



Article

Age-at-Death Estimation: Accuracy and Reliability of Common Age-Reporting Strategies in Forensic Anthropology

Christine Bailey ^{1,*} and Giovanna Vidoli ² ¹ Department of Physical Anthropology, Cleveland Museum of Natural History, Cleveland, OH 44106, USA² Department of Anthropology, University of Tennessee, Knoxville, TN 37996, USA

* Correspondence: cbailey@cmnh.org

Abstract: Forensic anthropologists build a biological profile—consisting of sex, age, population affinity, and stature estimates—to assist medicolegal stakeholders in the identification of unknown human skeletal remains. While adult age-at-death estimations can narrow the pool of potential individuals, a lack of standards, best practices, and consensus among anthropologists for method selection and the production of a final age estimate present significant challenges. The purpose of this research is to identify age-reporting strategies that provide the most accurate and reliable (i.e., low inaccuracy and low bias) adult age-at-death estimates when evaluated considering the total sample, age cohort (20–39; 40–59; 60–79), and sex. Age-reporting strategies in this study were derived from six age-at-death estimation methods and tested on 58 adult individuals (31 males, 27 females) from the UTK Donated Skeletal Collection. An experienced-based estimation strategy was also assessed. A paired-samples t-test was used to determine whether there was a significant difference ($p \leq 0.05$) between the mean estimated age and the actual age for all age-reporting strategies. Results show that the most accurate and reliable age-reporting strategy varied if the sample was evaluated as a whole, by age, or by sex. While none of the age-reporting strategies evaluated in this study were consistently the most accurate and reliable for all of the sample categories, the experience-based approach performed well for each group.

Keywords: forensic anthropology; age-at-death estimation; biological profile; standards; best practices; error

**Citation:** Bailey, C.; Vidoli, G.Age-at-Death Estimation: Accuracy and Reliability of Common Age-Reporting Strategies in Forensic Anthropology. *Forensic Sci.* **2023**, *3*, 179–191. <https://doi.org/10.3390/forensicsci3010014>

Academic Editors: Kanya Godde and Rebecca Taylor

Received: 31 January 2023

Revised: 7 March 2023

Accepted: 14 March 2023

Published: 20 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multiple adult age-estimation methods performed in forensic anthropology meet the admissibility standards set by the *Daubert v. Merrell Dow Pharmaceuticals* ruling [1], but most of these methods produce age ranges so large that they offer little probative value. Alternatively, determining a final age estimate (the range reported to medicolegal stakeholders) based on expert judgment can produce a narrow age range, but it often lacks reproducibility. Therefore, determining a final estimate that both narrows down the potential victim pool and meets the scientific standards poses a significant challenge to forensic anthropologists.

To better understand current practices in skeletal age estimation in a forensic context, Garvin and Passalacqua [2] administered a survey to 145 members of the Anthropology section of the American Academy of Forensic Science (AAFS). Their study assessed how anthropologists decide which skeletal region to evaluate, which method(s) to use, how to report statistical information, and how information from different methods are translated into a final age estimate. They found that practitioners prefer the pubic symphysis, fourth rib, and auricular surface methods when estimating age, and tend to rely on methods developed in the 1980s and 1990s. When reporting age using a single method, survey participants preferred to report the full age range provided by the method. When using information from multiple methods, respondents did not agree on the best way to combine

information, but using one's experience/familiarity with human skeletal variation was a deciding factor in producing a final age estimate. This survey demonstrates the lack of standardization in reporting adult age, especially when combining information from multiple methods.

Many studies have assessed the accuracy and reliability of commonly used skeletal aging methods [2–6], and some have evaluated ways to combine information from multiple methods and skeletal indicators [2–4,7–9] in order to reach a final age estimate to report to stakeholders. While these studies suggest several age-reporting strategies, these strategies have not been systematically tested for their accuracy, reliability, and practicality in a forensic setting. There are differing viewpoints regarding which age-reporting strategies are most appropriate to be used in order to arrive at a final age estimate. When relying on a single aging method to produce a final age estimate, one could simply report the age range provided by the method. However, these ranges can be extremely broad, encompassing over 30 years, rendering them ineffective in narrowing down potential matches. If multiple methods or indicators are relied upon for estimating age, for example, pubic symphyses and ribs, it is often difficult to decide which method or ranges would be best to report. Nawrocki [10] suggests that it is statistically appropriate to report age using the indicator with the lowest standard error when using multiple methods. Alternatively, several researchers [3,4,7,8] suggest using a two-step strategy in which the most appropriate method is determined based on a preliminary assessment of skeletal morphology that is associated with age-related skeletal changes. Finally, some anthropologists favor comprehensive strategies, such as the overlap of multiple age ranges or using professional judgment to combine information from multiple methods/indicators [2,6,8,9]. This study assesses various age-reporting strategies for their ability to produce accurate and reliable final age estimates. The accuracy and reliability of each strategy included in this study were assessed by the total sample, age, and sex. While the Organization of Scientific Area Committees for Forensic Science (OSAC) provides recommendations for reporting skeletal age, there are no current standards for reporting a final age estimate in forensic anthropology.

The goal of this study is to examine which age-reporting strategies provide the most appropriate final age estimate to include in a forensic anthropology case report. This project assessed whether (1) the two-step strategy is the most accurate and reliable when the sample is divided by age, (2) methods that are sex-specific provide more accurate and reliable age estimations than non-sex specific methods when the sample is separated by sex, and (3) experience-based final age estimates are accurate and reliable.

2. Materials and Methods

2.1. Sample

In this study, 58 individuals (31 males and 27 females) from the UTK Donated Skeletal Collection, were selected. The ages-at-death ranged from 21–79 years old, with roughly 10 individuals representing each decade of life (Table 1).

Table 1. Sample Demographics.

Age Group	Females	Males	Total
20–30	2	6	8
31–39	5	5	10
40–49	5	6	10
50–59	5	6	10
60–69	5	6	10
70–79	5	6	10
Total	27	31	58

2.2. Age-Estimation Methods

Six adult age-estimation methods (three original methods from the 1980s/1990s and three revised methods) were independently applied following the publication descriptions. The original methods included in this study were Suchey–Brooks [11], pubic symphysis (SBPS); Lovejoy [12] auricular surface (LJAS); and İşcan [13–15] fourth rib (ISR). The revised methods were Hartnett [16] pubic symphysis (HNPS); Buckberry–Chamberlain [17] auricular surface (BCAS) and Hartnett [18] fourth rib (HNR). A summary of each method and their age composition is shown in Table 2. To reduce the potential for bias, one method at a time was applied to all of the individuals in the sample before moving on to the next method, and chronological age was hidden from the observer until all data were collected. The primary author, whose experience at the time was limited— including only an undergraduate and graduate course in forensic anthropology— collected all the data.

Table 2. Overview of Methods.

Skeletal Indicator	Method	% of Sample Over Age 60 *	Phase/Component
Pubic Symphysis	Suchey–Brooks [11]	15%	Phase
Pubic Symphysis	Hartnett [16]	40%	Phase
Auricular Surface	Lovejoy [12]	13%	Phase
Auricular Surface	Buckberry–Chamberlain [17]	48%	Component
Sternal Rib End	İşcan [13,14]	19%	Phase
Sternal Rib End	Hartnett [18]	40%	Phase

* Percentages are from Merritt 2014 [4].

2.3. Age-Reporting Strategies

Sixteen final age estimates, such as those that would be reported in a forensic report, were produced for each of the 58 individuals using different age-reporting strategies (Figure 1) derived from the six methods used in this study; this included method range, lowest error, two-step, overlap, and experience. First, an age estimate based on the phase/score from each method was recorded, resulting in six estimates, one for each of the methods in this study. Previous survey data indicated that this strategy was preferred by anthropologists when using the results of a single method for age estimation [2].

Next, the method range with the lowest standard error when considering all methods was recorded. Then, the range with the lowest error by skeletal region was identified for each individual. These four age ranges reflect Nawrocki’s advice, which was to report the method with the lowest range of error [10]. Of the six methods used in this study, only İşcan’s [13–15] rib method provided the standard error value for each phase described in the method. Therefore, the standard error was calculated for the remaining methods using the following formula:

$$SE = \frac{S}{\sqrt{n}}$$

where SE = standard error, S = standard deviation, and n = sample size [19]. The standard error calculations are found in Table 3. Lovejoy [13,14] did not provide enough information to calculate the standard error and was not an option for this portion of the study.

Another strategy for estimating age is a “two-step” approach, in which a preliminary assessment of the skeletal indicator (pubic symphysis, sternal rib end, auricular surface) was made to decide which method’s range would be recorded as the final age estimate [3,4,7,8]. Each indicator was first assessed using the original method, and if it was assigned to a lower phase of the method (Suchey–Brooks I–III, İşcan 0–5, and Lovejoy 1–4), then that method’s results were recorded; however, if the individual was assigned to a higher phase of the original method, then the associated revised method (Hartnett pubic symphysis, Hartnett rib, or Buckberry–Chamberlain) was applied and its results were recorded. Using

the two-step strategy, three final age estimates were recorded for each individual—one for each indicator assessed in this study.

Table 3. Standard Error for Each Method.

Phase	SBPS Males	SBPS Females	HNPS Males	HNPS Female	LJAS	BCAS
1	0.02	0.38	0.52	0.59	-	0.88
2	0.42	0.71	0.50	1.06	-	2.74
3	0.91	1.22	1.11	1.02	-	2.79
4	0.72	1.75	1.01	1.03	-	2.56
5	0.90	2.20	0.89	0.70	-	1.62
6	0.87	1.74	1.38	1.24	-	1.86
7	-	-	0.95	0.99	-	3.67
8	-	-	-	-	-	-

Phase	ISR Males	ISR Females	HNR Males	HNR Females
1	0.25	-	0.32	0.63
2	0.59	0.68	0.38	0.44
3	0.85	0.74	0.71	0.68
4	1.11	1.46	0.43	0.67
5	1.93	2.96	0.40	0.59
6	2.71	3.52	0.45	0.80
7	2.31	2.81	0.76	0.82
8	2.97	2.66	-	-

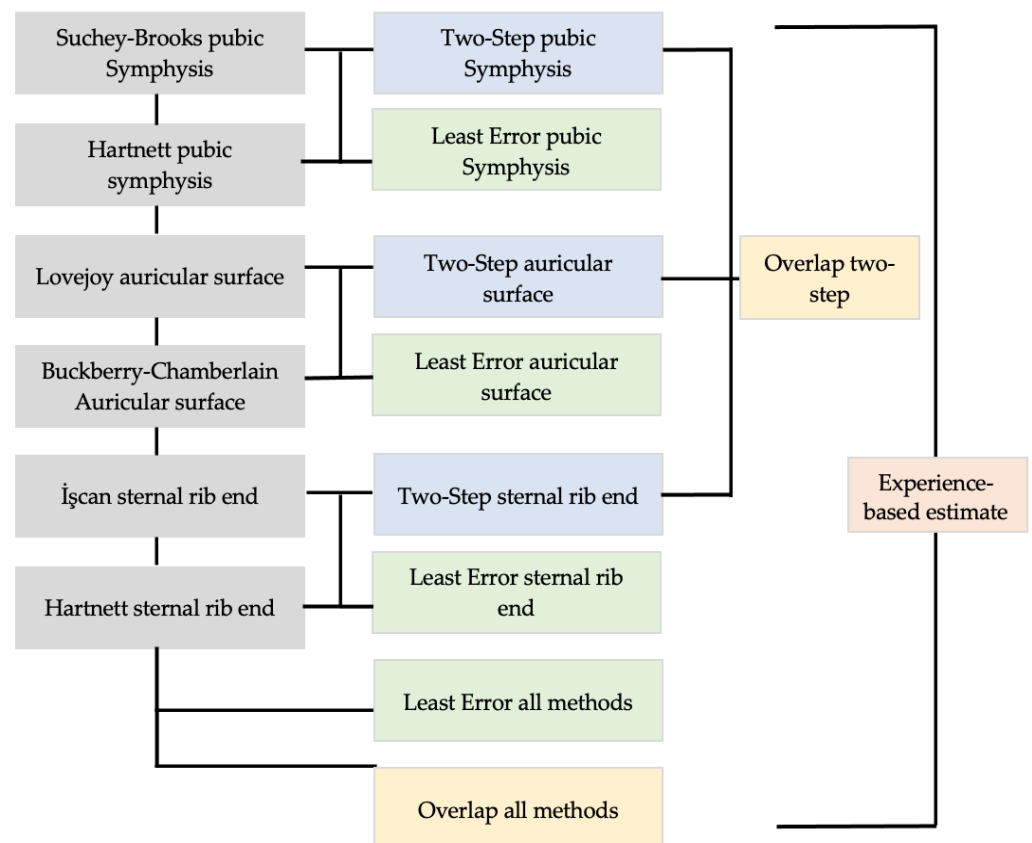


Figure 1. Age-reporting strategies.

Finally, forensic anthropologists often choose to include data from multiple methods/skeletal indicators in their final age estimates. Therefore, two comprehensive strategies, overlap and experience-based, were included. For the overlap strategy, the researcher chose a range based on the overlap of the six ranges derived from the aging methods used in

this study. Additionally, an estimate was produced using the overlap of the three methods identified through the two-step strategy. For the experience-based estimate, the researcher had the ability to delineate any range they felt appropriate given the results of the aging methods combined with the overall appearance/condition of the skeletal remains.

2.4. Statistical Methodologies

A paired-samples *t*-test was conducted in SPSS [20] and used to determine whether there was a significant difference ($p \leq 0.05$) between the mean estimated age and the actual age for all the age-reporting strategies, including ages derived from the six aging methods. Accuracy and reliability were also assessed for each strategy. These metrics were assessed for the whole sample, age cohort, and sex. Age cohorts were expanded from six, approximately ten-year, ranges to three, approximately twenty-year, ranges. The three age cohorts represent “younger” (20–39), “middle-age” (40–59), and “older” (60–79) individuals in the sample.

Accuracy is the number of individuals who were correctly assigned to an age range that included their actual age at death. The following equation was used:

$$Accuracy(\%) = \frac{\#Correct}{\#Total}$$

Reliability, as described by Meindl and colleagues [21], is low inaccuracy and minimal bias. Inaccuracy assesses the absolute difference between the estimated and the actual age without consideration for under/overestimation, and is calculated as follows:

$$\frac{[estimatedage - actualage]}{n}$$

where inaccuracy is the sum of the absolute value of the estimated age minus the actual age divided by the number of individuals in the sample.

Alternatively, bias is the mean under/overprediction of the individuals age and is calculated as follows:

$$\frac{(estimatedage - actualage)}{n}$$

where bias is the sum of the estimated age minus the actual age divided by the number of individuals in the sample. If the bias score is positive, then the age-reporting strategy overestimated age. If the bias score is negative, the age-reporting strategy underestimated age.

The thresholds for accuracy and reliability were arbitrarily delineated for this study due to a lack of standards for these measures. The goal of this study was to identify thresholds that would be rigorous enough to distinguish the best performing age-reporting strategies, but not too restrictive that none of the strategies met the standard. Age-reporting strategies were considered accurate if 80% of the individuals in the sample were correctly assigned to a range that included their age at death. Inaccuracy was considered low if the mean difference between the actual and estimated age was 10 years, and bias was considered minimal if the mean difference between the actual and estimated age was between -1 and 1 year. Accuracy, inaccuracy, and bias results were calculated in an Excel® spreadsheet. All percentages were rounded to the nearest whole number.

3. Results

3.1. T-Test Results Comparing Actual Age and Estimated Age

The results from the paired sample *t*-test, comparing the actual age and estimated age, are found in Table 4. There were significant differences in the mean values between the actual age and estimated age for Lovejoy (LJAS) ($t = 4.107$, $df = 57$, $p < 0.05$), Buckberry–Chamberlain (BCAS) ($t = -3.485$, $df = 57$, $p < 0.05$) and Least Error Auricular Surface (LEAS) ($t = -3.485$, $df = 57$, $p < 0.05$). Because only BCAS contributed to LEAS estimates, their results are identical and will be discussed together.

Table 4. Paired Sample *t*-test comparing actual and estimated age. Significant values are bolded.

		Paired Samples Test								
		Paired Differences					t	df	Sig. (2-Tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
					Lower	Upper				
Pair 1	Age—ISR	1.2500	11.6590	1.5309	−1.8156	4.3156	0.817	57	0.418	
Pair 2	Age—HNR	−2.7328	12.4405	1.6335	−6.0038	0.5383	−1.673	57	0.100	
Pair 3	Age—SBPS	2.0000	15.4056	2.0229	−2.0507	6.0507	0.989	57	0.327	
Pair 4	Age—LJAS	7.6810	14.2425	1.8701	3.9362	11.4259	4.107	57	0.000	
Pair 5	Age—BCAS	−6.4397	14.0728	1.8479	−10.1399	−2.7394	−3.485	57	0.001	
Pair 6	Age—LEall	−1.1810	11.1145	1.4594	−4.1034	1.7414	−0.809	57	0.422	
Pair 7	Age—LER	−2.7672	12.8838	1.6917	−6.1549	.6204	−1.636	57	0.107	
Pair 8	Age—LEPS	−0.6552	13.7883	1.8105	−4.2806	2.9703	−0.362	57	0.719	
Pair 9	Age—LEAS	−6.4397	14.0728	1.8479	−10.1399	−2.7394	−3.485	57	0.001	
Pair 10	Age—TSR	−0.4138	12.5164	1.6435	−3.7048	2.8772	−0.252	57	0.802	
Pair 11	Age—TSPS	−1.2586	14.3812	1.8883	−5.0400	2.5227	−0.667	57	0.508	
Pair 12	Age—TSAS	−2.6638	14.9556	1.9638	−6.5962	1.2686	−1.356	57	0.180	
Pair 13	Age—TSOL	−0.0086	10.1914	1.3382	−2.6883	2.6711	−0.006	57	0.995	
Pair 14	Age—overlap	0.1552	10.4748	1.3754	−2.5990	2.9094	0.113	57	0.911	
Pair 15	Age—experience	0.6293	8.2921	1.0888	−1.5510	2.8096	0.578	57	0.566	
Pair 16	Age—HNPS	−2.0690	12.3702	1.6243	−5.3215	1.1836	−1.274	57	0.208	

3.2. Accuracy and Reliability Results

Table 5 shows the accuracy and reliability of the six individual skeletal aging methods and is divided by sample categories (age cohort, sex, and total sample). Table 6 shows the accuracy and reliability of the ten additional age-reporting strategies, including two-step, least-error, overall, and experience-based. These results are also divided by the sample categories. The sixteen age-reporting strategies (the results from the six skeletal aging methods plus the ten derived results in Table 6) are included in the analyses comparing the performance of the different age-reporting strategies. Table 7 highlights the best performing strategy (most accurate and reliable) by the sample category for each criterion (accuracy, mean inaccuracy, and mean bias). The last column of Table 7 provides the strategy, if any, that met all the accuracy and reliability criteria set for this study.

Table 5. Accuracy and reliability (inaccuracy and bias) of individual methods.

Sample Group	Suchey–Brooks Pubic Symphysis	Hartnett Pubic Symphysis	İşcan Ribs	Hartnett Ribs	Lovejoy Auricular Surface	Buckberry–Chamberlain Auricular Surface
20–39 <i>n</i> = 18						
Accuracy	89%	67%	78%	39%	22%	78%
Inaccuracy	8.81	9.53	7.00	8.92	8.89	19.00
Bias	0.47	8.36	2.06	6.53	5.33	16.06

Table 5. Cont.

Sample Group	Suchey–Brooks Pubic Symphysis	Hartnett Pubic Symphysis	İşcan Ribs	Hartnett Ribs	Lovejoy Auricular Surface	Buckberry–Chamberlain Auricular Surface
40–59 <i>n</i> = 20						
Accuracy	95%	80%	70%	35%	25%	85%
Inaccuracy	14.10	11.25	10.08	10.78	7.95	12.07
Bias	7.35	4.60	1.68	2.73	−6.15	9.48
60–79 <i>n</i> = 20						
Accuracy	70%	85%	90%	50%	15%	100%
Inaccuracy	15.58	9.28	10.00	9.18	20.93	7.85
Bias	10.68	−6.13	−7.15	−0.68	−20.93	−5.25
Males <i>n</i> = 31						
Accuracy	74%	81%	77%	42%	29%	84%
Inaccuracy	12.50	10.58	8.10	9.18	12.58	13.18
Bias	4.02	−1.03	−4.97	4.08	−5.54	9.24
Females <i>n</i> = 27						
Accuracy	96%	74%	82%	41%	11%	93%
Inaccuracy	13.50	9.41	10.24	9.44	12.87	12.30
Bias	9.06	5.63	3.02	1.19	−10.13	3.30
Total <i>n</i> = 58						
Accuracy	84%	78%	79%	41%	21%	88%
Inaccuracy	12.97	10.03	9.10	9.65	13.75	12.77
Bias	6.36	2.07	−1.25	2.73	−8.72	6.44

Table 6. Accuracy and reliability (inaccuracy and bias) of age-reporting strategies.

Sample Group	Lowest Error All	Lowest Error Ribs	Lowest Error Pubic Symphysis	Lowest Error Auricular Surface	Two-Step Ribs	Two-Step Pubic Symphysis	Two-Step Auricular Surface	Overlap	Overlap Two-Step	Experience
20–39 <i>n</i> = 18										
Accuracy	44%	39%	83%	78%	67%	78%	56%	83%	72%	78%
Inaccuracy	7.36	9.83	8.44	19.00	6.61	10.19	14.56	5.72	6.64	5.42
Bias	4.92	7.56	8.22	16.06	0.33	9.36	11.00	4.00	4.03	4.58
40–59 <i>n</i> = 20										
Accuracy	35%	30%	80%	85%	30%	75%	60%	65%	85%	85%
Inaccuracy	8.70	10.20	10.68	12.08	10.53	12.20	14.43	8.05	6.28	5.63
Bias	1.80	3.30	3.98	9.48	2.98	3.55	4.48	0.05	1.78	−0.08
60–79 <i>n</i> = 20										
Accuracy	40%	45%	80%	100%	45%	80%	95%	50%	65%	75%
Inaccuracy	10.70	9.98	12.08	7.85	9.98	11.48	8.95	10.45	9.68	7.28
Bias	−2.80	−2.08	−9.48	−5.25	−2.08	−8.33	−6.65	−4.10	−5.38	−5.88
Males <i>n</i> = 31										
Accuracy	42%	42%	87%	88%	48%	77%	61%	61%	71%	71%
Inaccuracy	8.58	10.02	10.82	13.18	9.45	12.68	13.50	8.86	8.42	7.19
Bias	1.84	3.37	−3.47	9.24	0.58	−2.87	4.82	−4.68	−0.61	−0.84

Table 6. Cont.

Sample Group	Lowest Error All	Lowest Error Ribs	Lowest Error Pubic Symphysis	Lowest Error Auricular Surface	Two-Step Ribs	Two-Step Pubic Symphysis	Two-Step Auricular Surface	Overlap	Overlap Two-Step	Experience
Females <i>n</i> = 27										
Accuracy	37%	33%	74%	93%	44%	78%	82%	71%	78%	89%
Inaccuracy	9.43	10.00	10.06	12.30	8.74	9.78	11.52	7.35	6.57	4.91
Bias	0.43	2.07	5.39	3.30	0.22	6.00	0.19	0.20	0.72	−0.39
Total <i>n</i> = 58										
Accuracy	40%	38%	81%	88%	47%	78%	71%	66%	74%	80%
Inaccuracy	8.97	10.01	10.47	12.77	9.12	11.33	12.58	8.16	7.56	6.13
Bias	1.18	2.77	0.66	6.44	0.41	1.26	2.66	−0.16	0.01	−0.63

Table 7. Most accurate and reliable strategies by sample category.

Sample Group	Accuracy	Inaccuracy	Bias	Accuracy, Inaccuracy, and Bias
20–39 <i>n</i> = 18	Suchey–Brooks Pubic Symphysis (89%)	Experience (5.42)	Two-Step Ribs (0.33)	Suchey–Brooks Pubic Symphysis (89%, 8.81, 0.47)
40–59 <i>n</i> = 20	Suchey–Brooks Pubic Symphysis (95%)	Experience (5.63)	İşcan Ribs (1.68)	Experience (85%, 5.63, −0.08)
60–79 <i>n</i> = 20	Buckberry–Chamberlain (100%)	Experience (7.28)	Hartnett Ribs (−0.33)	None
Males <i>n</i> = 31	Buckberry–Chamberlain (88%)	Experience (7.19)	Two-Step Ribs (0.58)	None
Females <i>n</i> = 27	Suchey–Brooks Pubic Symphysis (96%)	Experience (4.91)	Two-Step Auricular Surface (0.19)	Experience (89%, 4.91, −0.39)
Total <i>n</i> = 58	Buckberry–Chamberlain (88%)	Experience (6.13)	Overlap Two-Step (0.01)	Experience (80%, 6.13, −0.63)

All strategy results are found in Tables 6 and 7. The strategy with the highest accuracy for the total sample (*n* = 58) was Buckberry–Chamberlain (BCAS)/Least Error Auricular Surface (LEAS) (88%). Lovejoy Auricular Surface (LJAS) (21%) was the least accurate strategy. Other strategy accuracies ranged from 38–85%. The strategy with the least amount of bias ($x \geq -1$, $x \leq 1$) was Overlap: Two-Step (TSOL) (0.01 years), while the strategy with the most amount of bias was LJAS (−8.72). Other method biases ranged from −0.63–6.44 years. Suchey–Brooks Pubic Symphysis (SBPS), İşcan Rib (ISR), LJAS, experience, and overlap tended to underestimate age and all other strategies tended to overestimate age. There was a significant difference ($p \leq 0.05$) in the bias results of LJAS (the strategy with the highest amount of bias) and all other strategies other than ISR ($p = 0.09$). Finally, the strategy with the lowest mean inaccuracy ($x < 10$ years) was the experience-based approach (6.13 years). The strategy with the highest mean inaccuracy was LJAS (13.75 years). All other strategies' mean inaccuracies ranged from 7.56–12.97 years. The experience-based strategy was the only one to meet all the accuracy and reliability criteria (accuracy = 80%, inaccuracy = 6.13 years, bias = −0.63 years) for the total sample.

3.3. Results by Age Cohort

The age cohorts in this study were defined as younger (20–39 years old), middle-age (40–59 years old), and older (60–79 years old). All accuracy and reliability (inaccuracy and bias) scores by age cohort are found in Tables 6 and 7. If the ANOVA results indicated that

there were differences in the mean bias or inaccuracies in the age-reporting strategies based on the age cohort, pairwise comparisons were carried out to examine the differences.

The accuracies for the younger, middle-age, and older cohorts ranged from 22–89%, 25–95%, and 15–100%, respectively. Suchey–Brooks Pubic Symphysis (SBPS) performed the best for the younger and middle-age cohorts, and BCAS/LEAS performed best for the older cohort. LJAS performed the worst for all the age cohorts. Mean inaccuracy ranged from 5.42–19 years, 5.63–14.43 years, and 7–20.93 years for the younger, middle-age, and older cohorts, respectively. The experience-based strategy had the lowest mean inaccuracies for all the age categories. LJAS performed the worst for the younger and older cohorts and BCAS/LEAS performed the worst for the middle-age category. Finally, mean biases ranged from 0.33–16.06 years, 0.05–9.48 years, and –20.93–12.08 years for each age category. Two-step Ribs (TSR) (0.33 years), Overlap (0.05 years), and Hartnett Ribs (HNR) (–0.33 years) performed best for the younger, middle-age, and older categories, respectively. BCAS/LEAS performed worst for the younger (16.06 years) and middle-age (9.48 years) cohorts, and LJAS (–20.93) performed the worst for the older category. Strategies tended to overestimate age for the younger and middle-age cohorts, but underestimate for the older cohort.

SBPS (accuracy = 89%, inaccuracy = 8.81 years, bias = 0.47 years) was the only strategy to meet all the criteria for accuracy and reliability for the younger-age cohort, the experience-based (accuracy = 85%, inaccuracy = 5.63 years, bias = –0.08 years) strategy was the only strategy to meet accuracy and reliability criteria for the middle-age cohort, and none of the strategies met the criteria for the older cohort.

3.4. Results by Sex

All accuracy and reliability (inaccuracy and bias) results for males and females can be found in Tables 6 and 7. The accuracy for females ranged from 11–96%, with SBPS performing the best and LJAS the worst. The accuracy for males ranged from 29–88%, with BCAS/LEAS performing the best and LJAS performing the worst.

Mean inaccuracies for females ranged from –13.50–4.97 years and males ranged between –13.50–7.19 years. The experience-based strategy performed the best for both cohorts. SBPS performed the worst for females and Two-step Auricular Surface (TSAS), for males. The strategy with the least bias for females was TSAS (0.19 years) and the most bias was LJAS (10.13 years). Other strategy biases ranged between –8.72–6.44 years. Overall, strategies tended to overestimate age for the female cohort. TSR (0.58 years) had the least bias for the male cohort and BCAS/LEAS (9.44 years) had the most. Other strategies ranged between –5.54–4.82 years. The strategies equally under and overestimated age for the male cohort.

Experience (accuracy = 89%, inaccuracy = 4.91 years, bias = –0.39 years) was the only strategy to meet the accuracy and reliability criteria for the female cohort, and none of the strategies met the criteria for the male cohort.

4. Discussion

This study compared the accuracy and reliability of sixteen final age-at-death estimates, which were derived from six common skeletal aging methods, as well as comprehensive strategies in order to identify the most appropriate final age estimate to report to medico-legal stakeholders. The results of this study reveal that only a few of the age-reporting strategies tested were both accurate and reliable for any single sample category (total sample, younger cohort, males, etc.), and none of the strategies were both accurate (<80%) and reliable (inaccuracy $x < 10$ years; bias $x \geq -1$, $x \leq 1$) for all categories. Four of the six sample categories had a single strategy that met both the accuracy and reliability criteria. When the sample was assessed as a whole, the most accurate and reliable final age estimate was produced using the experienced-based strategy. When the sample was divided by age cohort, Suchey–Brooks was most accurate and reliable method for the younger-age cohort, the experience-based strategy was most accurate and reliable for the middle-age

cohort, and no strategy was both accurate and reliable for the older cohort. When divided by sex, no strategy was accurate and reliable for the male cohort, and the experience-based strategy worked best for the female cohort.

Fourteen of the sixteen strategies had low inaccuracies for the younger-age cohort, demonstrating the strong correlation between chronological and skeletal age in the earlier years of life, even into adulthood [10]. None of the age-reporting strategies were both accurate and reliable for the older-age cohort. This finding supports the assertion that aging indicators become less accurate and reliable as chronological and skeletal age become less correlated [10]. Milner and colleagues [22] also found that the current aging methods are not failing everyone assigned to older-age categories, but rather those aged in their 40s through 70s. Similar to the results of this study, they found that experience-based estimates better captured the age-related variation in these individuals. This indicates that there are observable age-related changes occurring in the skeleton during these age intervals that our current methods simply do not capture.

In general, auricular surface methods performed poorly for each sample category. Lovejoy consistently had low accuracies and high inaccuracy and bias for all sample categories. The Lovejoy method provides small ranges (5–10 years) that do not overlap (i.e., 25–29, 30–34, etc.), leading to a greater chance of not including the decedent's actual age of death. Age estimates that do not include the decedent's age at death can greatly hinder the potential of positive identification. Not only was LJAS not accurate, it was also largely unreliable. Therefore, reporting an age-range based on the Lovejoy method in a forensic context is irresponsible as the estimate is likely to be incorrect and significantly different from the actual age of a decedent. While Buckberry–Chamberlain had relatively high accuracies, this method should also not be used in forensic casework. The method produces ranges so large that the mean ages were significantly different from the actual ages of the individuals in this study. As previously noted [4], the high accuracy of Buckberry–Chamberlain is directly explained by ranges that average fifty-year spans after the first two stages. These age ranges are so large that they provide little value for narrowing down identification. The t-test showed that both auricular surface methods performed poorly in this study and did not capture age-related variation that well. Therefore, neither are appropriate for forensic casework.

The two-step strategies employed in this study were not the most accurate and reliable when the sample was divided by age cohorts. Suchey–Brooks was the most accurate and reliable for the younger-age, experience was the most accurate and reliable for the middle-age, and none were both accurate and reliable for the older cohort. Similar to Merritt [4], the original methods used in this study tended to perform well for the younger individuals and the revised methods performed better for the older individuals. The superior performance of the Suchey–Brooks method for the younger cohort is likely due to its development on a young sample, a trend which has been noted by several scholars [3–5,9]. This result corresponds with those found by Martrille et al. [3], who concluded that SBPS was the most accurate method for aging individuals in their younger-age cohort (25–40 years old). Merritt [4] also shows that SBPS had the lowest bias scores and had among the lowest inaccuracies for the younger-age cohort (20–39 years old) in her study. While these trends, showing that original methods perform better for younger individuals and that the revised perform better for older individuals, support the fundamental concept of the two-step approach, the results of this study do not support the use of this strategy in producing a final age estimate for medicolegal stakeholders.

Sex-specific methods did not perform better than other strategies when the sample was separated by sex. The experience-based strategy was the most accurate and reliable for females and none of the strategies met the criteria for males. While the experience-based estimate did not meet the accuracy criteria for males, it did meet the reliability criteria. This suggests that experience can capture age-related changes associated with sex better than any of the sex-specific methods used in this study.

Finally, the results of this study support the use of practitioner experience in determining a final age estimate, in spite of the fact that the analyst had only been practicing for less than five years. For the experience-based approach, the researcher was able to consider multiple lines of evidence, including all method results and the overall condition of the skeleton. This approach contributed to a better approximation of age as more of the skeletal variation could be captured. The experience-based estimates were the most accurate and reliable for the middle-age cohort, female cohort, and total sample. The experience-based strategy also resulted in low inaccuracy scores for all the sample categories, but higher bias scores for the younger and older cohorts. Additionally, while the experience-based strategy did not meet the accuracy criteria (80%) for all of the sample categories, its accuracies were 70% or above for all groups. These results are consistent with those of Baccino et al. [7], who found that both observers in their study achieved high accuracies using the “global approach,” which resulted in experience-based estimates. Additionally, Parsons [23] found that age estimates documented in resolved case reports were 92% accurate, and she attributed this success to practitioners’ reliance on multiple methods for producing final age estimates.

5. Conclusions

This research provides forensic anthropologists with valuable insights regarding the efficacy of some of the strategies currently used to produce final age estimates. The results of this study show that the most accurate and reliable age-reporting strategy varied if the sample was evaluated as a whole, by age cohort, or by sex. While none of the strategies were consistently the most accurate and reliable for all sample categories, the experience-based approach performed well in each category. The experience-based strategy allowed the researcher to use the results of the individual aging methods and professional judgment to arrive at a final age estimate. Despite not having known, or potentially known, the error rates associated with it, the experienced-based strategy is recommended to better capture chronological age in forensic casework. The estimates produced using the auricular surface ranges were either too large to provide exclusionary power in a forensic case or did not produce accurate or reliable age-at-death estimates. It is recommended that auricular surface aging be avoided in forensic casework as age ranges provided by the two most commonly used methods are either unrealistically narrow and statistically invalid (LJAS), or so broad they contribute minimally to identification (BCAS).

Several limitations of this study should be noted. First, only three age indicators were evaluated in this study, despite the availability of aging methods that are focused on other regions of the skeleton, such as the teeth [24], acetabulum, [25] and sacrum [26]. Understanding the accuracy and reliability of reporting strategies that produce estimates from multiple areas of the skeleton is beneficial for deciding how to report age if certain elements are not recovered in a forensic situation. To address this, future studies evaluating reporting strategies should diversify and/or expand the number of methods included. Secondly, this study did not include transition analysis as an age-reporting strategy, despite its ability to combine aging indicators in a way that is statistically valid. Transition analysis was excluded from this study because it is not widely used in forensic practice [2,23]. However, it has been shown to perform well in validation studies [22] and has been included in the most updated version of the Data Collection Procedures manual [27]. It is possible that transition analysis could provide accurate and reliable age estimates that also meet the *Daubert* standards.

It is crucial to understand how to report a person’s age at death in a manner that is both accurate and reliable. This study was able to shed light on the performance of different age-reporting strategies and provide further support when relying upon multiple aging indicators in determining a final age estimate. Ultimately, many factors contribute to how final age estimates are produced, all of which cannot be included within a single research design. Therefore, more studies like this one can help with the pursuit of better age-at-death estimates, and ultimately increase the identification of unknown skeletal remains.

Author Contributions: Conceptualization, C.B. and G.V.; methodology, C.B. and G.V.; formal analysis, C.B.; investigation, C.B.; data curation, C.B.; writing—original draft preparation, C.B.; writing—review and editing, C.B. and G.V.; supervision, G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to acknowledge the individuals who donated their remains to the UTK Donated Skeletal Collection, making this research possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Daubert v. Merrell Dow Pharmaceuticals. 1993. Available online: <https://supreme.justia.com/cases/federal/us/509/579/> (accessed on 13 March 2023).
2. Garvin, H.M.; Passalacqua, N.V. Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *J. Forensic Sci.* **2011**, *57*, 427–433. [[CrossRef](#)] [[PubMed](#)]
3. Martrille, L.; Ubelaker, D.H.; Cattaneo, C.; Seguret, F.; Tremblay, M.; Baccino, E. Comparison of Four Skeletal Methods for the Estimation of Age at Death on White and Black Adults. *J. Forensic Sci.* **2007**, *52*, 302–307. [[CrossRef](#)]
4. Merritt, C.E. Testing the Accuracy of Adult Skeletal Age Estimation Methods: Original Methods versus Revised and Newer Methods. *Explor. Anthropol.* **2013**, *12*, 102–119.
5. Miranker, M. A Comparison of Different Age Estimation Methods of the Adult Pelvis. *J. Forensic Sci.* **2016**, *61*, 1173–1179. [[CrossRef](#)] [[PubMed](#)]
6. Saunders, S.; Fitzgerald, C.; Rogers, T.; Dudar, C.; McKillop, H. A Test of Several Methods of Skeletal Age Estimation using a Documented Archaeological Sample. *Can. Soc. Forensic Sci. J.* **1992**, *25*, 97–118. [[CrossRef](#)]
7. Baccino, E.; Ubelaker, D.H.; Hayek, L.-A.C.; Zerilli, A. Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. *J. Forensic Sci.* **1999**, *44*, 12019J. [[CrossRef](#)]
8. Baccino, E.; Sinfield, L.; Colomb, S.; Baum, T.P.; Martrille, L. Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. *Forensic Sci. Int.* **2014**, *244*, 247–251. [[CrossRef](#)]
9. Lovejoy, C.O.; Meindl, R.S.; Mensforth, R.P.; Barton, T.J. Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. *Am. J. Phys. Anthr.* **1985**, *68*, 1–14. [[CrossRef](#)]
10. Nawrocki, S.P. The Nature and Sources of Error in the Estimation of Age at Death from the Skeleton. In *Age Estimation of the Human Skeleton*; Latham, K.E., Finnegan, M., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2010; pp. 79–101.
11. Brooks, S.; Suchey, J.M. *Skeletal Age Determination Base on the Os Pubis: A Comparison of the Acsádi-Nemeskéri and Suchey-Brooks Methods*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 227–238.
12. Lovejoy, C.O.; Meindl, R.S.; Pryzbeck, T.R.; Mensforth, R.P. Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *Am. J. Phys. Anthr.* **1985**, *68*, 15–28. [[CrossRef](#)]
13. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Age Estimation from the Rib by Phase Analysis: White Males. *J. Forensic Sci.* **1984**, *29*, 11776J. [[CrossRef](#)]
14. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *Am. J. Phys. Anthr.* **1984**, *65*, 147–156. [[CrossRef](#)] [[PubMed](#)]
15. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Age Estimation from the Rib by Phase Analysis: White Females. *J. Forensic Sci.* **1985**, *30*, 11018J. [[CrossRef](#)]
16. Hartnett, K.M. Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample-Part I: Pubic Bone. *J. Forensic Sci.* **2010**, *55*, 1145–1151. [[CrossRef](#)] [[PubMed](#)]
17. Buckberry, J.; Chamberlain, A. Age estimation from the auricular surface of the ilium: A revised method. *Am. J. Phys. Anthr.* **2002**, *119*, 231–239. [[CrossRef](#)]
18. Hartnett, K.M. Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample-Part II: Sternal End of the Fourth Rib. *J. Forensic Sci.* **2010**, *55*, 1152–1156. [[CrossRef](#)] [[PubMed](#)]
19. Yule, G.U. On the Theory of Correlation. *J. R. Stat. Soc.* **1897**, *60*, 812. [[CrossRef](#)]
20. IBM Corp. Released 2016. *IBM SPSS Statistics for Windows, Version 24.0*; IBM Corp: Armonk, NY, USA, 2016.
21. Meindl, R.S.; Lovejoy, C.O.; Mensforth, R.P.; Walker, R.A. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. *Am. J. Phys. Anthr.* **1985**, *68*, 29–45. [[CrossRef](#)]
22. Milner, G.R.; Boldsen, J.L. Transition analysis: A validation study with known-age modern American skeletons. *Am. J. Phys. Anthr.* **2012**, *148*, 98–110. [[CrossRef](#)]

23. Parsons, H.R. *The Accuracy of the Biological Profile in Casework: An Analysis of Forensic Anthropology Reports in Three Medical Examiner Offices*; University of Tennessee: Knoxville, TN, USA, 2017.
24. Lamendin, H.; Baccino, E.; Humbert, J.F.; Tavernier, J.C.; Nossintchouk, R.M.; Zerilli, A. A Simple Technique for Age Estimation in Adult Corpses: The Two Criteria Dental Method. *J. Forensic Sci.* **1992**, *37*, 13327J. [[CrossRef](#)]
25. San-Millán, M.; Rissech, C.; Turbón, D. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *Int. J. Leg. Med.* **2016**, *131*, 501–525. [[CrossRef](#)]
26. Passalacqua, N.V. Forensic Age-at-Death Estimation from the Human Sacrum. *J. Forensic Sci.* **2009**, *54*, 255–262. [[CrossRef](#)] [[PubMed](#)]
27. Langley, N.R.; Jantz, L.M.; Ousley, S.D.; Jantz, R.L.; Milner, G. Data Collection Procedures for Forensic Skeletal Material 2.0. 2016. Available online: https://fac.utk.edu/wp-content/uploads/2016/03/DCP20_webversion.pdf (accessed on 13 March 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.