensic Sciences, Washington, DC. Colorado Springs, CO: American Academy of Forensic Sciences. pp: 434.

Stull K. 2013. An Osteometric Evaluation of Age and Sex Differences in the Long Bones of South African Children from the Western Cape. PhD Dissertation: University of Pretoria, South Africa.

Stull KE, L'abbé EN, Steiner S. 2013b. Measuring distortion of skeletal elements in Lodox Statscan-generated images. Clin Anat 26:780–786.

Stull KE, Tise ML, Ali Z, Fowler DR. 2014. Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. Forensic Science International 238:133–140.

Treiman DJ. 2007. The Legacy of Apartheid: Racial Inequalities in the New South Africa. In: Heath AF, Cheung SY, editors. Unequal chances: ethnic minorities in Western labour markets. Oxford: Published for the British Academy by Oxford University Press. p 401–447.

Ubelaker DH. 1987. Estimating Age at Death from Immature Human Skeletons: An Overview. Journal of Forensic Sciences 32:1254–1263.

Uhl N, Nawrocki S. 2010. Multifactorial Estimation of Age at Death from the Human Skeleton. In: Krista E. Latham, Finnegan M, editors. Age Estimation of the Human Skeleton. Springfield: Charles C. Thomas. p 243–261.

Ulijaszek S, Kerr D. 1999. Anthropometric Measurement Error and the Assessment of Nutritional Status. Br J Nutr 82:165–177.

Utermohle CJ, Zegura SL, Heathcote GM. 1983. Multiple observers, humidity, and choice of precision statistics: factors influencing craniometric data quality. Am J Phys Anthropol 61:85–95.

De Veaux RDD, Ungar LH. 1994. Multicollinearity: A tale of two nonparametric regressions. In: Oldford RW, editor. Selecting Models from Data. Lecture Notes in Statistics. Springer New York. p 393–402. Available from: http://link.springer.com/chapter/10.1007/978-1-4612-2660-4_40

WHO Multicentre Growth Reference Study Group. 2006. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 450:38–46.