

ASSESSING THE ACCURACY OF TRANSITION ANALYSIS 3.0 FOR ESTIMATING AGE IN A BLACK SOUTH AFRICAN POPULATION

Jennifer Tempest

2483228

A dissertation submitted to the Faculty of Health Sciences, University of the
Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of
Master of Science in Medicine

Supervisors:

Dr Anja Meyer

Prof Desiré Brits

Johannesburg, 2024



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I **Jennifer Tempest** (Student number: **2483228**) am a student registered for the degree of **Master of Science in Medicine** in the academic year **2024**.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature:  Date: 26 July 2024

Abstract

Age estimation is a complex part of the biological profile, influenced by various traits across the skeleton. The topic of combining multiple methods for estimating age is extensively debated, but the vitality of multifactorial age estimation methods is acknowledged. Transition analysis attempts to solve this problem using Bayesian statistics, particularly to avoid the effects of age mimicry. This study assessed the validity of Transition Analysis 3.0 (TA3) for the age estimation of black South Africans in a forensic context. The study sample consisted of the skeletal remains of 89 (44 females and 45 males) black South African individuals aged 15 to 105 years. This research assessed the validity of TA3 on dry bone via the analysis of trait-age correlations, sex-related and side-related differences, the correlation between true and estimated age, accuracy, inaccuracy, bias, and the effect of missing data.

None of the traits correlated strongly with age, but estimates improved when poorly correlated traits were removed from the analysis. There were significant sex-related differences and significant asymmetry in bony dimensions and the scoring of the pubic symphysis. The average accuracy was approximately 70%, with the highest accuracies observed in individuals aged 35-84 years. The lowest bias and inaccuracy were observed in middle-aged individuals, but adolescents and young adults were over-estimated, and older adults were under-estimated. No significant difference in the predictive capability was noted when full regions of traits were removed.

The low repeatability of some traits may be related to the fact that numerous traits used in TA3 are rarely used in age estimation, so practitioners are inexperienced in assessing them. The sex-related differences are consistent with the literature, but the findings of asymmetry in the pubic symphyseal traits are important to note. Problems related to accuracy, including the attraction to the middle, are likely related to the contents of the reference sample. In the case of TA3, the reference sample is primarily European-American and has a minimal representation of black African individuals.

TA3 requires further revision of the included traits and substantial expansion of the reference sample, particularly emphasising non-European-American samples. Additionally, further testing in other South African populations is required before its use in South African a forensic context.

This study showed that TA3 is neither reliable nor accurate enough for use in a South African forensic context and that the correlations between trait score and age are moderate at best. This study also highlighted problems with the methodology, software, and statistics.

Acknowledgements

I would like to extend my thanks to my supervisors, Dr Anja Meyer and Prof Desiré Brits, who provided substantial insight and support during all phases of this research. My thanks also to Mandy Mokoena, who did my inter-observer repeatability scoring, and my office mates who gave good advice when I ran into problems while writing. Many thanks to Dr Brendon Billings, Dr Nicholas Bacci, and Mr Mashudu Mulaudzi who allowed me to use the Raymond A. Dart Collection of Modern Human Skeletons for my research and who assisted me with retrieving the individuals who made up my sample.

My thanks to the National Research Foundation for funding this research in part, and to the University of the Witwatersrand for the Postgraduate Merit Award which also helped to fund this research.

I am also extremely grateful to my family and friends who provided a solid support network, across many time zones and through many rough patches.

I want to extend a special thanks to the individuals who made up my study sample. Thank you for your donation of your remains to further the education of many students, and to further research that assists in the identification of unknown victims of homicide in this country and across the world. Without you, none of this research would have been possible.

Table of Contents

ABSTRACT	III
ACKNOWLEDGEMENTS	V
LIST OF FIGURES.....	IX
LIST OF TABLES.....	XVII
NOMENCLATURE AND ABBREVIATIONS	XIX
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 LITERATURE REVIEW	4
2.1 Age Estimation	4
2.2 Adult age estimation methods	5
2.2.1 Suture, synchondrosis, and joint fusion	5
2.2.2 Epiphyseal fusion	8
2.2.3 Phase analysis	10
2.2.4 Degenerative changes	12
2.2.5 Other methods	18
2.3 Transition Analysis	20
CHAPTER 3 MATERIALS AND METHODS	28
3.1 Ethics	28
3.2 Materials	28
3.2.1 Sample	28
3.3 Methods	30
3.3.1 Data collection	30
3.3.2 Repeatability	31

3.3.3 Age estimation using the TA3 software	31
3.3.4 Data analysis	31
CHAPTER 4 RESULTS	36
4.1 Repeatability	36
4.2 Correlation of trait to true age	37
4.2.1 Correlations	37
4.2.2 Differences between males and females	48
4.2.3 Asymmetry	50
4.3 Correlation of estimated age to true age	50
4.4 Accuracy, inaccuracy, and bias	53
4.5 Sensitivity to missing data	61
CHAPTER 5 DISCUSSION	65
5.1 Repeatability	65
5.2 Correlations between traits and true age and true age with estimated age	70
5.3 Differences between the left and right sides	77
5.4 Accuracy, inaccuracy, and bias	79
5.5 Ability to work with incomplete information	82
5.6 Problems identified during data collection and analysis	84
5.7 Limitations and future research directions	87
CHAPTER 6 CONCLUSION	90
REFERENCES	92
APPENDIX A ETHICAL CLEARANCE	112

APPENDIX B SCORING MANUAL	114
APPENDIX C EXPANDED RESULTS TABLES & DIAGRAMS	145
C.1 Repeatability	145
C.2 Correlation of trait to true age	150
C.3 Differences between males and females	200
C.4 Differences between the left and right sides	206
APPENDIX D SIMILARITY REPORT	210

List of Figures

Fig. 4.1 Scatter plots for the traits with a moderate correlation with age in the pooled-sex sample ($n = 89$, $\alpha = 0.05$); part one	38
Fig. 4.2 Scatter plots for the traits with a moderate correlation with age in the pooled-sex sample ($n = 89$, $\alpha = 0.05$); part two	39
Fig. 4.3 Scatter plots for the traits with a moderate correlation with age in the male sample ($n = 45$, $\alpha = 0.05$); part one	42
Fig. 4.4 Scatter plots for the traits with a moderate correlation with age in the male sample ($n = 45$, $\alpha = 0.05$); part two	43
Fig. 4.5 Scatter plots for the traits with a moderate correlation with age in the female sample ($n = 44$, $\alpha = 0.05$); part one	45
Fig. 4.6 Scatter plots for the traits with a moderate correlation with age in the female sample ($n = 44$, $\alpha = 0.05$); part two	46
Fig. 4.7 Scatter plots for the traits with a moderate correlation with age in the female sample ($n = 44$, $\alpha = 0.05$); part three	47
Fig. 4.8: Scatterplots of true age vs point-estimated age with a Spearman's Rank Correlation Coefficient ($\alpha = 0.05$). (a) Female sample ($n = 44$), (b) Male sample ($n = 45$), (c) Pooled-sex sample ($n = 89$), (d) Female sample with poorly correlated traits removed from the analysis ($n = 44$), (e) Male sample ($n = 45$) with poorly correlated traits removed from the analysis, (f) Pooled sex sample with poorly correlated traits removed ($n = 89$).....	52
Fig. 4.9 Box plots of the age distribution for the female ($n = 44$, purple), male ($n = 45$, green), and pooled-sex ($n = 89$, grey) samples showing the differences between the true distribution, the first estimated distribution with all traits present (point estimate 1), and the second estimated distribution with poorly correlated traits removed (point estimate 2)	60
Fig. 4.10 Scatter plots of true age vs point estimate age for sub-samples in which a single scoring region was fully removed ($n = 68$)	62
Fig. 4.11 Scatter plots of true age vs point estimate age for sub-samples in which a single scoring region was fully removed ($n = 68$)	63
Fig. 5.1 The GUI for the scoring of vertebral lippling, candlewax, and DISH.....	85
Fig. 5.2 The GUI for the scoring of the humerus relative weight, showing the lack of bilateral scoring	85

Fig. 5.3 The GUI for traits associated with the ribs, showing the lack of bilateral scoring for all relevant rib traits except the fusion of rib one to the manubrium	86
Fig. 5.4 The age estimates for a 16-year-old male when (a) all lower limb traits were removed from the analysis, and (b) all upper limb traits were removed from the analysis	86
Fig. 5.5 The age estimate for a 15-year-old male for whom the male-specific poorly correlated traits were removed	86
Fig. 5.6 The age estimate for a 15-year-old female for whom the female-specific poorly correlated traits were removed	86
Fig. 5.7 The age estimate for a 17-year-old male for whom the pooled-sex sample's poorly correlated traits were removed	87
Fig. B.1 Parietal depression (a) 0, (b) 1	115
Fig. B.2 Spheno-occipital synchondrosis (a) 0, (b) 1	115
Fig. B.3 Occipital condyle lipping (a) 0, (b) 1, (c) 2	115
Fig. B.4 Femur fovea margin lipping (a) 0, (b) 1	118
Fig. B.5 Femur head surface extra bone (a) 0, (b) 1, (c) 2	118
Fig. B.6 Femur greater trochanter roughening (a) 0, (b) 1, (c) 2, (d) 3 lateral view, (e) 3 posterior view	118
Fig. B.7 Femur trochanteric fossa exostoses (a) 0, (b) 1	119
Fig. B.8 Femur trochanteric medial exostoses (a) 0, (b) 1	119
Fig. B.9 Fibula wings (a) 0, (b) 1	119
Fig. B.10 Clavicle medial epiphysis fusion (a) 0, (b) 1, (c) 2 (Boldsen <i>et al.</i> , 2021a), (d) 3	122
Fig. B.11 Clavicle medial epiphysis gravel (a) 0, (b) 1	122
Fig. B.12 Clavicle medial macroporosity (a) 1, (b) 2	123
Fig. B.13 Clavicle lateral macroporosity (a) 0, (b) 1	123
Fig. B.14 Scapula margin lipping (a) 0, (b) 1, (c) 2	123
Fig. B.15 Humerus lesser tubercle bumps (a) 0, (b) 1	124
Fig. B.16 Humerus lesser tubercle margin lipping (a) 0, (b) 1, (c) 2	124
Fig. B.17 Humerus greater tubercle pits (a) 0, (b) 1	124
Fig. B.18 Humerus medial epicondyle (a) 0, (b) 1	124
Fig. B.19 Humerus lateral epicondyle (a) 0, (b) 1	125
Fig. B.20 Radius tuberosity medial crest (a) 0, (b) 1	125
Fig. B.21 Ulna olecranon spur (a) 0, (b) 1	125

Fig. B.22 Trapezium lipping (a) 0, (b) 1	125
Fig. B.23 Sacro-iliac joint fusion (a) 0, (b) 1, (c) 2	128
Fig. B.24 Iliac crest fusion (a) 0, (b) 1, (c) 2	128
Fig. B.25 Iliac crest tuberculum ossification (a) 0 (insufficient size), (b) 1, (c) 2	129
Fig. B.26 Ilium anterior inferior iliac crest exostoses (a) 0, (b) 1 anterior view, (c) 1 lateral view	129
Fig. B.27 (a) Acetabulum posterior margin lipping & articular surface extra bone 0, (b) lipping 1, (c) extra bone 1	129
Fig. B.28 Acetabulum inferior joint lipping (a) 0, (b) 1	130
Fig. B.29 Ischial tuberosity superior margin spur (a) 0, (b) 1	130
Fig. B.30 Ischial tuberosity medial spur (a) 0, (b) 1	130
Fig. B.31 Ischial tuberosity bumps (a) The new scoring area highlighted in blue and green (Semimembranosus attachment in blue and semitendinosus & biceps femoris attachments in green), (b) 0, (c) 1, (d) 2	131
Fig. B.32 Auricular surface inferior surface porosity (a) 0, (b) 1	133
Fig. B.33 (a) Auricular surface superior posterior exostoses, inferior posterior exostoses, & posterior exostoses 0, (b) Auricular surface superior posterior exostoses 1, (c) Auricular surface inferior posterior exostoses & posterior exostoses 1, (d) Auricular surface posterior exostoses 2 (Boldsen <i>et al.</i> , 2021a)	133
Fig. B.34 Pubic symphyseal collar (a) 0, (b) 1	135
Fig. B.35 (a) Pubic symphyseal relief 0, Pubic superior protuberance 0, (b) Pubic symphyseal relief 1, (c) Pubic symphyseal relief 2, (d) Pubic superior protuberance 1, (e) Pubic superior protuberance 2	135
Fig. B.36 Pubic symphysis ventral margin & dorsal margin (a) 0, (b) 1, (c) 2 (Boldsen <i>et al.</i> , 2021a), (d) 3	136
Fig. B.37 (a) C1 lipping & eburnation 0, (b) C1 lipping 1, (c) C1 eburnation 1	139
Fig. B.38 L1 spinous process flattening (a) 0, (b) 1, (c) 2	139
Fig. B.39 L1-5 epiphyseal ring fusion (a) 0 lateral view, (b) 0 superior view, (c) 2, (d) 3	139
Fig. B.40 L1-5 surface morphology (a) 0, (b) 1, (c) 2	139
Fig. B.41 L5 margin shape (a) 0, (b) 1, (c) 2	140
Fig. B.42 S1 margin shape (a) 0, (b) 1, (c) 2 superior view, (d) 2 anterior view	140
Fig. B.43 S1-2 fusion (a) 0, (b) 1, (c) 2	140

Fig. B.44 Sacral elbow (a) 0, (b) 1, (c) anterior view of the sacral elbow and healed fracture	141
Fig. B.45 (a) Vertebral lipping & candlewax 0, (b) Vertebral lipping 1, (c) Vertebral candlewax 1	141
Fig. B.46 DISH 1	141
Fig. B.47 Rib 1 fusion (a) 0, (b) 1	143
Fig. B.48 Sternum dorsal ridge (a) 0, (b) 1	143
Fig. B.49 Rib sternal end rim profile (a) 0, (b) 1	143
Fig. B.50 Rib ossification pattern (a) 0, (b) 1, (c) 2, (d) 3.....	144
Fig. B.51 Shingle ribs (a) 0, (b) 1 (Boldsen <i>et al.</i> , 2021a).....	144
Fig. C.1 Scatterplots of trait score vs age for the left femur fovea head margin lipping, head extra bone, and greater trochanter roughening for the female (n = 44), male (n = 55), and pooled-sex (n = 89) samples from left to right.....	156
Fig. C.2 Scatterplots of trait score vs age for left femur trochanteric fossa exostoses, trochanteric medial exostoses, and right femur fovea margin for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	157
Fig. C.3 Scatterplots of trait score vs age for right femur head extra bone, greater trochanter roughening, and trochanteric fossa exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	158
Fig. C.4 Scatterplots of trait score vs age for right femur trochanteric medial exostoses, left tibia relative weight, and actual weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	159
Fig. C.5 Scatterplots of trait score vs age for left tibia length, right tibia relative weight, and actual weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	160
Fig. C.6 Scatterplots of trait score vs age for right tibia length, left fibula wings, and right fibula wings for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	161
Fig. C.7 Scatterplots for trait score vs age for left calcaneus relative weight, actual weight, and length for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	162
Fig. C.8 Scatterplots for trait score vs age for right calcaneus relative weight, actual weight, and length for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	163

Fig. C.9 Scatterplots of trait score vs age for left clavicle medial epiphysis gravel, medial macroporosity, and lateral macroporosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	164
Fig. C.10 Scatterplots of trait score vs age for right clavicle medial epiphysis gravel, medial macroporosity, and lateral macroporosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	165
Fig. C.11 Scatterplots of trait score vs age for left scapula glenoid fossa lipping, humerus relative weight, and actual weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	166
Fig. C.12 Scatterplots of trait score vs age for left humerus length, right humerus relative weight, and actual weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	167
Fig. C.13 Scatterplots of trait score vs age for right humerus length, left humerus lesser tubercle anterior surface bumps, and lesser tubercle margin shape for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	168
Fig. C.14 Scatterplots of trait score vs age for left humerus greater tubercle pits, medial epicondyle, and lateral epicondyle for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	169
Fig. C.15 Scatterplots of trait score vs age for right humerus lesser tubercle anterior surface bumps, lesser tubercle margin shape, and greater tubercle pits for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	170
Fig. C.16 Scatterplots for trait score vs age for right humerus epicondyle, lateral epicondyle, and left radius tuberosity medial crest for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right:	171
Fig. C.17 Scatterplots for trait score vs age for right radius tuberosity medial crest, left ulna olecranon spur, and right ulna olecranon spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	172
Fig. C.18 Scatterplots for trait score vs age for left trapezium lipping, right trapezium lipping, and left innominate relative weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	173
Fig. C.19 Scatterplots for trait score vs age for left innominate actual weight, length, and right innominate relative weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	174

Fig. C.20 Scatterplots for trait score vs age for right innominate actual weight, length, and left sacroiliac joint fusion for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	175
Fig. C.21 Scatterplots for trait score vs age for right sacroiliac joint fusion, left iliac crest fusion, and tuberculum ossification for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	176
Fig. C.22 Scatterplots for trait score vs age for left ilium anterior inferior iliac exostoses, right iliac crest fusion, and tuberculum ossification for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	177
Fig. C.23 Scatterplots for trait score vs age for right ilium anterior inferior iliac exostoses, left acetabulum posterior margin lipping, and posterior articular surface extra bone for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	178
Fig. C.24 Scatterplots for trait score vs age for left acetabulum inferior joint lipping, right acetabulum posterior margin lipping, and posterior articular surface extra bone for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	179
Fig. C.25 Scatterplots for trait score vs age for right acetabulum inferior joint lipping, left ischial tuberosity superior margin spur, and medial spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	180
Fig. C.26 Scatterplots for trait score vs age for left ischial tuberosity bumps, right ischial tuberosity superior margin spur, and medial spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	181
Fig. C.27 Scatterplots for trait score vs age for right ischial tuberosity bumps, left auricular surface inferior surface porosity, and superior posterior exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	182
Fig. C.28 Scatterplots for trait score vs age for left auricular surface inferior posterior exostoses, posterior exostoses, and right auricular surface inferior surface porosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	183
Fig. C.29 Scatterplots for trait score vs age for right auricular surface superior posterior exostoses, inferior posterior exostoses, and posterior exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	184

Fig. C.30 Scatterplots for trait score vs age for left pubic symphyseal collar, relief, and superior protuberance for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	185
Fig. C.31 Scatterplots for trait score vs age, for left pubic symphysis ventral margin, dorsal margin, and right pubic symphysis collar for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	186
Fig. C.32 Scatterplots for trait score vs age for right pubic superior protuberance, dorsal margin, and C1 lipping for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	187
Fig. C.33 Scatterplots for trait score vs age for C1 eburnation, L1 superior epiphyseal ring fusion, and inferior epiphyseal ring fusion for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	188
Fig. C.34 Scatterplots for trait score vs age for L1 superior surface morphology, inferior surface morphology, and L5 superior epiphyseal ring fusion for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	189
Fig. C.35 Scatterplots for trait score vs age for L5 inferior epiphyseal ring fusion, superior surface morphology, and inferior surface morphology for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	190
Fig. C.36 Scatterplots for trait score vs age for L5 superior margin shape, inferior margin shape, and S1 margin shape for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	191
Fig. C.37 Scatterplots of trait score vs age cervical lipping, candlewax, and thoracic lipping for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	192
Fig. C.38 Scatterplots for trait score vs age for thoracic candlewax, lumbar candlewax, and DISH for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	193
Fig. C.39 Scatterplots for trait score vs age for left rib 1 fusion, right rib 1 fusion, and sternum central dorsal ridge full sample for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	194
Fig. C.40 Scatterplots for trait score vs age for left rib 2 rim profile, ossification pattern, and right rib 2 rim profile for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right.....	195

Fig. C.41 Scatterplots for trait score vs age for right rib 2 ossification pattern, left ribs 3-10 rim profile, and ossification pattern for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	196
Fig. C.42 Scatterplots for trait score vs age for left ribs 2-10 shingle ribs, right ribs 3-10 rim profile, and ossification pattern for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	197
Fig. C.43 Scatterplots for trait score vs age for right ribs 2-10 shingle ribs, left parietal depression, and right parietal depression for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	198
Fig. C.44 Scatterplots for trait score vs age for spheno-occipital synchondrosis, left occipital condyle lipping, and right occipital condyle lipping for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right	199

List of Tables

Table 3.1 Sample size and distribution	28
Table 3.2 Modified agreement scale for Cohen's Kappa from Landis and Koch (1977)	32
Table 3.3 Agreement scale for Lin's CCC from McBride (2005)	32
Table 3.4 Modified agreement scale for a linear relationship from Chan (2003).....	33
Table 4.1 Traits in the pooled-sex sample (n = 89) with "moderate" strength of correlation with age	37
Table 4.2 Traits in the male sample (n = 45) with "moderate" strength of correlation with age.....	41
Table 4.3 Traits in the female sample (n = 44) with "moderate" strength of correlation with age.....	44
Table 4.4: Traits with significant difference between males (n = 45) and females (n = 44) for the correlation between trait and age as determined by a Fisher's Z transformation with $\alpha = 0.05$	48
Table 4.5 Traits with significant differences in scoring between males (n = 45) and females (n = 44) as determined by a Mann-Whitney U (Wilcoxon rank sum) test with an α of 0.05	49
Table 4.6 Traits with significant differences in the correlation between trait and age on the left and right sides as determined by a percentile bootstrap ($\alpha = 0.05$, n = 89, B = 2000).....	50
Table 4.7 Traits with significant differences in scoring between the left and right sides as determined by a Wilcoxon signed rank test (paired, n = 89, $\alpha = 0.05$).....	50
Table 4.8 Accuracy of age estimates with all traits by age cohorts for the pooled sex (n = 89), male (n = 45), and female (n = 44) samples	54
Table 4.9 Bias and inaccuracy by age cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples when all traits were included in the analysis	55
Table 4.10 Age estimate accuracy per cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples after poorly correlated traits were removed from the analysis	57
Table 4.11 Bias and inaccuracy per cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples after poorly correlated traits were removed from the analysis	58

Table 4.12: Spearman's Rank Correlation Coefficient ($\alpha=0.05$, $n=68$) for true age vs estimated age in a complete dataset and incomplete datasets (pooled-sexes)	61
Table 4.13: Fisher's Z transformation test ($\alpha = 0.05$, $n = 68$) for a significant difference in correlation between true age and estimated age in a full dataset vs incomplete dataset combinations	64
Table B.1 Cranial traits descriptions and scoring	114
Table B.2 Lower limb traits descriptions and scoring	116
Table B.3 Upper limb traits descriptions and scoring	120
Table B.4 Innominate traits descriptions & scoring	126
Table B.5 Auricular surface traits descriptions & scoring	132
Table B.6 Pubic symphysis traits descriptions & scoring	134
Table B.7 Vertebral traits descriptions and scoring	137
Table B.8 Ribs & sternum traits descriptions & scoring	142
Table C.1 Cohen's Weighted Kappa for intra-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$)	145
Table C.2 Lin's Concordance Correlation Coefficient for intra-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$) *density calculated by dividing actual weight by length as recommended in the trait scoring guide	147
Table C.3 Cohen's Weighted Kappa for inter-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$)	147
Table C.4 Lin's Concordance Correlation Coefficient for inter-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$) *density calculated by dividing actual weight by length as recommended in the trait scoring guide	149
Table C.5 The Spearman's Rank Correlation Coefficient for the correlation between trait and true age in the full ($n = 89$), male ($n = 45$), and female ($n = 44$) samples ($\alpha = 0.05$). (1According to Chan (2003)).....	150
Table C.6 Fisher's Z transformation ($\alpha = 0.05$) comparing male ($n = 45$) vs female ($n = 44$) age-trait correlations	200
Table C.7 Mann-Whitney U ($\alpha = 0.05$) test for male ($n = 45$) vs female ($n = 44$) trait scores with H_0 = no difference in distribution.....	203
Table C.8 Percentile bootstrap on a Spearman's rho test ($\alpha = 0.05$, $n = 89$, $B = 2000$) comparing the correlation of trait-age between the left and right side	206
Table C.9 Wilcoxon signed rank test (paired, $n = 89$, $\alpha = 0.05$) for left vs right trait scores.....	208

Nomenclature and abbreviations

AS – Auricular Surface

AIIS – Anterior Inferior Iliac Spine

CCC – Concordance Correlation Coefficient

CT – Computed Tomography

DF – Degrees of Freedom

DISH – Diffuse Idiopathic Skeletal Hyperostosis

DXA – Dual X-ray Absorptiometry

GUI – Graphic User Interface

HVIRU – Human Variation and Identification Research Unit

MSCT – Multi-Slice Computed Tomography

QCT – Quantitative Computed Tomography

TA2 – Transition Analysis 2

TA3 – Transition Analysis 3.0

Chapter 1 Introduction

In the event of the discovery of a single body or a mass grave, all efforts must be made to ascertain the identity of the individual(s) both to enact justice and to provide closure for their loved ones. To contribute to the identification of an individual based on their skeletal remains, forensic anthropologists establish a biological profile. This biological profile consists of estimations of age-at-death, sex, population affinity, and stature (İşcan and Steyn, 2013a). Age estimation is vital as it can notably reduce the pool of likely matches within a missing persons database, increasing the likelihood of an accurate identification (Algee-Hewitt, 2017). This is especially pertinent given South Africa's high homicide rate and relatively high rate of unidentified victims (Evert, 2011).

Age estimation is based on the association between biological and chronological age but is complicated by the considerable variation observed in biological age at any given chronological age (Boldsen and Milner, 2012; Algee-Hewitt, 2017; Langley *et al.*, 2017). Chronological age refers to the amount of time lived since the birth of an individual in years, months, weeks, and days (Acsádi and Neméskeri, 1970; Garvin *et al.*, 2012; İşcan and Steyn, 2013b). Biological age refers to the physiological state of the skeleton and is influenced by the environment, disease, and life history (Acsádi and Neméskeri, 1970; Garvin *et al.*, 2012; İşcan and Steyn, 2013b; Langley *et al.*, 2017). The rate and pattern at which an individual ages is not constant, so a single individual's ageing pattern will be a variant of the pattern observed in their population. Besides the variation seen within a population, variation exists between populations. The result of this is that a method developed with a reference sample from one population is unlikely to perform as well in another, distinct, population (Acsádi and Neméskeri, 1970; Oettlé and Steyn, 2000; Rissech *et al.*, 2007; Hens *et al.*, 2008; Garvin *et al.*, 2012; Jooste *et al.*, 2016; Mahon *et al.*, 2017; Marera and Satyapal, 2018; Meyer *et al.*, 2021; San-Millán and Rissech, 2023). As such, validation studies must be conducted to assess the performance of different methods in populations distinct from those used to develop those methods. Developing population- and sex-specific methods and variants is also essential.

A common problem with many age-estimation methods is 'age mimicry', where the age distribution of any given sample mimics that of the sample used to develop the method. This results in an estimated age distribution that differs from the actual age

distribution of the sample in question, which leads to a decrease in the accuracy of age estimates obtained from the method (Boldsen *et al.*, 2002; Işcan and Steyn, 2013b; Algee-Hewitt, 2017).

Transition Analysis (TA2) is a statistical method designed to avoid the effects of age mimicry while also providing a robust combination of multiple age-predicting traits. This is achieved via the use of Bayesian statistics and an informed prior distribution. (Boldsen *et al.*, 2002; Milner and Boldsen, 2012; Getz, 2020). This informed prior distribution is a large dataset with known ages-at-transition for numerous traits, with a known age distribution. Ideally, any given informed prior distribution should only be applied to a test sample it is representative of (Boldsen *et al.*, 2002; Milner and Boldsen, 2012).

Tests of TA2 have indicated several problems, including the lack of applicable population-specific prior distributions, relatively low accuracy, persistent under-estimation of age, and poor repeatability for some traits (Bethard, 2005; Milner and Boldsen, 2012; Bullock *et al.*, 2013; Jooste *et al.*, 2016; Hagelthorn *et al.*, 2019; Lopez-Cerquera and Casallas, 2019; Kim and Algee-Hewitt, 2022). These problems with TA2 are also common problems seen in many other age estimation methods (Kimmerle *et al.*, 2008; Hartnett, 2010a; Hens and Godde, 2016; Navega *et al.*, 2018b; Muñoz-Silva *et al.*, 2020). In response to the problems with TA2, a new version is under development, using notably more traits and a larger uniform prior distribution drawn from eight sample populations in the United States of America, Portugal, Thailand, and South Africa.

Any research's use in a medicolegal setting must be legally admissible in court. In forensic anthropology, a commonly accepted standard is that of the Daubert criteria, which include method testing, error rate estimation, peer review, and general acceptance of the method in question by practitioners in the field (Christensen and Crowder, 2009). Part of the satisfaction of these criteria requires the conduction of validation studies and the publication of standards to which any assessment can be compared (Dirkmaat *et al.*, 2008). A critical factor for a validation study is the use of a distinctly different sample than that used to develop a method to avoid repeating any results by chance (Brownlee, 2017; Valsecchi *et al.*, 2019). While the Daubert criteria

are not legally applicable in South Africa, the scientific principle of the validation of any given method remains vital.

Transition Analysis has been shown to perform worse than other single-region methods when applied to the South African population. As a result it is rarely used for age estimation in a forensic setting due to its poor performance in the South African context.

This study aimed to test the accuracy and repeatability of Transition Analysis 3.0 (TA3) in a sample of black South African adults. To achieve this aim, the following objectives had to be met:

- a) Assessing the inter- and intra-observer repeatability of the scoring of age-related features used in TA3.
- b) Assessing the correlation of each age feature or indicator with true age.
- c) Assessing whether there are differences in the degree of correlation of age indicators and increasing true age between males and females.
- d) Assessing whether there are differences in the degree of correlation of age indicators and increasing true age between the right and left sides.
- e) Assessing the degree of correlation between the estimated ages, generated using TA3, and true ages.
- f) Testing the accuracy, inaccuracy and bias of age estimates generated from TA3.
- g) Assessing and describing the sensitivity of TA3 to missing data.

Chapter 2 Literature Review

2.1 Age Estimation

Sub-adult age estimation is based on skeletal development, growth and maturation, which occurs in a rapid and known sequence, allowing for the generation of relatively narrow age estimates with high accuracies (İşcan and Steyn, 2013b; Langley *et al.*, 2017). Methods used include dental development and eruption (Willems *et al.*, 2001, 2018; AlQahtani *et al.*, 2010; Metsäniitty *et al.*, 2019), epiphyseal ossification and fusion (McKern and Stewart, 1957; Schaefer and Black, 2005; Cardoso, 2008; Cunningham *et al.*, 2016; Cardoso *et al.*, 2017), long bone diaphyseal length (Facchini and Veschi, 2004; Rissech *et al.*, 2013, 2014; Cardoso *et al.*, 2014) and the fusion of cranial sutures such as the fusion of the foramen magnum (Redfield, 1970; Bassed *et al.*, 2010; Cardoso *et al.*, 2013; Cunningham *et al.*, 2016).

In contrast, adult age estimation is primarily based on degenerative changes associated with increasing age. These changes are frequently associated with wear and tear and accumulate as age progresses (Garvin *et al.*, 2012; İşcan and Steyn, 2013b). Other causes for age-related changes to bone include metabolic, hormonal, and pathological changes as age progresses and the influence of the external environment over time (Garvin *et al.*, 2012; İşcan and Steyn, 2013b). Commonly used adult age estimation methods include the assessment of the pubic symphysis (Todd, 1920; Katz and Suchey, 1986; Brooks and Suchey, 1990; Hartnett, 2010a), auricular surface (Lovejoy *et al.*, 1985; Buckberry and Chamberlain, 2002; Osborne *et al.*, 2004), cranial sutures (Acsádi and Neméskeri, 1970; Meindl and Lovejoy, 1985; Nawrocki, 1998), and the sternal ends of the ribs (İşcan *et al.*, 1984; Russell *et al.*, 1993; Oettlé and Steyn, 2000; DiGangi *et al.*, 2009; Hartnett, 2010b).

It is widely considered that it is best to use multiple traits when estimating age and multiple methodologies to verify results (Buckberry, 2015; Márquez-Grant, 2015). Therefore, attempts have been made to combine different skeletal age indicators and develop multi-regional methods to account for the distribution of age-related changes throughout the skeleton. Unfortunately, there is a lack of consensus regarding which traits to include and how to weigh their contributions statistically (Acsádi and Neméskeri, 1970; Garvin and Passalacqua, 2012; İşcan and Steyn, 2013b; Algee-Hewitt, 2017). The lack of standardisation in both method combination and statistical

procedures leads to low repeatability and accuracy, which limits the use of these methods in a practical forensic setting (Garvin *et al.*, 2012; Buckberry, 2015; Algee-Hewitt, 2017).

Another well-documented problem affecting the accuracy of age estimation methods is age mimicry, in which a test sample's age distribution falsely imitates that of the sample used in the development of the method (Boldsen *et al.*, 2002; Işcan and Steyn, 2013b; Algee-Hewitt, 2017). Age mimicry typically results from a small sample size during method development, which results in the method failing to account for variation occurring naturally within and between populations (Boldsen *et al.*, 2002; Usher, 2002). The practical result of age mimicry is a decrease in accuracy and precision and an increase in bias when any method is applied to a population distinct from the original sample (Bocquet-Appel and Masset, 1982; Hoppa and Vaupel, 2002; Buckberry, 2015). As such, population-specific data is often required to use a specific method to account for the variation within the target population as well as inter-population differences in the timings of age-related changes (Boldsen *et al.*, 2002; Usher, 2002; Hens *et al.*, 2008; Garvin *et al.*, 2012; Mahon *et al.*, 2017; Meyer *et al.*, 2021; San-Millán and Rissech, 2023).

2.2 Adult age estimation methods

Transition analysis 3.0 (TA3) assesses age-related bony changes spread across the human skeleton. Some of these traits have been investigated before with other well-established methods. However, some traits have either not been formally investigated before or have been investigated in a different manner. This section discusses numerous age estimation methods spread throughout the skeleton compared to the traits and scoring methodologies of TA3. The methods are grouped according to the type of bony change occurring, such as epiphyseal fusion, degenerative change, and phasic changes.

2.2.1 Suture, synchondrosis, and joint fusion

The most commonly used methods for age estimation from the skull involve the cranial sutures despite their low reliability, low precision, and detrimental effects when included in multifactorial age estimation methods (Garvin *et al.*, 2012; Meyer *et al.*, 2021). This is because of their admittedly variable utility in combination with other methods and because some forensic or archaeological cases only include the cranium

(Todd and Lyon Jr., 1924; Garvin and Passalacqua, 2012; Garvin *et al.*, 2012). Meindl and Lovejoy (1985) indicated ten 1 cm long sites along the ectocranial sutures, which were scored on a four-point scale from completely unfused to completely fused. These scores were then added to form a composite score for the lateral-anterior and vault sutures. The composite scores were then compared to reference tables for the respective regions, which provided a mean age estimate and a potential age range. They observed substantial standard deviations and estimated age ranges (Meindl and Lovejoy, 1985).

Nawrocki (1998) expanded on Meindl and Lovejoy's (1985) work and developed a scoring method with 27 landmarks spread across the ectocranial and endocranial surfaces, as well as the palatine sutures. Nawrocki's (1998) correlation between the composite ectocranial score and age is higher than both correlations for Meindl and Lovejoy's (1985) vault and lateral-anterior composite scores with age. The largest impact of Nawrocki's (1998) research on cranial suture fusion is the extension of the maximum age estimate threshold, from 76-80 years in Meindl and Lovejoy (1985) to up to age 102 years, while reducing the standard error from approximately 12-20 years to 7-13.5 years. It should be noted, however, that Nawrocki's (1998) research sample only spanned the ages of 21 to 85. This artificial extension of the predictive bounds of the method is likely due to his use of regression analysis. It can be argued that the method's predictive age range should have been capped at the upper and lower bounds to reflect reality. Nawrocki's (1998) regression formulae also allow for a more case-specific age estimate based on how an individual is scored while taking their sex and population affinity into account instead of assigning a large age range to any individual irrespective of their sex or population affinity.

An additional cranial 'suture' investigated for age estimation is the spheno-occipital synchondrosis, which connects the occipital and sphenoid bones. The age at which it fuses and becomes obliterated is variable within the literature. Bassed *et al.* (2010) conducted a study on Computed Tomography (CT) scans of individuals between the ages of 15 and 25 years, and they found that fusion was generally completed by 17 years. Mahon *et al.* (2017) suggest that some of the variability in the literature may be due to the variety of methods used, as some studies used radiographic techniques, while others used dry bone and thus observed less detail.

Mahon and colleagues (2017) tested the method developed by Shirley and Jantz (2011) for scoring the stages of fusion of the spheno-occipital synchondrosis in a black South African population. They found the method was highly repeatable and that fusion was 100% complete by the age of 20 years in their sample. Additionally the females showed fusion three years before the males (Mahon *et al.*, 2017). McDonald and Miller (2022) found that older studies indicate fusion in late adolescence or early adulthood and more recent studies indicate fusion strictly in adolescence. However, there is a general trend of females beginning the process a few years before their male counterparts (McDonald and Miller, 2022). The current consensus is that the fusion of the spheno-occipital synchondrosis occurs during the second decade of life, but some mark may remain through the third decade of life (Bassed *et al.*, 2010; Cunningham *et al.*, 2016).

In addition to the fusion of the cranial sutures and synchondroses, the fusion of the sternal joints has been investigated for age estimation. Research on age estimation from the sternum typically involves the fusion of the manubrio-sternal and sterno-xiphoidal joints. Chandrakanth *et al.* (2012) scored the manubrio-sternal and sterno-xiphoidal joints as open or closed and found no significant relationship between age and the fusion of these joints. The age of fusion for each of the joints was also highly variable. Bacci *et al.* (2018) researched the fusion of the sternum in a black South African sample and scored fusion on a four-point scale from unfused to completely fused. They found no significant correlation between age and the fusion of either the manubrio-sternal or sterno-xiphoidal joints. Bacci *et al.* (2018) also found a high variation in the rate of joint fusion compared to age.

Monum *et al.* (2020) studied the fusion of costal cartilages, the manubrio-sternal joint, and the sterno-xiphoidal joint using CT scans instead of dry bone inspection. Sternal fusion was scored on a three-point scale from complete fusion. They found that manubrio-sternal joint fusion was not significantly correlated with age, but sterno-xiphoidal fusion was significantly correlated with age. When combined with the scores for the ribs in a composite score, the method could generate an “acceptable” age range estimate (Monum *et al.*, 2020). It should be noted, however, that the age estimate intervals generated by Monum *et al.* (2020) are very wide, which limits applicability in a forensic context. Overall, the fusion of these sternal junctions is highly variable and

inadequate for adult age estimation on its own (Chandrakanth *et al.*, 2012; Bacci *et al.*, 2018).

2.2.2 Epiphyseal fusion

Epiphyseal fusion is noted in various areas of the body and can contribute to age estimation by indicating whether an individual is past a certain age threshold. The progression and timing of these changes are well documented and can provide accurate age estimates for subadults and adults in some cases (Cunningham *et al.*, 2016).

The vertebral body epiphyses, or annular epiphyses, typically start fusion between the ages of 12-16 years and complete fusion around the age of 18 years but may take up to the age of 24 years to fully fuse (Cunningham *et al.*, 2016). Before epiphyseal fusion, the vertebral body has a very recognisable billowed texture typical of an unfused epiphyseal surface (Cunningham *et al.*, 2016). If an individual has fully unfused vertebrae, it is reasonable to estimate an age of 16 years or younger. If fusion has begun but is not complete the individual is likely between the ages of 12-24 years, and if fusion is complete then the individual is likely to be at least 18 years old (Cunningham *et al.*, 2016).

McKern and Stewart (1957) noted an association between sacral vertebrae fusion and age and stated that S1-2 fusion is typically completed between the ages of 23 and 32 years. Ríos *et al.* (2008) scored the fusion of sacral vertebrae on a five-point scale running from non-fusion to complete fusion, using McKern and Stewart's (1957) scoring criteria. They found significant correlations between age and sacral element fusion but only estimated age into four groups, with 20 and 50 years being the lower and upper bounds of age estimates (Ríos *et al.*, 2008). Mahon *et al.* (2018a) scored S1-2 fusion on a three-point scale of unfused, partially fused, and complete fusion in a black South African population. Significant correlations existed between the trait stages and age; non-fusion was typically found in individuals of 22 years and younger, partial fusion was seen in individuals of 30 years and younger, and complete fusion was found in individuals of 35 years and older (Mahon *et al.*, 2018a). Overall, there is an association between sacral vertebrae fusion and age, but the age ranges estimated are limited so it is recommended to use additional estimation methods in a forensic context (McKern and Stewart, 1957; Ríos *et al.*, 2008; Mahon *et al.*, 2018a).

In the upper limb, the medial end of the clavicle is used for age estimation based on multiple types of age-related events. The most common approach is an assessment of the degree of fusion of the medial epiphysis as it is one of the last epiphyses to close (McKern and Stewart, 1957; Black and Scheuer, 1996). McKern and Stewart (1957) proposed a scale for epiphyseal fusion with five stages running from no union to complete fusion. They found that union was not complete until the age of 23 years, and all individuals in the samples had completed union by the age of 30 years. Black and Scheuer (1996) indicate the lack of fusion until 14-17 years, the beginning of fusion from 15-22 years, full coverage of the epiphysis by 23-28 years, and obliteration of the line of fusion by 25-29 years. They also indicate that until fusion is essentially complete it cannot be a reliable indicator of age due to variation during puberty. However, once growth-related changes have stopped the medial end of the clavicle is a good indicator of age until around the age of 30 years (Black and Scheuer, 1996).

Langley-Shirley and Jantz (2010) developed a new method for age estimation from the fusion of the medial clavicle using Bayesian statistics, the McKern and Stewart (1957) scoring system, and a modified scoring system with McKern and Stewart's traits two through four (partial fusion, total fusion with remnant line, and complete fusion) combined into a single phase. The three-phase system had lower observer error, indicating less subjectivity and increased ease of use, but the five-phase system provided more precise age estimates (Langley-Shirley and Jantz, 2010). Mahon *et al.* (2018b) used a three-phase scoring system of unfused, partially fused, and complete fusion for the medial clavicle and found a significant correlation between trait score and age. Non-fusion was typically observed in individuals up to the age of 25 years, partial fusion was seen in individuals aged 23 to 29 years, and complete fusion was typically in individuals aged 30 years and older. They also found moderate to good observer error, with most errors occurring when distinguishing between partial and complete fusion (Mahon *et al.*, 2018b).

The fusion of the iliac crest is associated with age, much like other epiphyseal unions (Cunningham *et al.*, 2016). Webb and Suchey (1985) assessed the fusion of the iliac crest epiphysis in an American sample and described four stages of fusion. The first stage, non-union without an epiphysis, was observed up to the ages of 16 years in males and 11 years in females. Stage two was non-union with an epiphysis present and was observed in the age range of 13 to 19 years in males and 14 to 15 in females.

Partial union, stage three, was found in males and females between the ages of 14 and 23. Complete fusion was seen as young as 17 years in males and 18 years in females, with 100% of the sample exhibiting complete fusion above the age of 24 years. Wittschieber *et al.* (2013) conducted a similar study on the fusion of the iliac crest epiphysis on radiographs. They observed a similar pattern of epiphyseal fusion with advancing age. There was, however, a slight delay of three to four years in the female sample for the two categories of non-fusion. The age at which 100% of the sample had completely fused iliac crest epiphyses was 31 years, and the youngest age at which complete fusion occurred was 16 years (Wittschieber *et al.*, 2013). The fusion of the iliac crest can inform a forensic age estimate by distinguishing between sub-adults and young adults.

2.2.3 Phase analysis

Phase analysis is based on the concept that the appearance of a trait changes unilaterally as age progresses, with distinct visual appearances in specific age cohorts. The most commonly assessed bony regions for phasic changes include the pubic symphysis, the auricular surface of the ilium, and the sternal end of the fourth rib.

The most well-known age estimation methods involving the innominate (terminology according to Kiales (2020)), assess the pubic symphysis and auricular surface of the ilium. Todd (1920) developed a ten-phase scoring method for the full pubic symphysis, and McKern and Stewart (1957) developed a three-component system for the pubic symphysis with five stages for each component. McKern and Stewart's (1957) method only applies to males; however, Gilbert and McKern (1973) developed a corresponding method for females. Katz and Suchey (1986) evaluated both the Todd (1920) and McKern and Stewart (1957) methods and found that Todd's method systematically overages individuals. Neither method accounted for the full range of variation between individuals nor the full range of ages encountered in the context of forensic anthropological investigations, and the sample sizes were inadequate (Katz and Suchey, 1986). They recommended against a three-component system and suggested a modified version of the Todd (1920) method using six phases instead of ten. Brooks and Suchey (1990) further developed the phasic component-based method with much larger samples than Todd (1920) or McKern and Stewart (1957). The estimates generated were quite wide, and they warned that separate standards are needed for each sex due to parturition-related alterations (Brooks and Suchey,

1990). Hartnett (2010a) modified the Suchey-Brooks (1990) method by adding a seventh phase representing more advanced age and adapted the age ranges for each phase to better correspond with a contemporary sample. Currently, the Suchey-Brooks (1990) method is the most commonly used method for age estimation from the pubic symphysis (Garvin and Passalacqua, 2012). This method is also used in the Human Variation and Identification Research Unit (HVIRU) forensic anthropology laboratory as opposed to the Hartnett (2010a) method.

The auricular surface of the ilium was first suggested as a potential site for age estimation by Lovejoy *et al.* (1985) with the reasoning that it is unaffected by sex, population affinity, or activity. Unfortunately, Lovejoy *et al.*'s (1985) method is complicated to apply, and its five-year intervals are unrealistic in a forensic context (Buckberry and Chamberlain, 2002; Osborne *et al.*, 2004; White and Folkens, 2005). Buckberry and Chamberlain (2002) developed a component-based system based on the descriptions by Lovejoy *et al.* (1985), identifying five good indicators. It should be noted, however, that their sample population was archaeological and so may be of limited use in forensic circumstances due to secular changes in ageing rates. Osborne *et al.* (2004) also attempted to improve Lovejoy *et al.*'s method (1985) by altering the proposed stages and redoing the statistics, but they still yielded large age intervals.

Age estimation from the sternal ends of the ribs is well-researched, with the most common technique utilising three components of the sternal end of the fourth rib and eight phases (İşcan *et al.*, 1984, 1985). The components assessed are the pit depth, shape, and margins. The sternal rib end progresses from a shallow amorphous pit with regular or scalloped margins to a wide, deep pit with thin and irregular margins. Sometimes ossifications extend beyond the rib margin, either from the pit floor or from the margin itself (İşcan *et al.*, 1984; Russell *et al.*, 1993; Oettlé and Steyn, 2000; İşcan and Steyn, 2013b). Hartnett (2010b) slightly modified the İşcan *et al.* (1984) method for the fourth rib by accounting for bone density and quality and using seven phases as opposed to İşcan *et al.*'s (1984) eight phases. Hartnett's (2010b) seven phases span from age 18 to 97 years, while İşcan *et al.*'s (1984) eight phases span from age 17 to 85. Hartnett's (2010b) estimated intervals are generally narrower than İşcan *et al.*'s (1984), increasing the method's usefulness in a forensic context. Changes to the sternal rib ends have been strongly associated with age, but the calculated age intervals are very wide and significant sex- and population-specific variation exists

(İşcan *et al.*, 1984; Russell *et al.*, 1993; Oettlé and Steyn, 2000; Hartnett, 2010b; İşcan and Steyn, 2013b).

Kunos *et al.* (1999) developed an age estimation method from the first rib; however, the age estimates generated are singular point estimates with no error estimates or ranges given, which is not practical in a forensic context. Kurki (2005) tested Kunos *et al.*'s (1999) method and found low accuracy and limited applicability in a forensic context. DiGangi *et al.* (2009) developed a new method for age estimation from the first rib via a modification of Kunos *et al.*'s (1999) method in combination with Bayesian statistics. They provided a list of eleven traits, each with a scoring guide for identifying transition phases. They conducted an analysis with a combination of the two traits with the lowest correlation to each other (geometric shape of the costal face and tubercle facet texture). They also provided much of the raw statistics and computational data, enabling future researchers to select a different combination of traits if needed (DiGangi *et al.*, 2009). DiGangi *et al.*'s (2009) age estimates extended up to the age of 89 years, but age estimate intervals were very wide, frequently spanning multiple decades. They recommend using this method in combination with other methods using traits independent of the ribs.

Jooste and Steyn (2023) developed a regression-based method using three traits on the first rib (costal margins of face, rib head overall shape and edges of margins, and tubercle facet overall shape and edges of margins) modified from DiGangi *et al.*'s (2009) traits. They found significant differences between males and females and indicated that less than 50% of the change in these traits was related to age. They also observed consistent overestimation of age for the combined and female samples but an average bias of zero for the male sample. Age estimate accuracy based on 95% confidence intervals was low, between 11% and 33%, possibly due to the narrow age intervals of 8-13 years they used (Jooste and Steyn, 2023).

2.2.4 Degenerative changes

As the body ages, there are hormonal changes, a breakdown in homeostatic functions, and the effects of mechanical wear and tear build-up. There is usually an advancement in pathological conditions, such as DISH, osteoporosis, various haematopoietic conditions, and various kinds of arthritis (Maat *et al.*, 1995; Brennaman, 2011; Cawley and Paine, 2015; Curate *et al.*, 2019). Hormonal changes associated with the onset of

menopause can cause osteopenia or osteoporosis, and long-term repetitive movement can cause arthritic changes in joints (Jurmain, 1977, 1980; Riggs *et al.*, 1981; Burr *et al.*, 1983; Liao *et al.*, 2002; Listi and Manhein, 2012). These degenerative changes can affect any bone and joint surface in the body.

An uncommon pathological indicator of biological age is the thinning of the parietal bone, causing depressions in the cranial vault. This rare occurrence only appears in 0.25-0.8% of the population, but it is more common in females and almost exclusively found in the elderly (Dengsong *et al.*, 1988; Yiu Luk *et al.*, 2010). Lillie *et al.* (2016) conducted a study on CT scans of individuals aged 20 to 99 years, and they found a significant decrease in cortical thickness of the parietal, occipital, and frontal bones in females. The decrease in cortical thickness occurred in both the inner and outer tables of the skull and was associated with increasing age (Lillie *et al.*, 2016). It is possible that this thinning of the parietal bone, usually between the sagittal suture and parietal prominence, may be associated with the decrease in oestrogen in menopause and subsequent bone resorption (Dengsong *et al.*, 1988; Yiu Luk *et al.*, 2010; Lillie *et al.*, 2016). There are no guidelines for estimating age based on skull thickness, especially since the general trend is an increase in thickness associated with age (Lillie *et al.*, 2016). If an individual presents with parietal thinning, it may be reasonable to estimate that they are an elderly adult if the thinning is related to bone resorption.

Falys and Prangle (2015) developed an age estimation method based on degenerative changes to the medial surface of the clavicle, finding a strong correlation between surface topography and age. This method was tested by Blatt *et al.* (2020), looking at the surface topography, porosity, and osteophyte formation. They found the strongest correlation between topography and age, followed by porosity and age, and osteophytes and age. Overall, they found wide age ranges and a tendency to overage individuals over 40 years. They recommended further research with larger, more diverse samples, as well as the use of a Bayesian statistical approach (Blatt *et al.*, 2020)

Rissech *et al.* (2006) developed an age estimation method from the male acetabulum, assessing seven traits on the rim, articular surface, and fossa with four to seven stages per trait and age intervals of five years. They reported 89% accuracy within 10-year intervals, good repeatability, and good performance in individuals above the age of 60

years. Overall, they were optimistic that the method would be a potential solution to the common problem of older individuals being lumped into categories of 60+ years (Rissech *et al.*, 2006). Rissech *et al.*'s (2006) method was tested in Western European (Rissech *et al.*, 2007), Canadian (Calce and Rogers, 2011), South African (Botha *et al.*, 2016), European-American (Winburn, 2019), and Colombian (Muñoz-Silva *et al.*, 2020) samples. Common findings were low repeatability in porosity or texture-based traits and lower accuracy in populations more distant from the original population used in development. A common comment made was that the Rissech *et al.* (2006) trait definitions were too ambiguous, which may have led to the low repeatability observed (Calce and Rogers, 2011; Calce, 2012; Botha *et al.*, 2016; Navega *et al.*, 2018b).

Calce (2012) developed a new method for age estimation from the acetabulum in white American males using only three traits from the original Rissech (2006) method: the acetabulum groove, rim porosity, and apex activity. They used three very large age intervals comprising young adults, middle adults, and old adults, thereby re-introducing the problem Rissech *et al.* (2006) attempted to solve. The 'improved' method showed good repeatability and 81% accuracy at the cost of precision (Calce, 2012). When Navega *et al.* (2018b) tested Calce's (2012) method in a Portuguese population, they reported low repeatability and low accuracy (48.3%). They also found considerable overlap between the middle and old adult age groups, specifically with middle-aged adults frequently being over-aged. Navega *et al.* (2018b) suggested that Calce's (2012) attempt to simplify and clarify Rissech *et al.*'s (2006) complicated and unclear method may have gone too far and oversimplified it instead.

San-Millán *et al.* (2017) also revised Rissech *et al.*'s (2006) method in another Portuguese sample by redefining traits one through four (acetabular groove, rim shape, rim porosity, apex activity) and replacing traits five through seven (fossa outer edge activity, fossa activity, fossa porosities). They also used five-year age intervals and reported 74% accuracy within a 10-year age interval. Additionally, they developed the method for females in addition to males and recommended the use of sex-specific reference populations when possible (San-Millán *et al.*, 2017). When San-Millán *et al.*'s (2017) method was tested in a white American sample, they reported high repeatability and significantly higher accuracy than attained with the original method (Rissech *et al.*, 2006). Accuracy was lower than reported in the sample used in the development of the method (San-Millán *et al.*, 2019). The overall consensus on the

use of the acetabulum for age estimation is that it shows promise for the age estimation of older individuals. However, significant differences exist in the rate of ageing between males and females, as well as between geographically distant populations (Rissech *et al.*, 2007; Calce and Rogers, 2011; Botha *et al.*, 2016; Navega *et al.*, 2018b; San-Millán and Rissech, 2023).

A common indicator of old age is the development of osteoarthritis and osteophytosis, particularly on the vertebrae. Listi and Manhein (2012) found a significant association between osteoarthritis and osteophyte formation and increasing age; however, the association was not strong enough to provide an age prediction more accurate than a general estimate (Listi and Manhein, 2012). This association between age and osteophytosis and osteoarthritis is generally acknowledged in the literature, as is the high variation in age at onset (Jurmain, 1977, 1980; Garvin *et al.*, 2012; Rodriguez-Fontenla *et al.*, 2012; İşcan and Steyn, 2013b; MacDonald *et al.*, 2014; Langley *et al.*, 2017).

Typical age estimation methods involving the vertebrae look at osteophyte formation on the vertebral bodies (also known as osteophytosis) and osteoarthritic changes to the zygapophyseal joints. Stewart (1958) developed a five-point scoring system for vertebral osteophytes on both the inferior and superior borders of the vertebrae, which has been tested by a variety of researchers (Snodgrass, 2004; Van der Merwe *et al.*, 2006; Listi and Manhein, 2012). Watanabe and Terazawa (2006) developed an alternate four-point scale based on the dimensions of osteophytes on the superior and inferior margins of the cervical, thoracic, and lumbar vertebrae. These scores were then averaged for each region, and regression formulas were developed to estimate age (Watanabe and Terazawa, 2006). Praneatpolgrang *et al.* (2019) further adapted the methods put forward by Snodgrass (2004) and Watanabe and Terazawa (2006) in a Thai population. They developed a six-point scale based on specific measurements of vertebral osteophytes and provided regression equations for each vertebra from C2-L5 as well as for each spinal region (cervical, thoracic, lumbar) (Praneatpolgrang *et al.*, 2019). The presence of vertebral osteophytes may indicate that an individual is at least 50 years old if not older (Watanabe and Terazawa, 2006; Praneatpolgrang *et al.*, 2019).

Vertebral osteophytosis is usually associated with damage and degradation of the intervertebral disk as a reaction to particles in a typically sealed space (Prescher, 1998; Burt, 2013). The reaction can then cause degradation and erosion of the vertebral body's subchondral layer. Another theory is that osteophyte development results from a periosteal reaction to strain on the attachment for the anterior longitudinal ligament (Collins, 1949, as cited by Prescher, 1998). It is important to note that osteophytes occur below the anterior longitudinal ligament and are not the result of the ossification of said ligament (Prescher, 1998). Osteophytes develop adjacent to the edge of the vertebral body and may potentially meet osteophytes from the adjacent vertebrae, thereby bridging the intervertebral space. These "kissing" osteophytes may fuse and cause ankylosis of the joint (Maat *et al.*, 1995; Prescher, 1998; Burt, 2013). If the osteophyte bridges extend across four or more vertebrae, it can be mistaken for Diffuse Idiopathic Skeletal Hyperostosis (DISH).

In actuality, DISH results from the ossification of the anterior longitudinal spinal ligament and not from damage to the intervertebral disk as in osteophytosis (Belanger and Rowe, 2001; Burt, 2013). DISH typically involves four or more consecutive vertebrae with ossification occurring on the right antero-lateral aspect of the vertebral bodies. Ossification on the left lateral aspect is impeded by the abdominal aorta on the left of the vertebral column (Belanger and Rowe, 2001; Burt, 2013). DISH may also cause ligament ossification throughout the skeleton, in both the axial and appendicular skeleton (Belanger and Rowe, 2001; Burt, 2013).

Osteoarthritis is a frequent cause of bony degeneration throughout the human skeleton. Arthritic changes can happen at any joint surface, including the occipital condyles, trapezium articular facets, the head of the femur, and the vertebral zygapophyseal joints, amongst many others. On the femur head, arthritic osteophytes are commonly associated with age, and the fovea can experience marginal lipping associated with age progression (Burt, 2013; Villotte and Knüsel, 2013). Fusion of the sacroiliac joint can also be associated with age due to osteophyte growth across the joint (Burt, 2013). McKern and Stewart (1957) found an association between age and lipping of the acromion and glenoid fossa of the scapula, which is supported by the findings of Brennaman *et al.* (Brennaman, 2011; Brennaman *et al.*, 2017) and Zampetti *et al.* (2016).

Osteoarthritis is characterised by eburnation, marginal lipping, and osteophyte formation (Prescher, 1998; Burt, 2013). The theory is that it is caused by the combination of long-term wear and tear, joint structure (joints experiencing more movement will be more inclined towards osteoarthritic degradation), and disruption of chondritic homeostasis which leads to inflammation (Burt, 2013; Plomp *et al.*, 2015). The presence of arthritic changes to joints may indicate that an individual is at least middle-aged (Zampetti *et al.*, 2016; Brennaman *et al.*, 2017), and vertebral osteophytes typically indicate an individual over the age of 40 years but are most common after the age of 70 years (Burt, 2013). As previously mentioned, the development and progression of osteoarthritis is significantly associated with age progression. Unfortunately, that association is not strong enough to predict age-at-death, apart from providing an idea of whether the individual is middle-aged or old. The predictive power of osteoarthritis is further complicated by the onset of menopause, which affects bone growth and resorption (Listi and Manhein, 2012).

A less commonly studied class age of age-related degeneration is enthesal change, which is more frequently used for activity estimations in archaeological investigations (Weiss, 2007; Lieverse *et al.*, 2013; Palmer *et al.*, 2016). Bony changes at ligament attachment sites fall under the umbrella of enthesal changes and have an association with age, activity, and trauma (Benjamin *et al.*, 2006; Shaw and Benjamin, 2007; Villotte and Knüsel, 2013; Villotte *et al.*, 2016). These changes can include resorption of the bone at the insertion point and the development of bony spurs, exostoses, and alterations to rim profiles (Villotte and Knüsel, 2013). Any bony surface at which ligaments or tendons attach can experience these changes. Examples of such attachment sites in the upper limb include both tubercles and epicondyles of the humerus, the radial tuberosity, and the ulna olecranon process (Alves Cardoso and Henderson, 2010; Abrahams *et al.*, 2013). In the lower limb, the greater trochanter, trochanteric fossa, and trochanteric medial fossa of the femur are all important muscle insertion points, as is the distal fibula (Abrahams *et al.*, 2013). Affected sites in the human innominate include the ischial tuberosity, the anterior inferior iliac spine and the iliac crest tubercle, which are all attachment sites for ligaments (Abrahams *et al.*, 2013).

Many enthesal changes on the upper limb are related to activity patterns (Villotte *et al.*, 2010; Ibáñez-Gimeno *et al.*, 2013) as well as age, so the presence of

enthesophytes on the humerus, ulna, and radius are indicative of being at least middle-aged, but activity is a confounding factor. Enthesophytes on the femur and innominate are less affected by activity, so they are potentially more informative for age estimation (Weiss, 2004; Campanacho and Santos, 2013; Listi, 2016). Listi (2016) suggested that if severe enthesal activity is seen on the innominate and femur, the individual is likely at least 60 years old.

2.2.5 Other methods

There are a variety of other methods of estimation, including dental analysis (Lamendin *et al.*, 1992; Prince *et al.*, 2008), histology (García-Donas *et al.*, 2016; Maggio and Franklin, 2019), bone density assessment (Botha *et al.*, 2019; Curate *et al.*, 2022), the two-step method (Baccino *et al.*, 2014), and many others. These accurate and precise but sometimes destructive methods frequently require specialised equipment and technical expertise. Training, equipment, and reagents are costly, limiting the capabilities of poorly funded forensic facilities.

The teeth have been used for age estimation in various ways, some of which are destructive and some are not. Gustafson and Malmö (1950) developed a scoring system for longitudinally ground tooth sections assessing apex attrition, secondary dentin, periodontosis, cementum apposition, and root resorption. They reported high accuracy with an error of only 3.6 years (Gustafson and Malmö, 1950). Further research by Maples and Rice (1979), Lucy and Pollard (1995), and Kumar *et al.* (2015) questioned Gustafson and Malmö's (1950) regressions and error values, highlighting the larger-than-reported error in Gustafson and Malmö's (1950) method and its limited applicability in a forensic context.

Stott *et al.* (1982) developed an age estimation method that counts the number of tooth cementum annulation line pairs on a tooth root section and then adds the average age of eruption for the given tooth. Other studies have tested this method and have found high precision and accuracy, as well as suggesting that it performs notably better than Lamendin *et al.*'s (1992) method (Wittwer-Backofen *et al.*, 2004; Aggarwal *et al.*, 2008; Meini *et al.*, 2008; Le Cabec *et al.*, 2019; Ameena *et al.*, 2021; Pradeep *et al.*, 2021). A disadvantage of the Stott (1982) method is its destructive nature which may limit its useability in a practical forensic setting.

Lamendin *et al.* (1992) developed a non-destructive method utilising root transparency and periodontic heights compared to total root height. They found an accuracy of 46% for individuals below the age of 40 years and 90% for individuals above the age of 40 years, as well as a substantially lower minimum error than Gustafson and Malmö's (1950) method. Lamendin *et al.*'s (1992) method has been tested in the United States (Prince and Ubelaker, 2002), Colombia (González-Colmenares *et al.*, 2007), Austria (Meinl *et al.*, 2008), the Balkans (Prince and Konigsberg, 2008), and South Africa (Ackermann and Steyn, 2014). A common critique is that population-specific regression equations are required for higher accuracy. The method also consistently over-ages young adults and under-ages middle and old adults.

Other dental methods assess the pulp chamber proportion on radiographs (Kvaal *et al.*, 1995; Karkhanis *et al.*, 2014; Nayyar *et al.*, 2016), development of the third molar (De Luca *et al.*, 2016; Duangto *et al.*, 2017; Esan and Schepartz, 2019), amino acid racemization (Ohtani and Yamamoto, 2010; Wochna *et al.*, 2018), dental attrition (Smith, 1984; Lovejoy, 1985; Prince *et al.*, 2008; Hou *et al.*, 2019), and alveolar bone resorption (Mays, 2014).

The two-step method (Prodhan *et al.*, 2006; Baccino *et al.*, 2014) for age estimation combines two well-known age estimation methods sequentially. First, the Suchey-Brooks (1990) method for the pubic symphysis is applied. If the score is three or lower, the individual is classed as a young adult and age is estimated according to the Suchey-Brooks (1990) phases and interval. If the Suchey-Brooks score is four or higher, the individual is classed as an older adult and Lamendin *et al.*'s (1992) method for single-rooted teeth is used to estimate age. If the pubic bone is absent or damaged, İşcan *et al.*'s (1984, 1985) method for the fourth rib can be used as a substitute. When using the ribs, İşcan *et al.*'s (1984, 1985) phases zero through five are considered to be for young adults and phases six and higher are for older adults (Prodhan *et al.*, 2006; Baccino *et al.*, 2014). While the two-step method (Prodhan *et al.*, 2006; Baccino *et al.*, 2014) as a way of combining methods is advantageous for its simplicity, it is still subject to the same limitations as its component methods. An advantage of this two-step approach is the avoidance of the decrease in age estimation capability of the Suchey-Brooks (1990) method in older individuals and the poor age estimation capability of Lamendin *et al.*'s (1992) method in individuals below the age of 40 years.

Histological methods for age estimation include the assessment of osteon structure, dimensions, density, and layout in a variety of skeletal elements including the clavicle (Stout and Paine, 1992; Kranioti *et al.*, 2020), ribs (Stout and Paine, 1992; Crowder and Rosella, 2007; Kim *et al.*, 2007; García-Donas *et al.*, 2016; Goliath *et al.*, 2016; Dominguez *et al.*, 2020), humerus (Yoshino *et al.*, 1994), and femur (Lynnerup *et al.*, 2006; Chan *et al.*, 2007; Keough *et al.*, 2009; Martrille *et al.*, 2009; Goliath *et al.*, 2016; Botha *et al.*, 2018; Maggio and Franklin, 2019; Michener *et al.*, 2020).

A loss of bone density in relation to age is common, particularly in post-menopausal women. It affects the entire skeleton and may be related to osteoporosis, general mineral loss, and degenerative bone disease (Prescher, 1998; Agarwal and Grynpas, 2009; Burt, 2013; İşcan and Steyn, 2013b; Johannesdottir *et al.*, 2022). Osteoporosis is more commonly observed in women, with particular loss of trabecular bone in the spine (Foss and Byers, 1972; Riggs *et al.*, 1981). It also plays a part in the texture of bone due to the loss of cortical density and trabecular density, further causing pitting and erosion (Prescher, 1998; Anderson, 2003).

Osteoporosis can be assessed by examining the cortical index (Meulendyk *et al.*, 2011; Curate *et al.*, 2019), but it is most commonly diagnosed by assessing bone mineral density, which has a strong negative correlation with age (Kranioti *et al.*, 2019). Methods to assess bone mineral density in forensic anthropology include dual-energy X-ray absorptiometry (DXA), quantitative computed tomography (QCT) and Multi-Slice Computed Tomography (MSCT) of the lumbar vertebrae, femur, innominate, radius, and clavicle (Riggs *et al.*, 1981; Agarwal and Grynpas, 2009; Navega *et al.*, 2018a; Botha *et al.*, 2019; Dubourg *et al.*, 2020; Bascou *et al.*, 2021; Curate *et al.*, 2022; Johannesdottir *et al.*, 2022; Toutin *et al.*, 2022; Luo *et al.*, 2023). Botha *et al.* (2019) noted significant differences in the relationship between the bone mineral density of the femur and advancing age between black and white South Africans, indicating population-specific differences in the rate and pattern of the ageing process.

2.3 Transition Analysis

A different statistical approach, designed to reduce the effects of age mimicry as well as to combine information from multiple traits, is Transition Analysis as developed by Boldsen and colleagues (Boldsen *et al.*, 2002; Boldsen and Milner, 2016; Getz, 2020). This method uses population-specific data related to the timing of age-related changes

for each skeletal region assessed (age of transition), as well as prior distributions reflecting uniform or informed prior age-at-death structures (Boldsen *et al.*, 2002; Milner and Boldsen, 2012).

Transition Analysis calculates age based on transition curves and Bayesian prediction. The transition curves illustrate the distribution of ages-at-transition between trait phases within a sample. They are generated from a generalized linear model (logit) based on the data provided for each trait in a large reference sample. This is done by taking the natural log of the odds (Equation 2.3) of a trait being present given an individual is a specific age 'x' (Boldsen *et al.*, 2002).

The probability that an individual is a specific age (a) given a specific phase of a trait is present (y_j) can be represented as $p(a|y_j)$ (Equation 2.1). The odds of this (Equation 2.2) are equal to the probability of that event occurring (p) divided by the probability of the event not occurring (p-1).

$$\text{Equation 2.1} \quad \text{probability} = p = \frac{\text{event}_j}{\text{all events}} = p(a|y_j)$$

$$\text{Equation 2.2} \quad \text{odds} = \frac{p}{1-p} = \frac{p(a|y_j)}{1-p(a|y_j)}$$

The logit or natural log of the odds (Equation 2.3) models the relationship between the independent (a) and dependent (y_j) variables. It is equal to the sum of a constant (β_0), the product of the predictor variable (a) and an estimated coefficient (β_1), and an error (ϵ). It should be noted that β_1 is a non-linearly distributed log-likelihood, while the logit of p is linearly related to the predictor variable (a). The estimated coefficient (β_1) and the constant (β_0) are calculated with iterative maximum likelihood estimations (Equation 2.4).

$$\text{Equation 2.3} \quad \text{logit}(p) = \ln(\text{odds}) = \beta_0 + \beta_1 a + \epsilon$$

$$\text{Equation 2.4} \quad \text{likelihood}(\beta_0, \beta_1) = \frac{\prod_{j=1}^n \exp[(\beta_0 + \beta_1 a_j) y_j]}{\prod_{j=1}^n [1 + \exp(\beta_0 + \beta_1 a_j)]}$$

The maximum likelihood coefficients from that calculation are used to create a cumulative logit (Equation 2.5) that represents the probability of all trait phases ($y_j \geq i$) given all ages (a_j). The cumulative logit is converted into a probability density function

(Equation 2.6) with a forward continuation model via Bayesian inversion (Baldsen *et al.*, 2002).

$$\text{Equation 2.5} \quad \text{cumulative logit} = p(y_j \geq i | a_j) = \frac{\exp(\beta_0 + \beta_1 a_i)}{1 + \exp(\beta_0 + \beta_1 a_i)}$$

$$\text{Equation 2.6} \quad \begin{array}{l} \text{probability density function} \\ \text{for age conditioned on a trait} \end{array} = p(a | y_j = i) = \frac{p(y_j = i | a)}{\int_0^\infty p(y_j = i | a)} da$$

This probability density function is proportional to the log-likelihood function (Equation 2.7) for age given the presence of the trait.

$$\text{Equation 2.7} \quad \text{likelihood}(a | y_j = i) \propto p(a | y_j = i)$$

The likelihood ratio test statistic has the same distribution as a χ^2 distribution with one degree of freedom (DF), for which the critical value is 3.841. The maximum likelihood value of this probability distribution is the point estimate for age, and the 95% bounds are calculated by taking the intercepts of a line 1.9205 units below the peak of the log-likelihood function and the log-likelihood function itself (Getz, 2017).

A complete set of traits is not required for an estimate because the cumulative log-likelihood function can be calculated from any number of traits, unlike many regression methods that require a complete set of specified values. Currently, there is no way to account for sex and population specificity since the only available prior distribution is uniform. There is also no statistical correction for correlated traits at the current time (Getz, 2017; Getz and Byrnes, 2021).

Godde and Hens (2012) compared the statistical method of transition analysis to the Suchey-Brooks (1990) method in an early 20th-century Italian sample. They found that transition analysis outperformed the Suchey-Brooks (1990) method, irrespective of whether an informed or uniform prior distribution was used. The estimates generated with the uniform prior distribution were less accurate than those generated with the Terry and Balkan prior distributions. However, the Suchey-Brooks (1990) estimates were the least accurate (Godde and Hens, 2012). They suggested that without a suitable prior, Bayesian analysis with a uniform prior is a better alternative than a method known to exhibit age mimicry. In addition, the rationale behind transition analysis can be applied to other well-known age estimation methods, potentially increasing both accuracy and precision.

Initially, Transition Analysis 2 (TA2) made use of 33 traits (accounting for bilateral traits) in three skeletal regions, which included the cranial sutures, the pubic symphysis, and the auricular surface of the ilium (Boldsen and Milner, 2016). For TA2, Boldsen and colleagues (Boldsen *et al.*, 2002; Boldsen and Milner, 2012; 2016; Milner and Boldsen, 2012) developed a scoring system and computer software program called ADBOU. ADBOU generates an age estimate based on the degree of morpho-macroscopic degeneration observable on the pubic symphysis, auricular surface, and cranial sutures. The method has an estimated accuracy of 95%, dependent on the number of traits for which a score is recorded and which sets of traits are utilised (Boldsen *et al.*, 2002).

The problem of combining traits that may be correlated with each other due to being in the same or adjacent regions is circumvented by assuming that any correlation between them is the result of age. Thus, the traits can be considered independent if conditioned on age (Boldsen *et al.*, 2002). Unfortunately, the assumption of no correlation between traits if conditioned on age is unrealistic in practice. Traits within direct proximity to each other will all be subjected to the same strains and forces and thus are likely to be at least weakly correlated. Multiple traits that are caused by a single activity will also be at least weakly correlated with each other. While all these traits' degrees of expression may progress as age progresses, the other processes affecting the related traits do not necessarily progress at the same rate. The result is that the assumption of trait independence when conditioned on age is not as effective as it appears, and inter-trait correlation may cause bias in the age estimation process.

TA2 made use of three different prior distributions drawn from the Terry Collection of the Smithsonian Institute, representing black and white North American individuals (the uniform distribution), a 17th-century Danish parish population (the archaeological distribution), and a selection of forensic cases (the forensic distribution) (Boldsen *et al.*, 2002; Milner and Boldsen, 2012).

Milner and Boldsen (2012) conducted a validation study for TA2 using a primarily white American sample from the William Bass collection and the Mercyhurst forensic cases. They found that when each of the three regions was used independently, the interval widths were high, especially for the cranial sutures. The cranial sutures severely underestimated the ages of individuals above the age of 40 years, while the auricular

surface did the same from the age of 60 years. A common problem encountered was the underestimation of the ages of individuals aged 70 years and above, irrespective of the prior distribution used (Milner and Boldsen, 2012).

Bullock *et al.* (2013) compared TA2 to well-known age estimation methods for the pubic symphysis and auricular surface in two historic Mexican samples. The cranial sutures generally under-aged individuals or massively over-aged them, causing an unrealistic clustering of individuals with estimated ages of 110 years. The pubic symphysis was relatively accurate until the breakdown phase began, after which it became less informative. The auricular surface followed a similar pattern but breakdown occurred much later in life than in the pubic symphysis, making the auricular surface a valuable site for the age estimation of older individuals (Bullock *et al.*, 2013). TA2 yielded notably more realistic age distributions for both populations as compared to “traditional methods”, particularly when the archaeological prior distribution was used. Overall, TA2 improved the estimation of age within the two samples. However, accuracy and precision were still lower than hoped for, and a persistent problem is the lack of genuinely applicable prior distributions (Bullock *et al.*, 2013).

Jooste and colleagues (2016) conducted a validation study for the use of TA2 in an adult black South African sample. They found that, while applicable, there were many limits to the use of this method in the South African sample. They observed a decrease in estimate accuracy after 40 years of age and generally very large age range estimates. This may be due to inter-population differences in ageing rates and patterns between the black South African population and the two sample populations used to develop TA2. Jooste and colleagues (2016) also found problems in the repeatability of scoring some traits like the ventral margin of the pubic symphysis and the symphyseal texture. They suggested that traits with low repeatability should be removed from the analysis and that not all traits were equally informative, such as the ectocranial sutures (Jooste *et al.*, 2016).

TA2 was compared to other cranial (Meindl and Lovejoy, 1985) and pelvic methods (Lovejoy *et al.*, 1985; Brooks and Suchey, 1990; Buckberry and Chamberlain, 2002) in a Greek population by Xanthopoulou *et al.* (2018). They found moderate to substantial inter- and intra-observer repeatability for all methods. The age estimates

yielded by TA2 were typically less accurate than the other methods tested, but the highest accuracy for TA2 occurred when all traits were used. The pubic symphysis was most accurate for males, the auricular surface was most accurate for females, and the lowest accuracy was when only the cranial traits were used. The correlations between TA2's traits and age were highest for the pubic symphysis and auricular surface and lowest for the cranial sutures. However, all TA2 traits had lower correlations with age than the traits for the other methods tested (Xanthopoulou *et al.*, May).

Lopez-Cerquera and Casallas (2019) tested TA2 in a modern Colombian sample population, finding a high correlation between actual and estimated age. The pubic symphyseal traits were the most accurate for older adolescents and young adults, but older adults tended to be overestimated. The auricular surface traits showed the same pattern, but the cranial traits were the least accurate, overestimating the ages of adolescents and underestimating the ages of the remaining individuals. A common problem for individuals of middle age or older was the extreme width of estimation intervals, both for the composite estimate and for each region's estimates (Lopez-Cerquera and Casallas, 2019).

Hagelthorn *et al.* (2019) tested TA2 on a sample of the coloured South African population and found that TA2 severely under-estimated the ages of individuals above the age of 50 years. None of the trait regions were strongly correlated with age, with the highest correlation strength being moderate. Strangely, the pubic symphysis had the lowest correlation with age, even though it was the least biased. The auricular surface was the most strongly correlated with age and had the second-lowest bias. The cranial sutures had the second-highest correlation with age and the highest bias. The method worked best for individuals aged 18-39 years, and the worst accuracy was found in individuals aged 60-89 years (Hagelthorn *et al.*, 2019).

Kim and Algee-Hewitt (2022) tested TA2 in three contemporary Asian samples (Japanese, Filipino, and Thai), finding that the cumulative maximum likelihood estimates using the archaeological distribution were the most accurate, followed by the uniform distribution and that estimates were more accurate for males than females. Under-estimation of age was a common problem for females. In contrast, the overall ages of male young adults and middle-aged adults were relatively accurate, but under-

estimation increased with age. They found that the maximum likelihood estimates had a bimodal distribution with the under-representation of middle-aged individuals. Japanese males and Filipino individuals were estimated to have a similar distribution to real life. However, significant differences in age-at-death distribution were seen in Japanese females and Thai individuals (Kim and Algee-Hewitt, 2022).

The findings of lower accuracy for female age estimates are, unfortunately, relatively common across many age estimation methodologies (Calce and Rogers, 2011; Listi and Manhein, 2012; Işcan and Steyn, 2013a; Botha *et al.*, 2016; Navega *et al.*, 2018b; Kim and Algee-Hewitt, 2022; Jooste and Steyn, 2023; Xanthopoulou *et al.*, May). The reason for this is uncertain but may be related to the fact that most methods were initially developed on an only male sample and that females are under-represented in research samples. This severe sex bias in development may inadvertently introduce bias into the method itself because it does not reflect realistic population demographics.

Common findings of studies testing TA2 are that the best age estimates were found when all regions were used but that the cranial sutures often negatively impacted the age estimates for older individuals. An advantage of TA2 was the ability to estimate age even when some traits were missing, and the more realistic age distributions generated when estimating with the uniform and archaeological priors. Common problems with TA2 include the lack of applicable prior distributions, relatively low accuracy, persistent underestimation of age, and the very wide 95% confidence intervals.

Given the limitations of TA2, a newer version is being developed. The new Transition Analysis 3.0 version is still in the Beta phase of development and requires extensive testing. In addition to the increased number of traits assessed, TA3 is developed from a total of eight different population samples, including a black South African sample from the Pretoria Bone Collection housed at the University of Pretoria in South Africa (Getz, 2017, 2020; Boldsen *et al.*, 2021a). Some problematic traits from TA2 such as the cranial sutures have been removed, and numerous other traits such as fibula wings and vertebral epiphyseal ring fusion (see Appendix C) have been added to TA3 (Boldsen *et al.*, 2002, 2021a; Boldsen and Milner, 2016). These new traits were added because when Boldsen and Milner (2012) conducted a validation study of TA2, they

noted that experience-based estimates fared better than TA2's estimates. This indicated that age estimations can be informed by many more traits than the standard methods use and that other "low information" traits may be helpful in aggregate.

TA3 utilises 141 traits (accounting for bilateral traits) in eight regions (the cranium, the vertebrae, the ribs and sternum, the upper limb, the lower limb, the pubic symphysis, the auricular surface of the ilium, and the rest of the innominate) (Boldsen *et al.*, 2021a). Both TA2 and TA3's scoring is based on an age indicator's morphological features and its transition between different phases as an individual ages (Boldsen *et al.*, 2002; Getz, 2020). Ideally, these indicators should exhibit distinct and unilateral morphological or textural changes in a phasic fashion, with phases transitioning at known temporal intervals (Boldsen *et al.*, 2002). TA3 currently only uses a single (uniform) prior distribution, with transition curves constructed from a much larger sample than TA2 (Boldsen *et al.*, 2002; Milner and Boldsen, 2012; Getz, 2020).

Due to the expansion in the scope of traits in TA3 vs TA2, the increased number of sample populations, and the use of an exclusively uniform prior as opposed to a choice between two informed priors or a uniform prior, it is vital for TA3 to undergo validation before its use in the South African context. The results of this validation study may inform the suitability of TA3 in the South African forensic context concerning its accuracy and repeatability.

Chapter 3 Materials and methods

3.1 Ethics

Approval for inclusion under the School of Anatomical Sciences' Human Research Ethics Committee (Medical) Clearance Certificate: W-CBP-220504-01 was granted by the Head of School of the School of Anatomical Sciences, Faculty of Health Sciences (Appendix A).

3.2 Materials

3.2.1 Sample

The sample was selected from the Raymond A. Dart Collection of Modern Human Skeletons housed at the University of the Witwatersrand. It consisted of the skeletal remains of 89 black South African individuals of known age and sex who died between 1926 and 1992 and between the ages of 15 and 105 years. This is the same age range used in the development of TA3 (Getz, 2017, 2020). The current study sample was equally distributed for sex and per age cohorts as far as possible (Table 3.1). The age spread of the sample was selected to ensure the representation of all age groups incorporated in the developmental dataset for TA3 (Getz, 2017, 2020). Only black South African individuals formed the sample because they are the largest population group in South Africa (Statistics South Africa, 2021) and they are the population group most frequently victim to crime in South Africa, thus making up the largest proportion of forensic anthropological cases (Statistics South Africa, 2019). Only the Raymond A. Dart Collection of Modern Human Skeletons was used because the Pretoria Bone Collection at the University of Pretoria was used to develop TA3.

Table 3.1 Sample size and distribution

Age cohort (years)	Males	Females	Total
15-24	8	8	16
25-34	6	6	12
35-44	5	5	10
45-54	5	5	10
55-64	5	5	10
65-74	5	5	10
75-84	5	5	10
85-94	4	3	7
95-105	2	2	4
Total	45	44	89

The remains of individuals who died before 1959 may be affected by commingling and water damage as a side effect of the flooding of the collection in 1959 (Dayal *et al.*,

2009). There may also be differences in lifestyle and environmental influences on the ageing process for individuals who died in the earlier half of the century compared to individuals who died in the latter half. It should be noted that their inclusion may have influenced the assessment of accuracy, sex-related differences, and asymmetry. The sample still included individuals from this period due to the scarcity of individuals who satisfied all the inclusion and exclusion criteria, particularly for the older age cohorts.

The higher sample size for the ages of 15 to 34 years and the lower sample size for the ages of 85 to 105 years result from limitations in the number of individuals who satisfied the inclusion and exclusion criteria. The cohorts of 75-84, 85-94, and 95-105 years were pooled for some analyses due to the small sample size impacting the statistical validity of the tests.

The sample consisted of both crania and the full post-crania because the traits assessed by TA3 are found throughout the human skeleton (Boldsen *et al.*, 2021a) instead of only the cranium and innominate like in the previous version of Transition Analysis (Boldsen and Milner, 2016). When a trait was bilateral, both sides were scored separately. When the programme had a single input for a bilateral trait, the left side was used, and the right side was used if the left was unavailable. Additionally, the mass/weight and maximum length were recorded for all elements for which a relative weight was recorded, even though the metric values were not utilized in the program. This was done to ensure that every single trait in the scoring manual (Boldsen *et al.*, 2021a) was assessed.

Individuals for whom less than 50% of traits were scoreable due to trauma, pathology, or post-mortem damage were excluded. The absence of two or more 'major' regions (e.g., entire upper limb, entire innominate) qualified as more than 50% absence. If the cranium, ribs, and vertebrae were all absent in addition to an entire major region then the individual was considered to be missing more than 50% of traits. As long as at least one side of a bilateral trait was present, the trait was considered present in the individual. Any forms of pathology or trauma that may have affected an age-related trait indirectly were also criteria for exclusion.

The 50% threshold was chosen to maximise the number of individuals who could make part of the sample as a whole while attempting to make as thorough an analysis as possible. Many sets of remains in the collection are damaged and incomplete because of maceration damage, research-related damage, and general long-term handling.

Requiring a higher percentage completion would have substantially decreased the available sample size.

Ideally, only complete skeletons would have been used, but that is unrealistic in the context of the collection for multiple reasons. There were very few elderly black individuals in the collection, and many individuals were subjected to some form of maceration damage. Given that the collection consists of individuals who died due to natural causes, it is not uncommon to find some evidence of pathology on their remains. An effort was made to include as many complete individuals as possible, but when that failed, individuals with missing skeletal elements were assessed provided they satisfied the above conditions. Individuals missing an entire region of traits had to be excluded from the analysis of TA3's ability to work with incomplete information.

Where measurements were necessary, sliding callipers, a Paleo-Tech field osteometric board, and a standard baking scale were used (trait-specific use is indicated in Appendix B). The element length and weight measurements were made once and redone if they seemed inconsistent with other element measurements, as the software did not utilise those values. Measurements of the extent of bony changes were made at least twice in all possible dimensions to ascertain whether the bony changes passed the specified thresholds.

3.3 Methods

TA3 calculates age based on transition curves, probability density functions, and Bayesian statistics in combination with a uniform prior distribution. The age estimate is the maximum likelihood value of a cumulative log-likelihood function generated from the selected combination of traits. The 95% prediction interval is based on the assumption that the log-likelihood function has an equivalent distribution to a χ^2 (DF = 1, $\alpha = 0.05$), where the 95% bounds are the points at which a horizontal line corresponding to half of the critical value intersects with the log-likelihood curve. A complete dataset is not required to provide an age estimate since the estimation curve is a composite function of as many or as few traits that are selected (Getz, 2017; Getz and Byrnes, 2021).

3.3.1 Data collection

Details of the scoring system can be found in Appendix B. The original scoring manual is by Boldsen *et al.* (2021a), and the only trait modified is the scoring site of the ischial tuberosity bumps due to anatomical incompatibilities. Appendix B also contains

photographs of the presentation of each trait's stages within this study's sample, with a few exceptions where a photograph was not obtainable. The sequence of skeletal elements assessed was as follows: cranium, femur, tibia, fibula, calcaneus, clavicle, scapula, humerus, radius, ulna, trapezium, innominate, vertebrae, and ribs. All scoring was done blindly, with the demographic details of each individual hidden from the observer. The left element was assessed first, followed by the right. For traits involving multiple joints, such as the ribs and vertebrae (C1-C7, T1-T12, L1-L5), traits were scored as present when apparent on at least one rib per side or at least one vertebral surface.

3.3.2 Repeatability

For intra- and inter-observer repeatability, a subset of ten complete skeletons representing five age cohorts (15-24 years, 25-34 years, 35-54 years, 55-74 years, and 75+) years from the original study sample were rescored by both the primary investigator and for the first time by a colleague. For the intra-observer repeatability testing, the subsample was rescored several weeks after the initial scoring to prevent researcher bias. The repeatability sample was not selected randomly but instead selected to provide a sufficient representation of the age cohorts. The associated statistics are in section 3.3.4, "Data analysis", under the subheading "Repeatability".

3.3.3 Age estimation using the TA3 software

Scores were individually inputted into the TA3 software (Boldsen *et al.*, 2021b), which then generated an estimated age (maximum likelihood value), upper- and lower-95% bounds, and the standard error for each individual. When a bilaterally scored trait only had a single applicable data entry point in the software, the score for the left side was used. If the left side was unavailable, the right side was used, and if one side was excluded from the analysis, then the other side's score was used. An individual did not have to have a complete data set to be analysed as long as it satisfied the exclusion criteria.

3.3.4 Data analysis

The software used for data collection, processing, and analysis were Microsoft Excel (Microsoft Corporation, 2023), R (R Core Team, 2022a), and TA3 (Boldsen *et al.*, 2021b). The R packages used included: abind (Plate and Heiberger, 2016), cocor (Diedenhofen, 2022), coin (Hothorn *et al.*, 2023), DescTools (Signorell, 2023), ggplot2 (Wickham *et al.*, 2023a), dplyr (Wickham *et al.*, 2023b), ggpubr (Kassambara, 2023), irr (Gamer *et al.*, 2019), openxlsx (Schauberger and Walker, 2023), pacman (Rinker

and Kurkiewicz, 2019), patchwork (Pedersen, 2023), readxl (Wickham and Bryan, 2023), reshape2 (Wickham, 2020), rio (Becker *et al.*, 2023), WRS (Wilcox and Schönbrodt, 2023a), WRSedit (Wilcox and Schönbrodt, 2023b), and other base packages for R (R Core Team, 2022b).

Repeatability

The repeatability of scoring of ordinal traits was assessed with Cohen's Weighted Kappa (Cohen, 1960), using linear weighting as suggested by Vanbelle (2016), and interpreted using the agreement scale put forward by Landis & Koch (1977) (Table 3.2). The repeatability of the measurement of numeric traits was assessed with Lin's Concordance Correlation Coefficient (Lin, 1989), with a confidence interval of 0.95, and interpreted with the agreement scale suggested by McBride (2005) (Table 3.3).

Table 3.2 Modified agreement scale for Cohen's Kappa from Landis and Koch (1977)

Kappa statistic	Strength of Agreement
0.81-1.00	Almost perfect
0.61-0.80	Substantial
0.41-0.60	Moderate
0.21-0.40	Fair
0.00-0.20	Slight
<0.00	Poor

Table 3.3 Agreement scale for Lin's CCC from McBride (2005)

Lin's CCC	Strength of Agreement
>0.99	Almost perfect
0.95-0.99	Substantial
0.90-0.95	Moderate
<0.90	Poor

Correlation of traits to true age and true age to estimated age

A Spearman's Rank Correlation Coefficient (Spearman, 1904) ($\alpha = 0.05$, $n = 89$) was used to assess the correlation between each trait (bilaterally where applicable) and actual age as recorded in the Raymond A. Dart Collection of Modern Human Skeletons catalogue, as well as for the correlation between true and estimated age. This was done for the pooled-sex sample and two subsets of males ($\alpha = 0.05$, $n = 45$) and females ($\alpha = 0.05$, $n = 44$) respectively. Significance was determined using the critical value table by Ramsey (1989).

The rho values were interpreted with guidelines by Chan (2003) (Table 3.4) for the strength of correlation, as suggested by Akoglu (2018).

Table 3.4 Modified agreement scale for a linear relationship from Chan (2003)

Correlation Coefficient value (r)	Strength of relationship
1	Perfect
≥ 0.8	Strong
$0.6 \leq r < 0.8$	Moderate
$0.3 \leq r < 0.6$	Fair
< 0.3	Poor
0	None

The correlation between true and estimated age was also done for eight sub-samples ($\alpha = 0.05$, $n = 68$), where all scores for a single region were removed from the analysis and the age was re-estimated.

Accuracy, inaccuracy, bias

The accuracy of the age ranges generated by TA3 was assessed by calculating the proportion of individuals whose true age fell within the estimated 95% confidence interval. This was done for the pooled-sex sample and male and female subsamples per age cohort, as well as for 75-105 years and the sample overall. The cohort of 75-105 years was created due to the low sample sizes in cohorts 85-94 and 95-105, which may have skewed results. Lin's Concordance Correlation Coefficient (Lin, 1989) was calculated to assess the agreement between the true age and point estimate age ($\alpha = 0.05$, $n = 89$) and interpreted using McBride's (2005) agreement scale (Table 3.3).

Bias (Equation 3.1) was calculated by taking the average of the difference between an individual's true age and the corresponding point-estimate age. This was done for the pooled-sex, male, and female samples per age cohort and 75-105 years.

$$\text{Equation 3.1} \quad \text{Bias} = \sum \frac{\text{Estimated age} - \text{True age}}{n}$$

Inaccuracy (Equation 3.2) was calculated by taking the average of the absolute difference between an individual's true age and the corresponding point-estimate age, as in Buckberry (2015). This was done for the pooled-sex, male, and female samples per age cohort and 75-105 years.

$$\text{Equation 3.2} \quad \text{Inaccuracy} = \sum \frac{|\text{Estimated age} - \text{True age}|}{n}$$

Traits which exhibited a poor correlation with true age were removed from the analysis, all ages were re-estimated, and the above was repeated for the new set of age

estimates. Different traits were removed for the sex-pooled, male, and female samples based on the specific traits that were poorly correlated for each sample. This was done to assess the effect of removing poorly correlated traits on age estimates' accuracy, bias, and inaccuracy.

Differences between males and females

The correlations between true age and trait for males vs females were compared using Fisher's Z transformation (Fisher, 1925) ($\alpha = 0.05$, $n(\text{male}) = 45$, $n(\text{female}) = 44$). The correlations used were drawn from the male and female subsamples mentioned above. Traits with undefined correlation coefficients due to a standard deviation of zero were excluded. The null hypothesis was that there was no significant difference in trait-age correlations between males and females.

The difference in trait scores for males and females was assessed with a Mann-Whitney U test (Mann and Whitney, 1947) with $\alpha = 0.05$. Traits for which a p-value could not be determined due to an over-abundance of ties were given the value of NA. The null hypothesis was that the distribution of trait scores was equal for males and females.

Differences between the left and right sides

The correlations between true age and trait for the right side versus the left side were compared using a percentile bootstrap on a Spearman's rho test ($\alpha = 0.05$, $n = 89$, $B = 2000$) with the "twoDcorR" function from a modification of the package WRS by Wilcox (Wilcox, 2016). Correlations were drawn from the pooled-sex dataset and the null hypothesis was that there was no significant difference between the age-trait correlations for the left and right sides.

The symmetry of trait scores for the left and right sides was assessed with a Wilcoxon signed rank test (Wilcoxon, 1945) for two paired samples with $\alpha = 0.05$. Traits for which a p-value could not be determined due to an over-abundance of ties were given the value of NA. The null hypothesis was that there was no significant difference between the scores of the left and right sides.

Sensitivity to missing data

To test TA3's ability to work with an incomplete dataset, individuals with adequate representation of all regions (at least 50% of traits in every region were scoreable)

were selected ($n = 68$). All individuals already missing regions were excluded from this analysis, as were individuals with only a single trait observed within a region.

The age of each individual in the identified sub-sample (pooled-sex, $n = 68$) was re-estimated eight times. Each time age was re-estimated, a new .ta3 file was created where all the traits from a single region (e.g., the entire lower limb) were removed from the original .ta3 file. The re-estimated ages were compiled in a single spreadsheet per region, and that dataset was used to calculate the new correlations between true and estimated age. The correlation between true and estimated age was also re-calculated for the sample of 68 individuals, as it would be a different value from the correlation previously calculated with 89 individuals.

The correlation between true and estimated age was calculated using Spearman's Ranked Correlation Coefficient ($\alpha = 0.05$, $n = 68$) for each subsample (nine in total). The ability of the method to work without a full dataset was assessed via a comparison of correlation coefficients for each combination of skeletal elements with a Fisher's Z transformation test ($\alpha = 0.05$, $n = 68$) (Fisher, 1925).

Chapter 4 Results

4.1 Repeatability

The only trait with a poor and negative intra-observer agreement was the parietal depression ($\kappa = -0.190$, $p = 0.389$). Three traits showed only “slight” intra-observer agreement: ilium anterior inferior iliac exostoses, pubic symphyseal collar, and occipital condyle lipping. Four traits presented with “fair” intra-observer agreement: the humerus’ lateral epicondyle exostoses, acetabulum posterior margin lipping, L1’s inferior surface morphology, and the sternum central dorsal ridge. Intra-observer agreement could not be assessed for the ischial tuberosity medial spur because there was no variation in the scoring, leading to a standard deviation of zero and an undefined kappa score. This statistical outcome may also reflect perfect agreement as there is no difference between observations. However, it is not possible to determine how the agreement compares to what would be expected by random chance. All other traits (eighty-three out of ninety-two) had an intra-observer agreement rating of “moderate” to “almost perfect”. Sixty-two of these traits were rated “substantial” to “almost perfect”, which indicates generally good intra-observer repeatability for the scoring method. Cohen’s Weighted Kappa and Lin’s Concordance Correlation Coefficient outputs and strength of agreement tables for the intra-observer repeatability can be found in Appendix C (Tables C.1 & C.2).

Inter-observer repeatability was lower than the intra-observer repeatability. Seven traits were rated to have “poor” agreement: the humerus relative weight, the acetabulum posterior margin lipping, the acetabulum posterior articular surface’s extra bone, L1’s inferior surface morphology, ribs 2-10 shingle ribs, the calcaneus’ relative density, and the calcaneus’ actual weight. There were nineteen traits with “slight” inter-observer agreement, many of which had a score of exactly zero, indicating no more agreement than random chance. There were fourteen traits with “fair” inter-observer agreement (Tables C.3 & C.4). Similar to the intra-observer repeatability for the ischial tuberosity of the medial spur, the inter-observer agreement could not be assessed because the output was undefined due to a standard deviation of zero. All other traits (fifty-one out of ninety-two) had an inter-observer agreement rating of “moderate” to “almost perfect”, of which thirty-three were rated “substantial” to “almost perfect”. The proportion of “moderate” to “almost perfect” traits vs “poor” to “fair” was much lower for the inter-observer repeatability, with only 55% of the traits showing adequate inter-

observer repeatability of “moderate” ($\kappa = 0.41 - 0.60$, $p = 0.001 - 0.107$) or higher repeatability. Cohen’s Weighted Kappa and Lin’s Concordance Correlation Coefficient outputs and strength of agreement tables for inter-observer repeatability can be found in Appendix C (Tables C.3 & C.4).

4.2 Correlation of trait to true age

4.2.1 Correlations

Pooled sexes sample

Of the 141 traits assessed (including bilateral traits) in the complete pooled sexes sample, 58 traits were poorly correlated with age or had an undefined correlation, many of which were element weight and length (Table C.5). The Spearman’s rho for ribs two through ten shingle ribs on the right side was undefined because all the scores were identical so no variation or standard deviation could be calculated.

No traits had a “strong” correlation with age, but thirteen traits had a “moderate” correlation with age (Table 4.1, Figs. 4.1 & 4.2). The scatter plots of the non-moderate strength traits can be found in Appendix C. Interestingly, all vertebral traits had at least “fair” strength, and some had a moderately strong correlation.

Table 4.1 Traits in the pooled-sex sample ($n = 89$) with “moderate” strength of correlation with age

Trait	Rho	p-value
Clavicle medial epiphysis fusion (L)	0.644	<0.001
Clavicle medial epiphysis fusion (R)	0.640	<0.001
Scapula glenoid fossa lipping (L)	0.611	<0.001
Scapula glenoid fossa lipping (R)	0.676	<0.001
Ischial tuberosity bumps (L)	0.604	<0.001
Pubic symphyseal relief (R)	0.669	<0.001
Pubic symphysis ventral margin (R)	0.644	<0.001
L1 spinous process flattening	0.657	<0.001
L5 superior margin shape	0.696	<0.001
S1-2 fusion	0.739	<0.001
T lipping	0.603	<0.001
L lipping	0.625	<0.001
L candlewax	0.625	<0.001

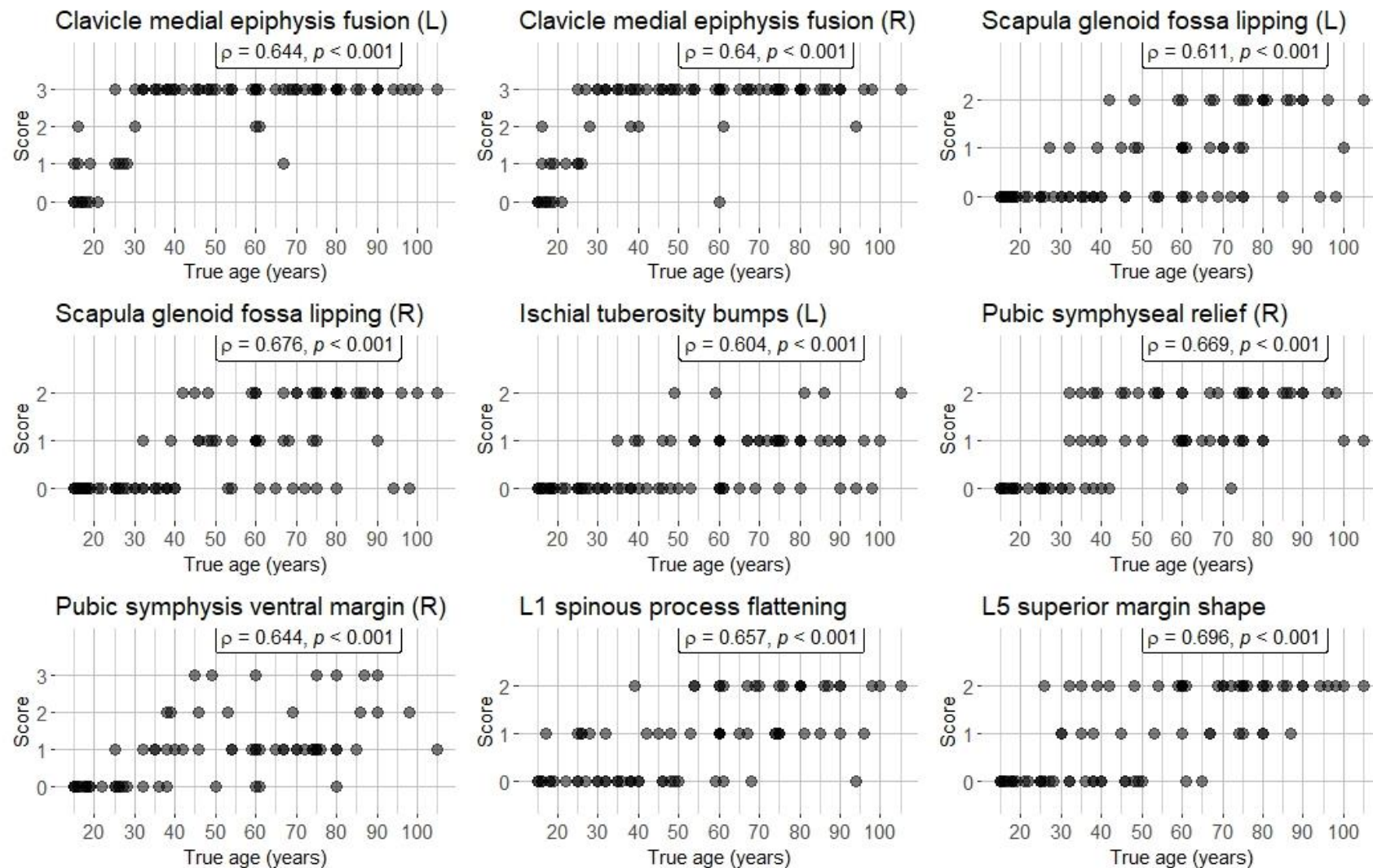


Fig. 4.1 Scatter plots for the traits with a moderate correlation with age in the pooled-sex sample ($n = 89$, $\alpha = 0.05$); part one

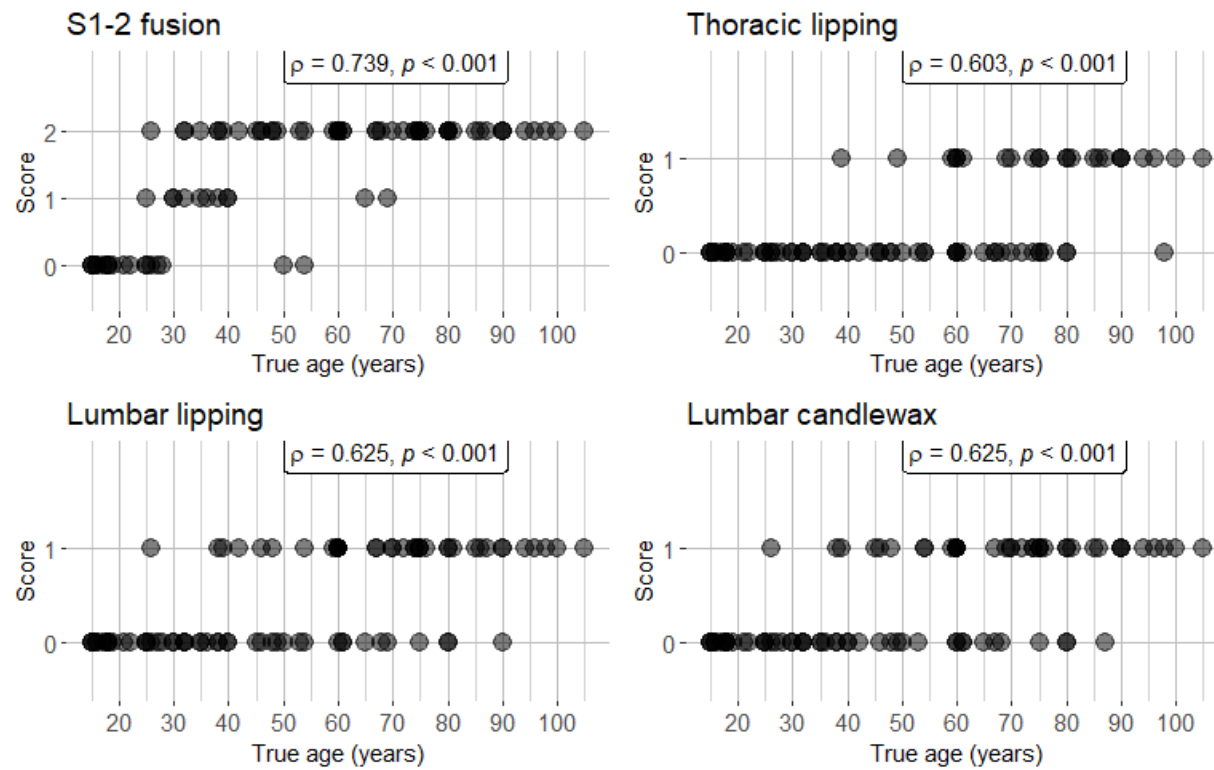


Fig. 4.2 Scatter plots for the traits with a moderate correlation with age in the pooled-sex sample ($n = 89$, $\alpha = 0.05$); part two

The only trait in the pooled-sex sample with a significant negative correlation with age was the sternum's dorsal ridge (Fig. C.39), but the correlation was poor ($p = -0.232$ to -0.317). The traits associated with the femur (Figs. C.1-4), clavicle (Figs. C.9,10), humerus (Figs. C.11-16), iliac crest (Figs. C.21,22), acetabulum (Figs. C.23-25), auricular surface (Figs. C.27-29), pubic symphysis (Figs. C.30-32), vertebrae (Figs. C.32-38), and to a lesser extent the ribs (Figs. C.39-43), generally showed "fair" to "moderate" correlations with age (Table C.5). Some traits showed a significant correlation but "poor" strengths of correlations (Table C.5).

Male sample

Of the 141 traits (including bilateral traits) assessed in the males, 52 were poorly correlated with age or had an undefined correlation and included most elements' relative and actual weights (Table C.5, Figs. C.4,5,7,8,11,12,18,19). The Spearman's rho for ribs two through ten shingle ribs on the right side, rib one fusion on the right side, and sternum central dorsal ridge were undefined because all the scores for each trait were identical. As a result, the variation in trait scoring could not be calculated.

No traits had a "strong" correlation with age, but nineteen had a "moderate" correlation with age (Table 4.2, Figs. 4.3 & 4.4). The scatter plots for non-moderate strength traits can be found in Appendix C. Again, all vertebral traits were significantly correlated with true age. Several traits (Table C.5) were negatively, poorly, and not significantly correlated and included the tibia and calcaneus actual weight (Figs. C.4,5,7,8), humerus relative weight (Figs. C.11,12) and shingle ribs (Fig. C.42).

Table 4.2 Traits in the male sample (n = 45) with “moderate” strength of correlation with age

Trait	Rho	p-value
Femur fovea margin lipping (L)	0.696	<0.001
Femur fovea margin lipping (R)	0.608	<0.001
Femur head surface extra bone (R)	0.634	<0.001
Clavicle medial epiphysis fusion (L)	0.625	<0.001
Clavicle medial epiphysis fusion (R)	0.666	<0.001
Scapula glenoid fossa lipping (L)	0.655	<0.001
Scapula glenoid fossa lipping (R)	0.662	<0.001
Acetabulum posterior margin lipping (L)	0.653	<0.001
Acetabulum posterior margin lipping (R)	0.659	<0.001
Pubic symphyseal relief (R)	0.675	<0.001
Pubic symphysis ventral margin (R)	0.677	<0.001
Pubic symphysis dorsal margin (R)	0.686	<0.001
L1 spinous process flattening	0.604	<0.001
L5 superior margin shape	0.772	<0.001
S1-2 fusion	0.732	<0.001
L lipping	0.672	<0.001
L candlewax	0.702	<0.001
R3-10 rim profile (L)	0.651	<0.001
R3-10 rim profile (R)	0.623	<0.001

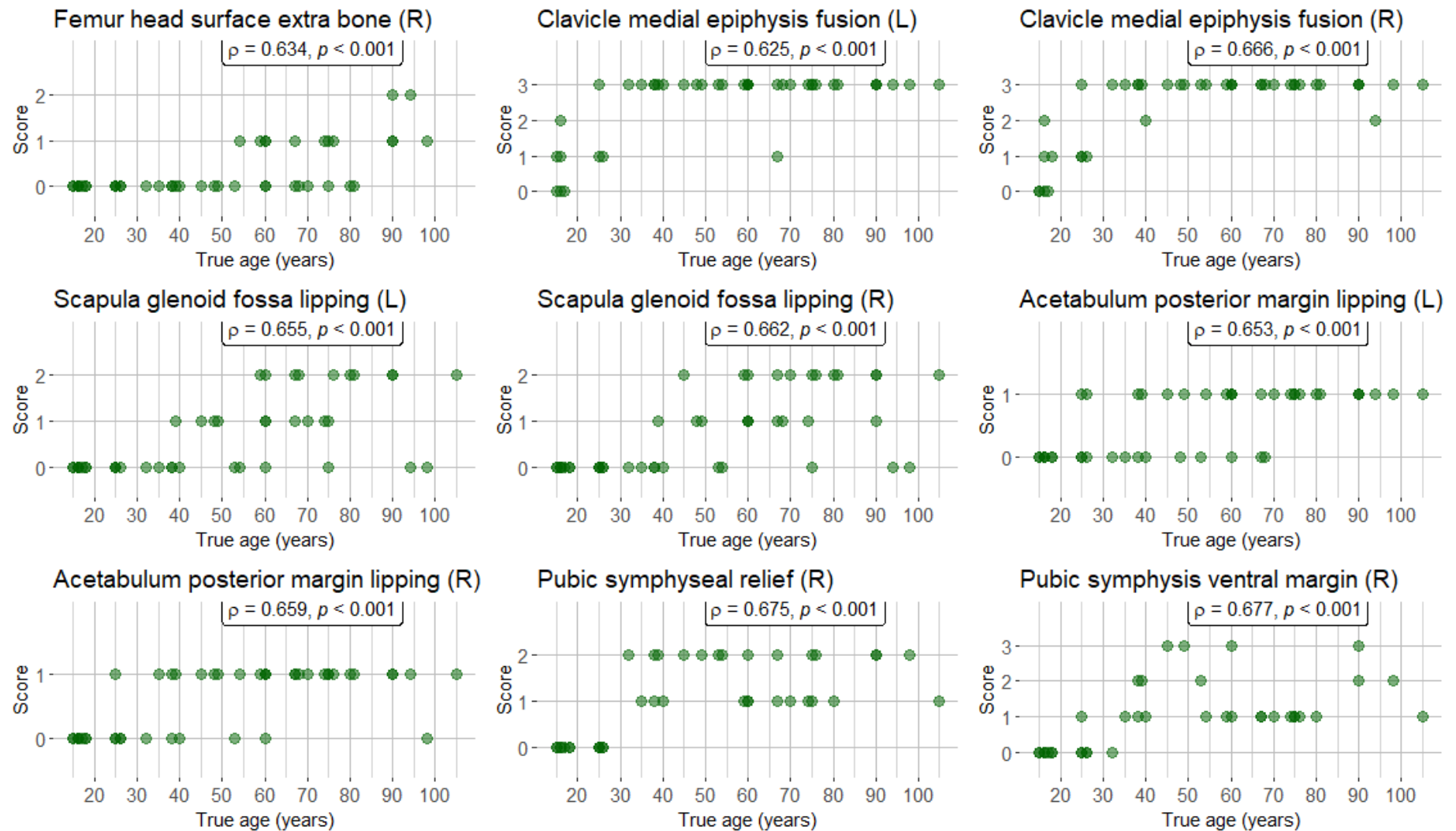


Fig. 4.3 Scatter plots for the traits with a moderate correlation with age in the male sample ($n = 45$, $\alpha = 0.05$); part one

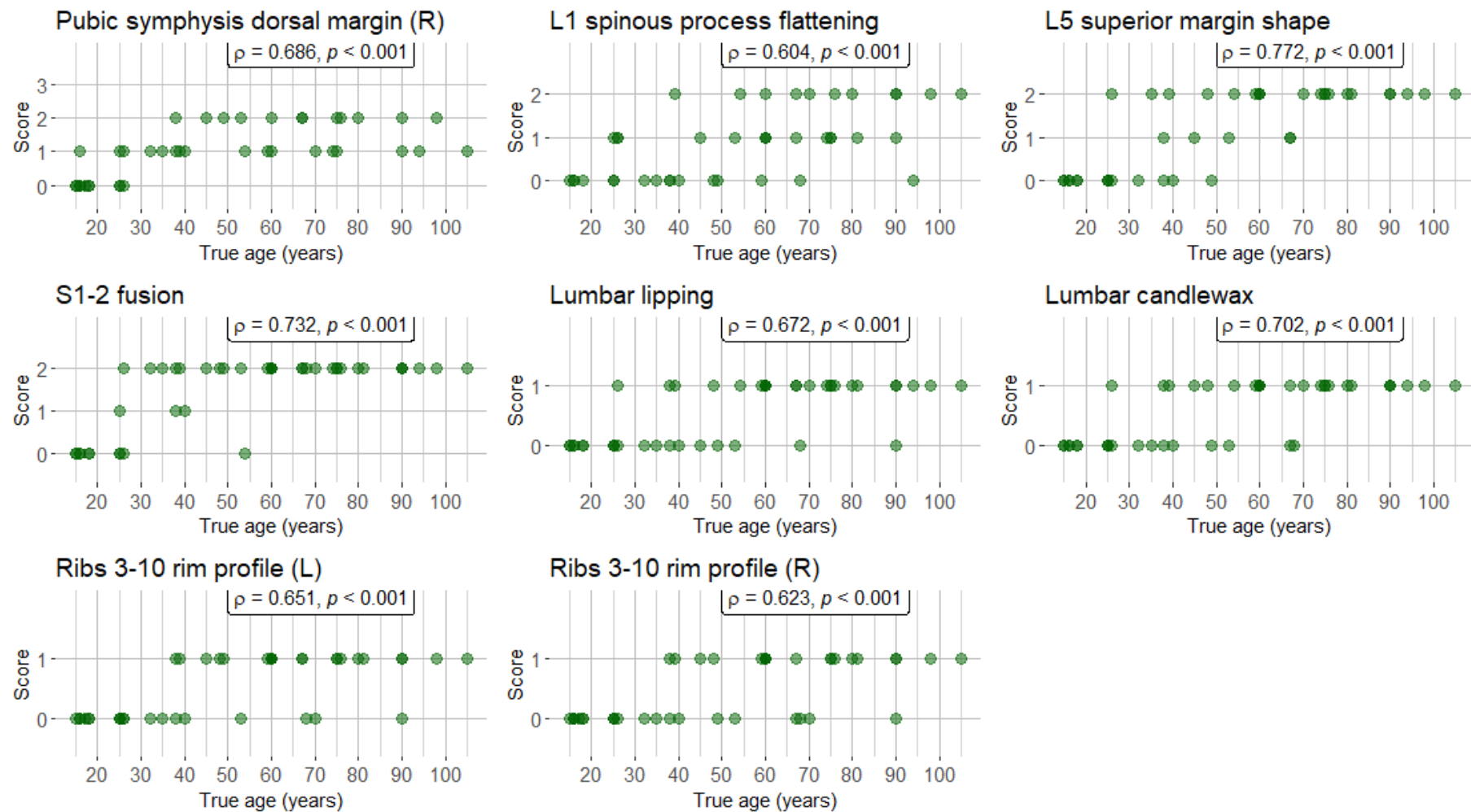


Fig. 4.4 Scatter plots for the traits with a moderate correlation with age in the male sample (n = 45, $\alpha = 0.05$); part two

Female sample

Of the 141 traits assessed (including bilateral traits) in the female sample, 57 were poorly correlated with age or had an undefined correlation. The Spearman's rho for the clavicle medial macroporosity on the left side, the ischial tuberosity medial spur bilaterally, the pubic symphyseal collar on the right side, and ribs two through ten shingle ribs bilaterally were undefined because there was no variation in the scores for each trait. No traits had a "strong" correlation with age, but twenty had a "moderate" correlation with age (Table 4.3, Figs. 4.5-4.7). The scatter plots of the non-moderate strength traits can be found in Appendix C. In the females, traits with significant negative correlations included all of the traits requiring actual weights of bony elements (Figs. C.4,5,7,8,11,12,19,20), as well as the sternum dorsal ridge (Fig. C.39), all of which had "fair" strength of correlation, except for the right humerus' actual weight which had a "moderate" strength. All other traits with a significant correlation were positively correlated with true age (Table C.5).

Table 4.3 Traits in the female sample (n = 44) with "moderate" strength of correlation with age

Trait	Rho	p-value
Femur greater trochanter roughening (L)	0.611	<0.001
Clavicle medial epiphysis fusion (L)	0.669	<0.001
Clavicle medial epiphysis fusion (R)	0.604	<0.001
Scapula glenoid fossa lipping (R)	0.696	<0.001
Humerus actual weight (R)	-0.698	<0.001
Iliac crest fusion (L)	0.609	<0.001
Iliac crest fusion (R)	0.612	<0.001
Ischial tuberosity bumps (L)	0.712	<0.001
Ischial tuberosity bumps (R)	0.652	<0.001
Pubic symphyseal relief (L)	0.603	<0.001
Pubic symphyseal relief (R)	0.696	<0.001
Pubic superior protuberance/apex (L)	0.612	<0.001
Pubic superior protuberance/apex (R)	0.641	<0.001
Pubic symphysis ventral margin (L)	0.686	<0.001
Pubic symphysis ventral margin (R)	0.649	<0.001
L1 spinous process flattening	0.716	<0.001
L5 superior margin shape	0.643	<0.001
S1-2 fusion	0.756	<0.001
T lipping	0.665	<0.001
L lipping	0.608	<0.001

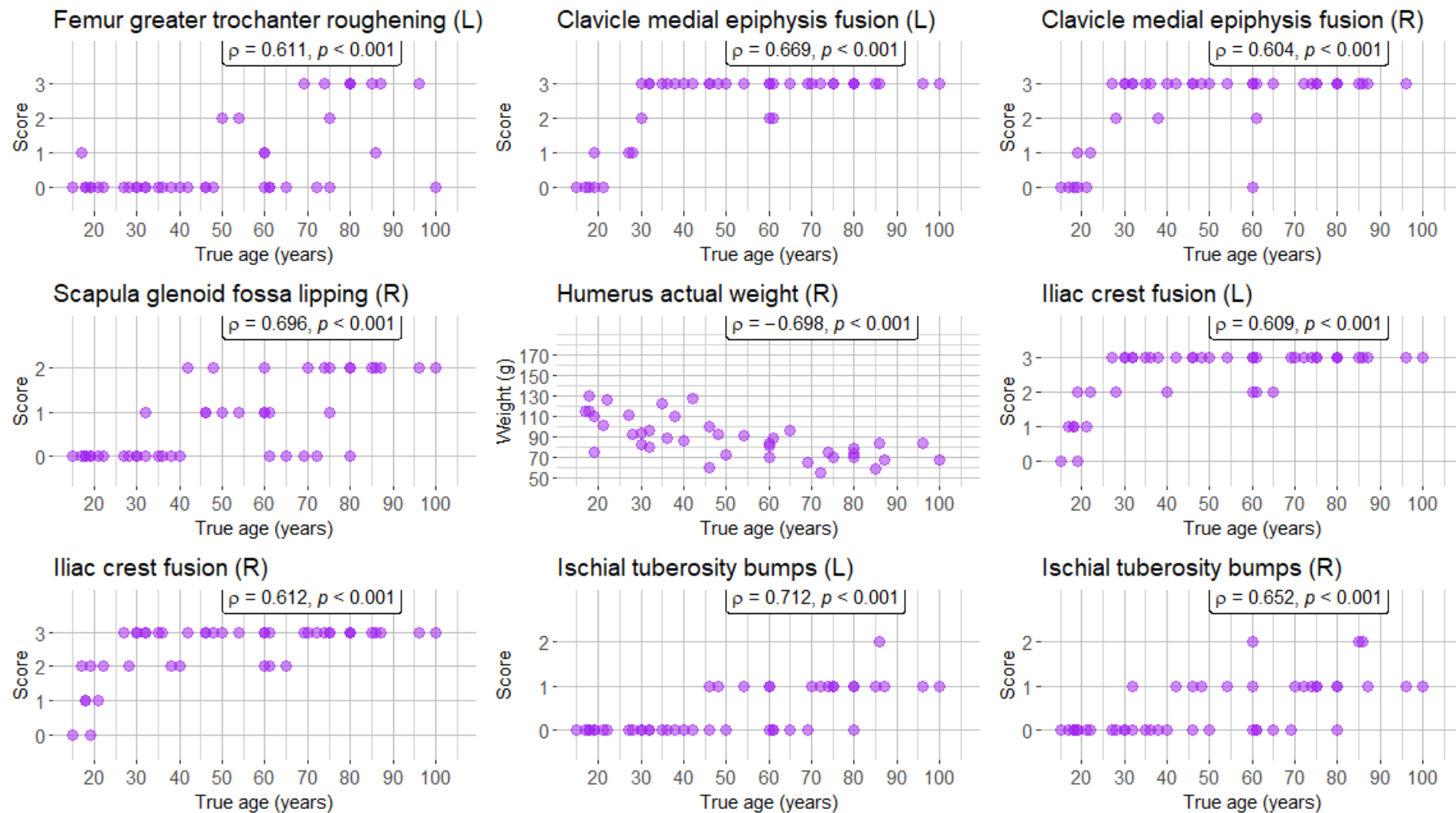


Fig. 4.5 Scatter plots for the traits with a moderate correlation with age in the female sample (n = 44, $\alpha = 0.05$); part one

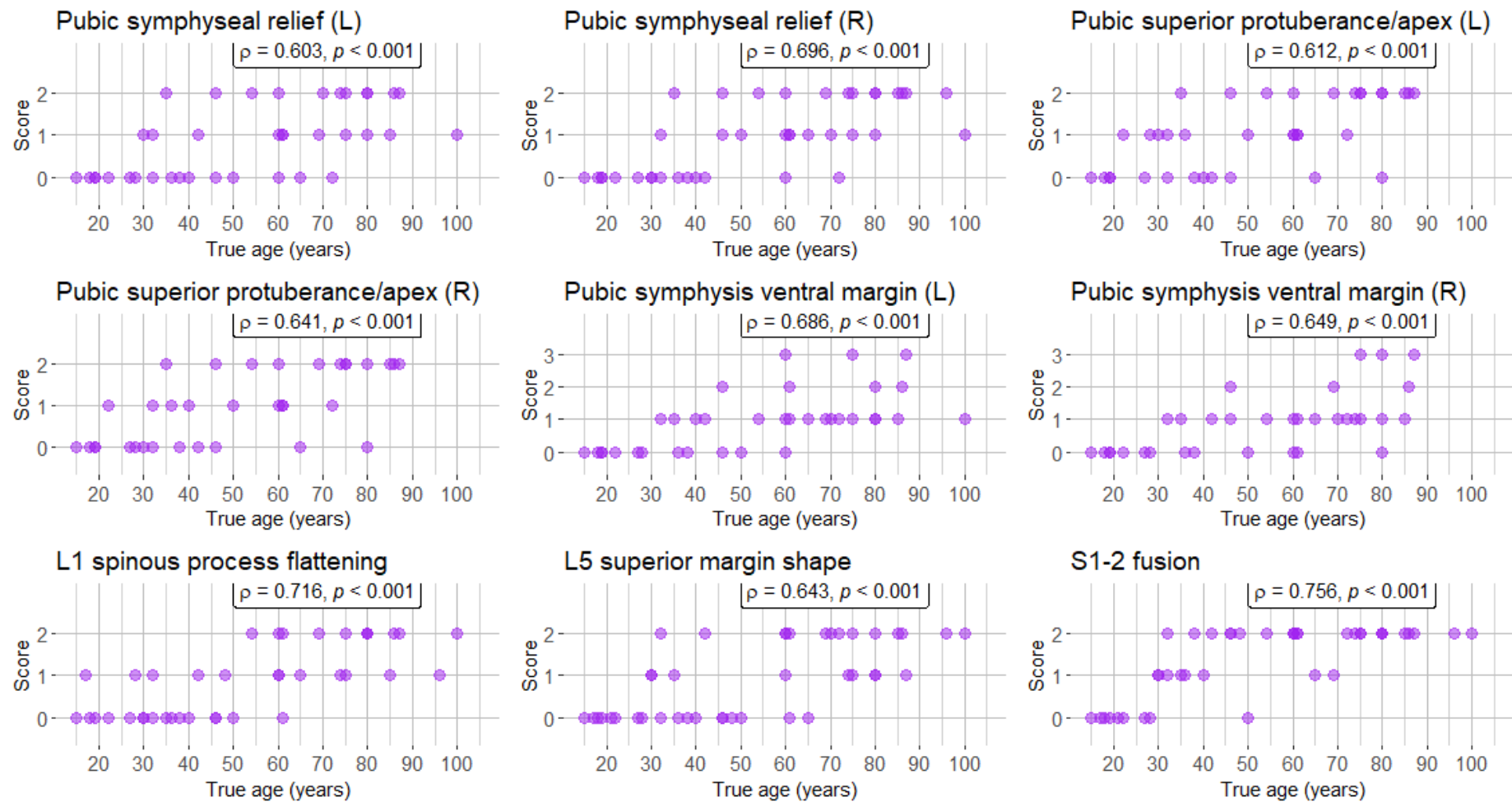


Fig. 4.6 Scatter plots for the traits with a moderate correlation with age in the female sample ($n = 44$, $\alpha = 0.05$); part two

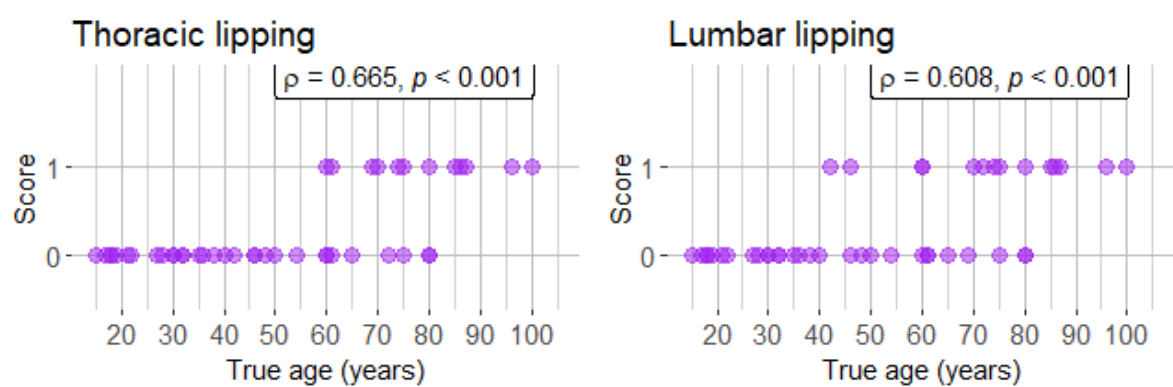


Fig. 4.7 Scatter plots for the traits with a moderate correlation with age in the female sample ($n = 44$, $\alpha = 0.05$); part three

4.2.2 Differences between males and females

When males and females were compared regarding the correlations between each trait and true age, most traits (120 out of 141) showed no significant difference in their correlation with age between males and females. The traits with significant differences are listed in Table 4.4 below.

Table 4.4: Traits with significant difference between males (n = 45) and females (n = 44) for the correlation between trait and age as determined by a Fisher's Z transformation with $\alpha = 0.05$

Trait	Fisher's Z	p-value
Femur fovea margin lipping (L)	3.026	0.003
Femur head surface extra bone (R)	2.509	0.012
Femur greater trochanter roughening (L)	-2.292	0.022
Femur trochanteric medial exostoses (L)	-2.598	0.009
Tibia actual weight (L)	2.358	0.018
Tibia actual weight (R)	2.507	0.012
Tibia length (L)	2.277	0.023
Tibia length (R)	2.571	0.010
Clavicle medial epiphysis gravel (R)	2.254	0.024
Humerus actual weight (L)	3.314	0.001
Humerus actual weight (R)	4.012	<0.001
Humerus length (L)	1.99	0.046
Humerus length (R)	2.375	0.018
Innominate actual weight (L)	3.798	<0.001
Innominate actual weight (R)	3.383	0.001
Innominate length (L)	2.419	0.016
Innominate length (R)	2.587	0.010
R2 ossification pattern (R)	2.393	0.018
R3-10 ossification pattern (L)	2.868	0.004
R3-10 ossification pattern (R)	2.156	0.031
R Occipital condyle lipping	2.430	0.015

Males had a significantly higher correlation than females for the traits on the femoral head (fovea margin lipping $Z = 3.026$, $p = 0.003$; extra bone $Z = 2.509$, $p = 0.012$), the clavicle (medial epiphysis gravel $Z = 2.254$, $p = 0.024$), and the length of the innominate (left $Z = 2.419$, $p = 0.016$; right $Z = 2.587$, $p = 0.010$). Females had a significantly higher correlation than males for the rest of the traits that were significantly differently correlated with age (Table 4.4).

Most traits (107 out of 141) were not significantly different in the degree of expression between males and females except for the traits listed below in Table 4.5.

Table 4.5 Traits with significant differences in scoring between males (n = 45) and females (n = 44) as determined by a Mann-Whitney U (Wilcoxon rank sum) test with an α of 0.05

Trait	p-value
Femur fovea margin lipping (L)	0.013
Femur head surface extra bone (L)	<0.001
Femur head surface extra bone (R)	0.006
Femur greater trochanter roughening (L)	0.024
Tibia relative weight (R)	0.005
Tibia actual weight (L)	<0.001
Tibia actual weight (R)	<0.001
Tibia length (L)	<0.001
Tibia length (R)	<0.001
Calcaneus relative weight (L)	<0.001
Calcaneus relative weight (R)	<0.001
Calcaneus actual weight (L)	<0.001
Calcaneus actual weight (R)	<0.001
Calcaneus length (L)	<0.001
Calcaneus length (R)	<0.001
Clavicle medial macroporosity (L)	0.012
Humerus relative weight (L)	<0.001
Humerus relative weight (R)	<0.001
Humerus actual weight (L)	<0.001
Humerus actual weight (R)	<0.001
Humerus length (L)	<0.001
Humerus length (R)	<0.001
Humerus medial epicondyle (R)	0.032
Radius tuberosity medial crest (L)	0.033
Innominate relative weight (L)	<0.001
Innominate relative weight (R)	<0.001
Innominate actual weight (L)	<0.001
Innominate actual weight (R)	<0.001
Innominate length (L)	<0.001
Innominate length (R)	<0.001
Sacral elbow	0.046
C candlewax	0.027
L lipping	0.031
L candlewax	0.010

When the degree of trait expression is considered, all relative weights, actual weights, and lengths were significantly different (Table 4.5), with the male sample tending towards the larger end of the spectrum. The significant difference in the degree of trait expression for the vertebral traits is interesting, particularly for the lumbar vertebrae.

The full tables of Fisher's Z and Mann-Whitney U values are in Appendix C (Tables C.6 and C.7).

4.2.3 Asymmetry

Using a percentile bootstrap on a Spearman's rho test with $n = 89$, $\alpha = 0.05$, and 2000 repetitions, most bilateral traits (52 out of 58) showed no significant difference in age-trait correlations between the left and right side except for the traits in Table 4.6 below.

Table 4.6 Traits with significant differences in the correlation between trait and age on the left and right sides as determined by a percentile bootstrap ($\alpha = 0.05$, $n = 89$, $B = 2000$)

Trait	p-value
Calcaneus length	0.044
Humerus relative weight	0.035
Humerus actual weight	0.022
Pubic symphyseal collar	<0.001
Pubic symphyseal relief	<0.001
Pubic symphysis dorsal margin	0.028

Most traits (54 out of 58) were also not significantly different in the degree of expression between the left and right sides, except for the traits listed below in Table 4.7.

Table 4.7 Traits with significant differences in scoring between the left and right sides as determined by a Wilcoxon signed rank test (paired, $n = 89$, $\alpha = 0.05$)

Trait	Z	p-value
Tibia actual weight	-2.437	0.015
Scapula glenoid fossa lipping	-2.846	0.004
Humerus actual weight	-5.607	<0.001
Humerus length	-6.035	<0.001

In this sample, asymmetry was observed in the tibiae ($Z = 2.437$, $p = 0.015$), calcanei ($p = 0.044$), humeri (weight correlation $p = 0.022$; weight expression $Z = -5.607$, $p < 0.001$; length $Z = -6.035$, $p < 0.001$), scapulae ($Z = 2.846$, $p = 0.004$), and pubic symphysis (collar $p < 0.001$; relief $p < 0.001$; dorsal margin $p = 0.028$). The pubic symphysis differences are consistent with the observation of asymmetry in the pubic symphysis scores during data collection and may merit further investigation.

The full tables of the percentile bootstrap and Wilcoxon signed rank values can be found in Appendix C (Tables C.8 and C.9).

4.3 Correlation of estimated age to true age

Age estimates were calculated using the Transition Analysis 3.0 (TA3) software, which provided a point estimate and upper and lower 95% confidence interval bounds. All

available traits were used for the initial age estimations. In contrast, the second age estimations were done with all poorly correlated traits removed to see whether results improved for the pooled-sex, male, and female samples, respectively. The software used all traits selected and did not exclude any on its own or provide indicators of outliers or atypical traits.

The correlation between true ages and the point estimate ages for the pooled sexes sample ($\alpha = 0.05$, $n = 89$) was significant with a Spearman's rho of 0.871, and after poorly correlated traits were removed, a Spearman's rho of 0.859. Both correlations are considered strong correlations. It is interesting to note that the correlation slightly decreased when the poorly correlated traits were removed. The correlations between true and point estimate ages for the male sample ($\alpha = 0.05$, $n = 45$) and female sample ($\alpha = 0.05$, $n = 44$) were significant with a Spearman's rho of 0.858 and 0.878, respectively, which are also considered strong correlations. After ages were re-estimated for the male and female samples, the correlations between the true ages and estimated ages were 0.833 and 0.874 respectively. Scatterplots for the pooled-sex sample, pooled-sex sample re-estimated without poorly correlated traits, male sample, and female sample can be found below in Fig. 4.8.

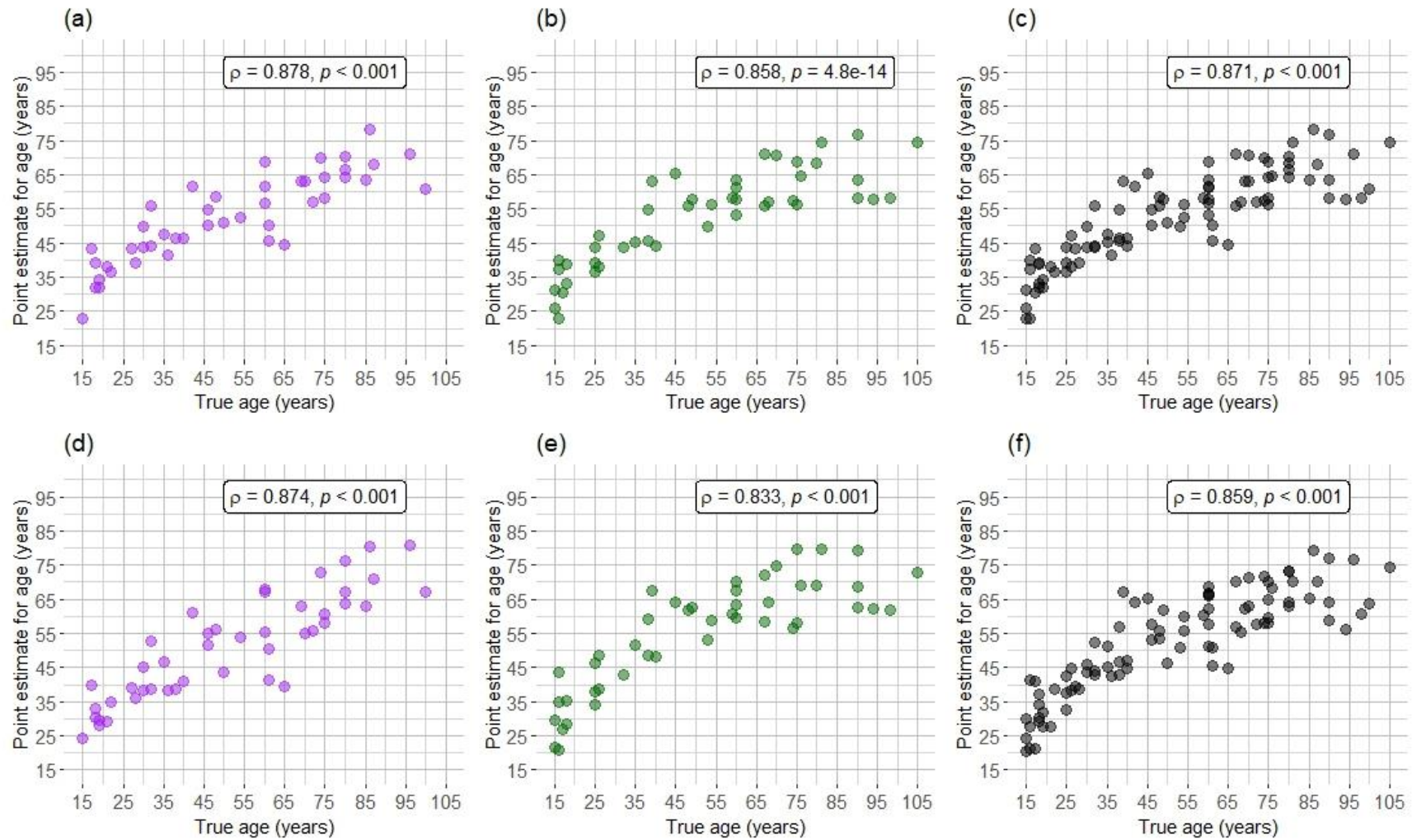


Fig. 4.8: Scatterplots of true age vs point-estimated age with a Spearman's Rank Correlation Coefficient ($\alpha = 0.05$). (a) Female sample ($n = 44$), (b) Male sample ($n = 45$), (c) Pooled-sex sample ($n = 89$), (d) Female sample with poorly correlated traits removed from the analysis ($n = 44$), (e) Male sample ($n = 45$) with poorly correlated traits removed from the analysis, (f) Pooled sex sample with poorly correlated traits removed ($n = 89$)

4.4 Accuracy, inaccuracy, and bias

The accuracy of the age estimates generated by the TA3 program was determined by taking the percentage of individuals in a cohort for whom their true age fell within the 95% confidence interval. The interval width is based on the 95% confidence interval of the composite log-likelihood function for an individual based on the traits available. As such, the interval is not necessarily symmetrical around the point estimate or of a constant width.

The accuracy of the pooled-sex sample, when all traits were included, across all age cohorts was 59.6%, with 29.2% of ages over-estimated and 11.2% under-estimated. The only age cohort in the pooled-sex sample with 100% accuracy was the group aged 55-64 years. The age cohorts below this had varying degrees of overestimation, and those above this had varying degrees of underestimation (Table 4.8). Accuracy was different when males and females were considered separately. The male age cohorts of 55-64 and 65-74 had 100% accuracy, and the female age cohorts of 45-54, 55-64, and 75-84 years had 100% accuracy. The average accuracy ranged from 55.6% (males) to 63.6% (females), with the highest accuracy obtained in the fifth (female sample) and sixth decades (all samples). Younger age cohorts (aged 54 years and below) were over-aged, and older cohorts (aged 55 years and above) were under-aged. The accuracy was generally higher for the female sample than for the male and pooled-sex samples (Table 4.8).

The average bias ranged from -1.139 (females) to -0.998 (males), and the smallest bias was observed in the age cohort of 55-64 years (-1.1 years, male sample). Average inaccuracy ranged from 12.962 (females) to 13.634 (males) years, with the lowest inaccuracy observed in the age cohort 55-64 years in the pooled-sex and male samples and 45-54 years in the female sub-sample (Table 4.9). Interestingly, while overall inaccuracy is lowest for the female sample and highest for the male sample, the opposite is observed for overall bias. The degree of bias and inaccuracy increased as the true age grew closer to the extremes of the age range of 15-105 years (Table 4.9).

Table 4.8 Accuracy of age estimates with all traits by age cohorts for the pooled sex (n = 89), male (n = 45), and female (n = 44) samples

Cohort (years)	Pooled-sex sample				Male sample				Female sample			
	n	Within 95%	Over 95%	Under 95%	n	Within 95%	Over 95%	Under 95%	n	Within 95%	Over 95%	Under 95%
15-105	89	59.6%	29.2%	11.2%	45	55.6%	31.1%	13.3%	44	63.6%	27.3%	9.1%
15-24	16	0%	100%	0%	8	0%	100%	0%	8	0%	100%	0%
25-34	12	41.7%	58.3%	0%	6	33.3%	66.7%	0%	6	50%	50%	0%
35-44	10	80%	20%	0%	5	80%	20%	0%	5	80%	20%	0%
45-54	10	90%	10%	0%	5	80%	20%	0%	5	100%	0%	0%
55-64	10	100%	0%	0%	5	100%	0%	0%	5	100%	0%	0%
65-74	10	90%	0%	10%	5	100%	0%	0%	5	80%	0%	20%
75-84	10	90%	0%	10%	5	80%	0%	20%	5	100%	0%	0%
85-94	7	42.9%	0%	57.1%	4	25%	0%	75%	3	66.7%	0%	33.3%
95-105	4	0%	0%	100%	2	0%	0%	100%	2	0%	0%	100%
75-105	21	57.1%	0%	42.9%	11	45.5%	0%	54.5%	10	70%	0%	30%

Table 4.9 Bias and inaccuracy by age cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples when all traits were included in the analysis

Cohort (years)	Pooled-sex sample			Male sample			Female sample		
	n	Bias (years)	Inaccuracy (years)	n	Bias (years)	Inaccuracy (years)	n	Bias (years)	Inaccuracy (years)
15-105	89	1.067	13.380	45	0.998	13.634	44	1.139	12.962
15-24	16	16.125	16.125	8	16.063	16.063	8	16.188	16.188
25-34	12	15.567	15.567	6	14.950	14.950	6	16.183	16.183
35-44	10	11.480	11.480	5	12.580	12.580	5	10.380	10.380
45-54	10	5.870	6.850	5	7.200	8.520	5	4.540	5.180
55-64	10	-2.480	5.480	5	-1.100	2.940	5	-3.860	8.020
65-74	10	-8.660	9.580	5	-6.800	8.640	5	-10.520	10.520
75-84	10	-12.140	12.140	5	-10.920	10.920	5	-13.360	13.360
85-94	7	-22.300	22.300	4	-26.925	26.925	3	-16.133	16.133
95-105	4	-33.600	33.600	2	-35.200	35.200	2	-32.000	32.000
75-105	21	-19.614	19.614	11	-21.155	21.155	10	-17.92	17.92

When ages were re-estimated with poorly correlated traits removed, the accuracy generally increased in all groups (Table 4.10). The overall accuracy ranged from 64.44% (males) to 70.79% (pooled-sex sample), with high accuracies for cohorts 35-44 years to 75-84 years in all samples. Younger cohorts (54 years and below in the pooled-sex and female samples, 65 years and below for males) were over-estimated, and older cohorts (55 years and above for the pooled sex and female samples, 65 years and above for the males) were underestimated. The female sample exhibited higher accuracy across most age cohorts than the male and pooled-sex samples.

Bias and inaccuracy were reduced for the female and pooled-sex samples, while the point at which bias shifted from positive to negative, or the 'turning point', remained the same (Table 4.11). The turning point for the male sample (64 years) was different from both the original turning point (54 years) as well as from the pooled-sex (54 years) and female samples (54 years). Overall bias ranged from -0.448 (females) to 3.244 (males) years, with the smallest bias observed in the age cohort of 55-64 years in the sex-pooled and female samples and the cohort of 65-74 years in the male sample. Average inaccuracy ranged from 11.439 (females) to 13.338 (males) years, with the lowest inaccuracy observed in the cohort aged 45-54 years for the male and female samples and 55-64 years for the pooled-sex sample. Slightly different to the original estimations, females had the lowest average bias and inaccuracy, and males had the largest average bias and inaccuracy. The re-estimated male sample also had an unexpected increase in bias in the age cohort of 35-44 years.

Table 4.10 Age estimate accuracy per cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples after poorly correlated traits were removed from the analysis

Cohort	Pooled-sex sample				Male sample				Female sample			
	n	Within 95%	Over 95%	Under 95%	n	Within 95%	Over 95%	Under 95%	n	Within 95%	Over 95%	Under 95%
15-105	89	70.8%	20.2%	9%	45	64.4%	24.4%	11.1%	44	70.5%	20.5%	9.1%
15-24	16	6.3%	93.8%	0%	8	0%	100%	0%	8	0%	100%	0%
25-34	12	83.3%	16.7%	0%	6	66.67%	33.3%	0%	6	83.3%	16.7%	0%
35-44	10	90%	10%	0%	5	80%	20%	0%	5	100%	0%	0%
45-54	10	100%	0%	0%	5	100%	0%	0%	5	100%	0%	0%
55-64	10	100%	0%	0%	5	100%	0%	0%	5	80%	0%	20%
65-74	10	90%	0%	10%	5	100%	0%	0%	5	80%	0%	20%
75-84	10	100%	0%	0%	5	100%	0%	0%	5	100%	0%	0%
85-94	7	57.1%	0%	42.9%	4	25%	0%	75%	3	66.7%	0%	33.3%
95-105	4	0%	0%	100%	2	0%	0%	100%	2	50%	0%	50%
75-105	21	66.7%	0%	33.3%	11	54.6%	0%	45.5%	10	80%	0%	20%

Table 4.11 Bias and inaccuracy per cohort for the pooled-sex (n = 89), male (n = 45), and female (n = 44) samples after poorly correlated traits were removed from the analysis

Cohort (years)	Pooled-sex sample			Male sample			Female sample		
	n	Bias (years)	Inaccuracy (years)	n	Bias (years)	Inaccuracy (years)	n	Bias (years)	Inaccuracy (years)
15-105	89	0.940	12.599	45	3.244	13.338	44	-0.448	11.439
15-24	16	12.694	12.694	8	13.788	13.788	8	12.463	12.463
25-34	12	13.675	13.675	6	14.883	14.883	6	11.733	11.733
35-44	10	12.710	12.710	5	17.000	17.000	5	6.860	6.860
45-54	10	6.700	7.840	5	10.240	0.000	5	5.860	0.000
55-64	10	-0.540	6.740	5	4.460	4.620	5	-4.060	9.940
65-74	10	-8.510	9.410	5	-4.040	8.000	5	-12.800	12.800
75-84	10	-11.230	11.230	5	-6.360	8.240	5	-12.840	12.840
85-94	7	-21.614	21.614	4	-22.90	22.900	3	-14.567	14.567
95-105	4	-30.875	30.875	2	-34.250	34.250	2	-24.000	24.00
75-105	21	-18.433	18.433	11	-17.445	18.300	10	-15.590	15.590

The overall trend for the over- and under-estimation of age in males and females differed between the original estimations and the re-estimations when poorly correlated traits were removed from the analysis. In the original estimations, females were over-aged more than males for ages 15-34 years and then under-aged more for ages 65-84 years. When poorly correlated traits were removed, males were more over-aged than females for ages 15-64 years and more under-aged for ages 85-105 years.

It is important to take note of the differences between the average bias values and the cohort-specific bias values. The overall score is not representative of the actual bias seen in a real population, and strong biases at different ends of the age spectrum mathematically cancel each other out, so the resultant bias value is deceptively small. This is discussed further in the discussion chapter.

To illustrate the difference in distribution between true ages, estimated ages calculated using all traits (point estimate 1) and estimated ages calculated after poorly correlated traits were removed (point estimate 2), box plots of the age distributions in the pooled-sex, male, and female samples were generated (Fig. 4.9). The true range of ages in the sample was much broader than that of the estimated ages. The median age of the sample was lower for the true age distribution than for both estimated ages, implying an attraction to the middle for the estimated ages.

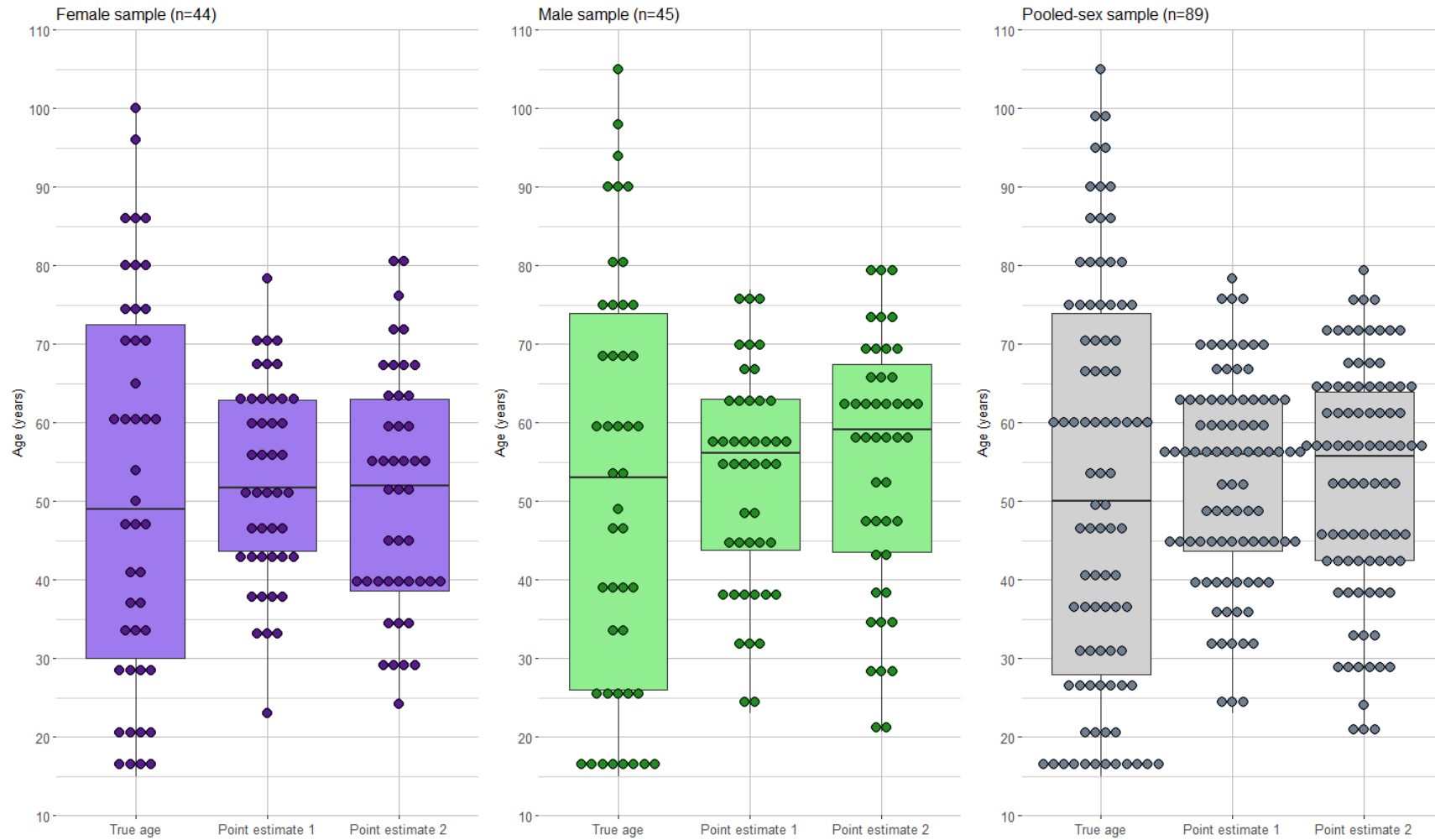


Fig. 4.9 Box plots of the age distribution for the female ($n = 44$, purple), male ($n = 45$, green), and pooled-sex ($n = 89$, grey) samples showing the differences between the true distribution, the first estimated distribution with all traits present (point estimate 1), and the second estimated distribution with poorly correlated traits removed (point estimate 2)

4.5 Sensitivity to missing data

The sample used to assess TA3's ability to work with absent data was a pooled-sex sample. All sub-samples for region absence exhibited a significant correlation between true age and estimated age (Table 4.12; Figs. 4.10 & 4.11) and satisfied the null hypothesis that there is no significant difference in the correlation between true age and estimated age between a sample with complete region representation and a sample missing an entire region of data points (i.e., upper limb) (Table 4.13).

Table 4.12: Spearman's Rank Correlation Coefficient ($\alpha=0.05$, $n=68$) for true age vs estimated age in a complete dataset and incomplete datasets (pooled-sexes)

Sub-sample	Rho	p-value
All elements present	0.832	<0.001
Upper limb absent	0.801	<0.001
Lower limb absent	0.820	<0.001
Pubic symphysis absent	0.808	<0.001
Auricular surface absent	0.824	<0.001
Cranium absent	0.828	<0.001
Ribs absent	0.834	<0.001
Vertebrae absent	0.785	<0.001
Innominate absent	0.831	<0.001

When the ribs were removed from the analysis, the correlation between true age and estimated age ($\rho = 0.834$, $p < 0.001$) became slightly higher than the correlation when all elements were present ($\rho = 0.832$, $p < 0.001$). The largest decrease in correlation occurred when the vertebrae were removed ($\rho = 0.785$, $p < 0.001$). The smallest decreases in correlation occurred when the cranium ($\rho = 0.828$, $p < 0.001$), lower limb ($\rho = 0.820$, $p < 0.001$), and traits on the innominate (innominate $\rho = 0.831$, auricular surface $\rho = 0.824$, $p < 0.001$) except for the pubic symphysis were removed.

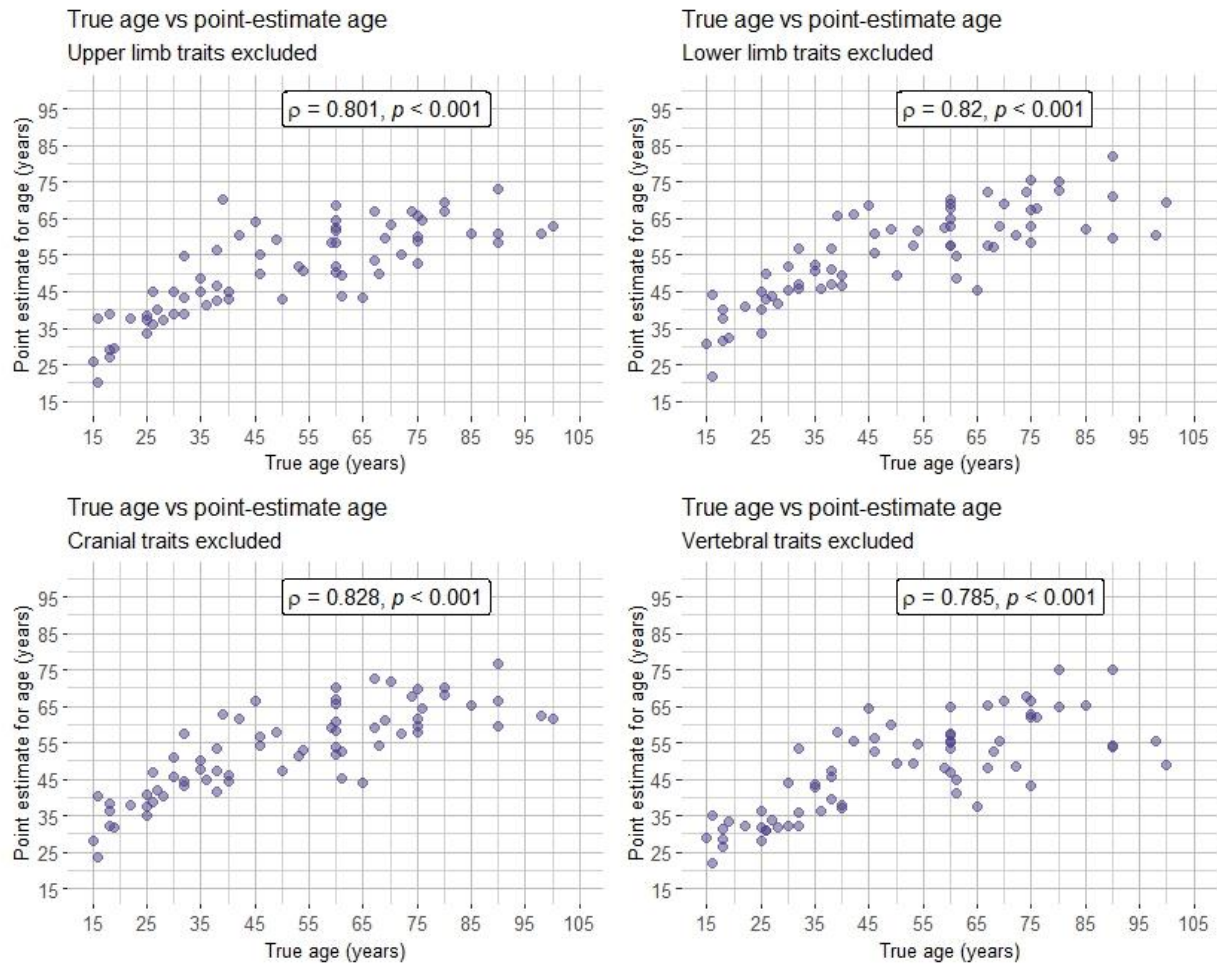


Fig. 4.10 Scatter plots of true age vs point estimate age for sub-samples in which a single scoring region was fully removed (n = 68)

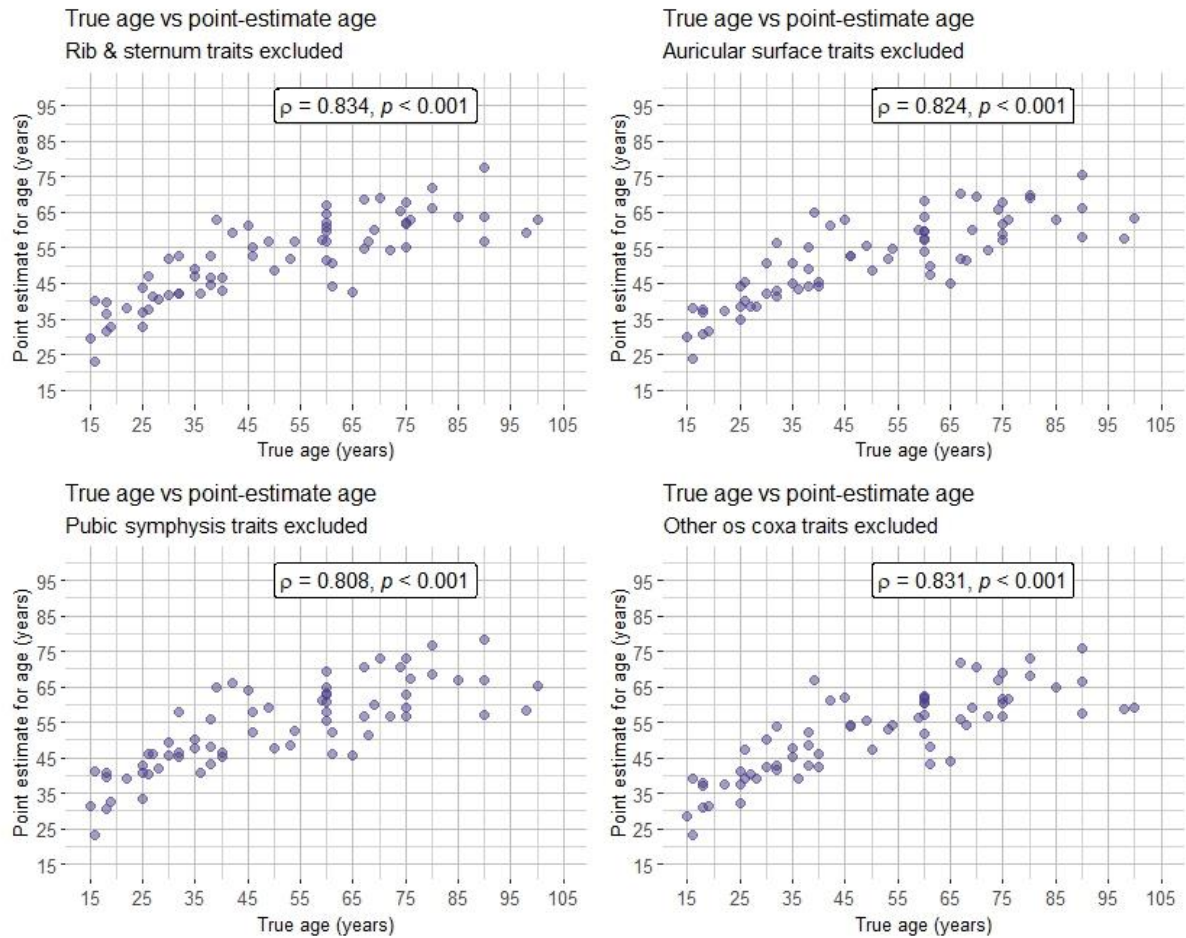


Fig. 4.11 Scatter plots of true age vs point estimate age for sub-samples in which a single scoring region was fully removed ($n = 68$)

Removing specific regions did not result in significant differences in the correlations between true age and estimated age. However, there were noticeable decreases in the correlation when the pubic symphysis, vertebrae, and upper limb elements were excluded. This suggests that these skeletal regions play a crucial role in accurately predicting age using this method. These results indicate that it is best to use all available traits, but there are problems in a practical setting.

Table 4.13: Fisher's Z transformation test ($\alpha = 0.05$, $n = 68$) for a significant difference in correlation between true age and estimated age in a full dataset vs incomplete dataset combinations

Sub-sample	Fisher's Z	p-value
Upper limb removed	0.555	0.579
Lower limb removed	0.228	0.820
Pubic symphysis removed	0.435	0.664
Auricular surface removed	0.155	0.877
Cranium removed	0.095	0.925
Ribs & sternum removed	-0.011	0.991
Vertebrae removed	0.789	0.430
Innominate removed	0.045	0.964

Chapter 5 Discussion

The goal of this study was to assess the performance of the TA3 method (Getz, 2017; Getz *et al.*, 2017; Boldsen *et al.*, 2021a; Getz and Byrnes, 2021; Getz and Galimany, 2022) in a black South African sample to ascertain whether it is a suitable and appropriate method to use in a South African forensic setting. Parameters assessed included repeatability, correlations between skeletal trait and age, correlations between true and estimated ages, accuracy, inaccuracy, bias, sex differences, asymmetry, and the method's ability to function with an incomplete dataset.

Unlike the previous version of Transition Analysis (TA2) (Boldsen *et al.*, 2002; Milner and Boldsen, 2012; Boldsen and Milner, 2016), the reference sample for TA3 includes South African black individuals, and the age-informing trait list has been substantially expanded. These changes to the method, its reference sample, and the lack of information regarding its function make it incredibly important that TA3 is tested and validated before its use in a practical forensic capacity. In order to validate TA3, the following aspects need to be assessed in different populations: trait-age correlations, asymmetry, sex differences, and population differences associated with TA3, including accuracy and trait-age correlation, which provides the need for this study.

5.1 Repeatability

Intra-observer repeatability was much higher than inter-observer repeatability, which may indicate the level of experience required to use the scoring method associated with TA3 effectively. Additionally, many of the traits used in TA3 have either not been used as an indicator before or have been minimally used and found lacking in isolation (Listi and Manhein, 2012; Listi, 2016; Getz, 2017; Botha *et al.*, 2019; Blatt *et al.*, 2020; Muñoz-Silva *et al.*, 2020). As such, practitioners are less likely to be familiar with the newer traits, possibly impacting their repeatability.

The clavicular traits had particularly high intra-observer repeatability, as did those found on the ischial tuberosity, vertebrae, ribs, scapula, spheno-occipital synchondrosis, and humerus. The femur exhibited high inter-observer repeatability, while the aforementioned bones/sites did not. The high intra-observer repeatability of scoring for the clavicle is consistent with the literature (Mahon *et al.*, 2018b; Blatt *et al.*, 2020), as it is for the spheno-occipital synchondrosis (Bassed *et al.*, 2010; Mahon *et al.*, 2017), ribs (Oettlé and Steyn, 2000; Hartnett, 2010b), and scapula (Brenneman,

2011; Zampetti *et al.*, 2016). The low inter-observer repeatability for the sphenoccipital synchondrosis may be related to it not being a common trait used in age estimation, unlike the more traditional cranial suture methods (Todd and Lyon Jr., 1924, 1925a, 1925b, 1925c; Acsádi and Neméskeri, 1970; Meindl and Lovejoy, 1985).

The auricular surface's lower inter-observer repeatability is in contrast to Lovejoy *et al.* (1985) and Buckberry and Chamberlain's (2002) findings, which indicated inter-observer correlations of 0.76 to 0.81 and a Cohen's Kappa of 0.66 respectively. The same is true for the lower inter-observer repeatability observed for the pubic symphysis compared to Hartnett (2010b), who reported good intra- and inter-observer repeatability. The low inter-observer repeatability for the pubic symphysis is also in contrast with Merrit's (2014) findings of a Cohen's Kappa of 0.913 for both the Suchey-Brooks (1990) and Hartnett (2010b) methods. Hens and colleagues (Hens *et al.*, 2008; Hens and Godde, 2016) did, however, indicate that Lovejoy *et al.*'s (1985) method is complicated to use for several reasons. Seriation was not possible due to space limitations, and it is not possible to seriate forensic cases (Hens *et al.*, 2008), which is a disadvantage of Lovejoy *et al.*'s (1985) method in a forensic setting as opposed to a palaeodemographic setting. The changes to the auricular surface are relatively complex, including varying degrees of exostosis and porosity development. As Lovejoy *et al.* (1985) point out, different parts of the auricular surface of the ilium may correspond with conflicting age cohorts. This results in the practitioner having to "force" an overall score based on their own subjective interpretation of the auricular surface as a whole since Lovejoy *et al.*'s (1985) method is based on the phase of the entire area of interest, instead of phases of components of that area (Hens and Godde, 2016).

The low inter-observer repeatability for the auricular surface and pubic symphysis is, however, in agreement with Jooste *et al.*'s (2016) findings of low repeatability for the same regions in the previous version of Transition Analysis (TA2). Similar to Jooste and colleagues' (2016) experience with the previous version of Transition Analysis, there was confusion over the exact location of retro-auricular traits and the scoring of the pubic symphysis ventral margin. Another problem in this study with the auricular surface was the difficulty distinguishing between developmental topography or maceration damage and exostoses development in the case of the auricular surface. For the pubic symphysis, there was a noted confusion between stage one (rampart

development) and three (rim breakdown) of the ventral and dorsal margins. This problem has also been noted by Kimmerle *et al.* (2008). The intra-observer repeatability ($\kappa = 0.690$, $p = 0.004$) for scoring the symphyseal relief was substantial. However, the inter-observer repeatability ($\kappa = 0.297$, $p = 0.096$) was only of “fair” strength, indicating that the trait was relatively subjective. Uncertainty typically occurs concerning stage one (residual billows), which describes the presence of at least two consecutive billows on the symphyseal face, but what one person may consider to be late-stage billowing (stage zero) may be considered by another to be residual billowing at best. A similar problem may also occur when distinguishing between residual billows and a flat surface. One individual may identify at least two consecutive but very faint billows, while another may consider those “billows” too faint and instead classify the symphyseal surface as flat (stage two).

A similar but inverted problem occurred when scoring the pubic symphysis superior protuberance, which had moderate intra-observer repeatability ($\kappa = 0.407$, $p = 0.062$) and slight inter-observer repeatability ($\kappa = 0.180$, $p = 0.236$). A late stage zero (serrated) and two (flat) looked very similar except in the early stages of serration, which is quite distinct. Confusion frequently occurred when some traits on the pubic symphysis indicated a more advanced age, whereas others on the same individual indicated the opposite. Such confusion is a common problem with phase-based traits and has been commented on by many authors for several traits (Lovejoy *et al.*, 1985; İşcan *et al.*, 1985, 1992; Brooks and Suchey, 1990; Oettlé and Steyn, 2000; Hartnett, 2010a; Hens and Godde, 2016). Many traits on the innominate were difficult to score, particularly in the beginning, due to a lack of reference images that did not illustrate the extremes of expression of some traits (e.g., iliac crest tuberculum ossification, Fig. C.25), poor image quality overall, and trait locations and phase descriptions that were frequently unclear and subjective, which is a common problem with phase-based methods (Garvin and Passalacqua, 2012; Márquez-Grant, 2015; Shirley and Ramirez Montes, 2015).

The acetabulum traits had lower repeatability than expected. However, the traits assessed using the TA3 scoring criteria were not the same as the traits assessed in previous literature (Calce, 2012; San-Millán *et al.*, 2017, 2019; Navega *et al.*, 2018b; San-Millán and Rissech, 2023). Inter-observer repeatability was particularly bad for the posterior margin traits, likely due to uncertainty regarding the exact location of the

scoring area and potential identification of the margin edge from which to measure lipping. Furthermore, additional bone deposition was frequently found just outside the scoring area described in the scoring manual (Boldsen *et al.*, 2021a). Given its proximity to the scoring area, this introduced confusion over whether that should be scored as present.

The low inter-observer scores for the vertebrae are likely due to differing opinions on what qualifies as “candlewax” (Burt, 2013), identification of the vertebral body margin, and unclear reference images for surface morphology. It is difficult to compare the repeatability of the vertebral traits to the literature as most studies (e.g., Van der Merwe *et al.*, 2006; Watanabe and Terazawa, 2006; Listi and Manhein, 2012; Praneatpolgrang *et al.*, 2019) do not provide repeatability results, only correlations, however, Cardoso and Ríos (2011) did provide repeatability information for the fusion of the vertebral body epiphyses. The intra-observer repeatability scores for both the superior and inferior epiphyses of L1 and L5 ($\kappa = 1.000$, $p = 0.002$) are higher than those of Cardoso and Ríos (2011), but the inter-observer repeatability scores ($\kappa < 0.001$) are considerably worse, except for the inferior epiphysis of L1 ($\kappa = 1.000$, $p = 0.002$). The differences in inter-observer repeatability between this study and Cardoso and Ríos (2011) may be related to the difficulty distinguishing between the remnant line and fully fused phases, as discussed later in this section. Given that the intra-observer repeatability was high for the vertebral traits, the low inter-observer repeatability indicates high subjectivity in the scoring criteria.

Relative weight has been mentioned by Hartnett (2010a, 2010b) as an influencing factor in the scoring of ribs and the pubic symphyses for age estimation. This is partially supported by the high intra-observer repeatabilities for the tibia, calcaneus, and innominate relative weights and the high inter-observer repeatabilities for the tibia and calcaneus relative weights. It should be noted, however, that the relative weight scoring itself was highly subjective, requiring extensive experience in the target population. Additionally, Botha *et al.* (2019) found that the South African black population is less prone to the development of osteoporosis and develops osteoporosis later in life than when compared to European populations. It also failed to account for unexpectedly low bone mass resulting from over-maceration, and for bones that were unexpectedly heavy given their size. The actual weight and length

measurements seem inconsequential as their values are not involved in the estimation, nor is the arbitrarily calculated density.

Traits such as the ulna olecranon spur and iliac crest tuberculum ossification were difficult to assess due to difficulty measuring the length of the ossification. Determining the point to measure from was a common problem, as was the physical measurement due to the size of the calliper head versus the available space. Many ulnae had spurs that were not quite long enough to be scored as present, but they may have been scored differently if the user-identified starting point for measurement was different. A different approach to scoring traits with “depth” problems like the ulna olecranon process or the iliac crest tuberculum ossification is to extend the back of the calliper into the scoring area and then measure the distance between the calliper jaws to estimate the length of the bony protrusion.

Poor maceration was a common problem, along with damage from over-maceration, sampling, and handling, which likely contributed to the poor repeatability results by damaging sites of interest and obscuring some sites of interest with unremoved tissue. These problems may be an artefact of the age of the Raymond A. Dart Collection of Modern Human Skeletons, as maceration methods have changed and improved over time. Additionally, some individuals in the study sample were among the first 100 skeletons accessioned into the collection. Scratching from maceration was sometimes difficult to distinguish from bone “roughness” as described for the greater trochanter roughening, especially since there were only reference images for the final stage and not the three preceding stages. Residual tissue also interfered with the scoring of the femur trochanteric fossa traits and rib traits, particularly when assessing the depth of the pit and measuring the length of a “cone” originating from the base of the pit.

The high intra- and inter-observer repeatability of certain enthesal traits such as the humerus medial epicondyle (intra-observer $\kappa = 1.000$, $p = 0.002$; inter-observer $\kappa = 0.615$, $p = 0.035$), ischial tuberosity superior margin spur (intra-observer $\kappa = 1.000$, $p = 0.002$; inter-observer $\kappa = 1.000$, $p = 0.002$), and ulna olecranon spur (intra-observer $\kappa = 0.821$, $p = 0.002$; inter-observer $\kappa = 0.688$, $p = 0.009$), is consistent with the literature (Weiss, 2003, 2004; Villotte *et al.*, 2010; Henderson *et al.*, 2016). The high repeatability of these traits may be related to the dimensions of bony outgrowth required for the trait to be scored as present. However, the low repeatability of other

enthesal traits, such as the ilium anterior inferior iliac exostoses (intra-observer $\kappa < 0.001$, $p = \text{undefined}$; inter-observer $\kappa < 0.001$, $p = \text{undefined}$) and the humerus lateral epicondyle (intra-observer $\kappa = 0.375$, $p = 0.236$; inter-observer $\kappa = 0.375$, $p = 0.236$), indicates that more specific definitions and descriptions are required.

Another common problem was differentiating between the remnant line and fully fused phases of epiphyseal traits. This problem has also been observed by Stevenson (1924), Langley-Shirley and Jantz (2010), Cardoso and Ríos (2011), and Belcastro *et al.* (2019). Langley-Shirley and Jantz (2010) assessed two scoring systems for the fusion of the medial clavicle epiphysis, and they found that the simpler three-phase system of unfused, fusing, and completely fused was less subjective and easier to apply for individuals of any experience level as opposed to the five-phase system by McKern and Stewart (1957). Langley-Shirley and Jantz (2010) noted particularly high bias in the scoring of active vs recent fusion by a less experienced member of the study. In addition, many methods utilise a three-phase approach for the fusion of the medial clavicle (Schaefer and Black, 2005; Cardoso, 2008; Langley-Shirley and Jantz, 2010; Milenkovic *et al.*, 2013).

This study scored the fusion of the medial clavicle on a four-point scale where the second phase of Langley-Shirley and Jantz's (2010) scoring system was split into two phases, partial fusion and remnant line presence. Milenkovic *et al.* (2013) tested the three-phase method for scoring the fusion of the clavicle's medial epiphysis and found a Cohens Kappa value of 0.851, which is almost perfect agreement. This study's intra-observer repeatability for the fusion of the medial clavicle was slightly lower than that found by Milenkovic *et al.* (2013) but also almost perfect agreement ($\kappa = 0.825$). The inter-observer repeatability for this trait in the current study was much lower, with a Cohen's Kappa value of 0.271 ($p = 0.058$), which is only a "fair" strength of agreement. The intra-observer repeatability is two levels of strength above the inter-observer repeatability, indicating that the trait scoring is relatively subjective. Clavicle fusion may have higher inter-observer repeatability if the "partially fused" and "remnant line" phases are collapsed into a single phase, so there are three phases.

5.2 Correlations between traits and true age and true age with estimated age

The high correlation between age and clavicle epiphysis fusion ($\rho = 0.640$ to 0.644 , $p < 0.001$) is consistent with the findings of Black and Scheuer (1996), Langley-Shirley

and Jantz (2010) and Mahon *et al.* (2018b). The high correlation is related to the lack of mechanical strain the clavicle experiences after development is complete, particularly for the medial epiphysis encapsulated in a synovial joint with the manubrium (Cunningham *et al.*, 2016). The lack of mechanical strain provides an advantage compared to other traits subject to external factors, as an individual's relative strength and activity patterns are unlikely to affect the medial clavicle. The medial clavicle is typically the last epiphysis to fuse, with South African black individuals occasionally presenting as unfused at the age of 25 years, with partial fusion between the ages of 20 and 28 years, and complete fusion between the ages of 21 and 40 years (Mahon *et al.*, 2018b). The combination of the slow maturation of this epiphysis and the lack of mechanical strain on the clavicle (under normal circumstances) makes it a useful indicator of age. The primary influence on its appearance is age, and epiphyseal fusion is a consistent and well-documented age estimator across many populations (Cunningham *et al.*, 2016).

The poor correlation between age and the clavicle medial macroporosity and gravel contrasts the findings of Falys and Prangle (2015) and Blatt *et al.* (2020). This may be related to the difficulty in scoring “gravel” due to poor reference images and descriptions. In addition, it is difficult to differentiate between macroporosity, pitting due to variation in the development of the surface, and pits caused by maceration damage. Many porosity-related traits were not correlated with age in this study, which may result from the aforementioned population-level difference in loss of bone density (Botha *et al.*, 2019) and the wide range of causes for osteoporosis. Osteoporosis and associated macroporosity can be caused by a variety of factors, including mineral deficiencies, haematopoietic diseases, neoplastic conditions, and a lack of strain on bone due to immobilisation which causes increased resorption of calcium (Brickley and Ives, 2008).

Other traits highly correlated with age, in agreement with the literature, include the scapula glenoid fossa lipping (McKern and Stewart, 1957; Brenneman, 2011; Zampetti *et al.*, 2016), pubic symphyseal relief and dorsal and ventral margins (Brooks and Suchey, 1990; Hens *et al.*, 2008; Hartnett, 2010a), sacrum fusion (Ríos *et al.*, 2008; Passalacqua, 2009; Colarusso, 2016; Mahon *et al.*, 2018a), and cervical and lumbar vertebrae lipping and candlewax (Snodgrass, 2004; Van der Merwe *et al.*, 2006; Watanabe and Terazawa, 2006; Praneatpolgrang *et al.*, 2019).

Interestingly, most of the highly correlated traits in this study involved bony deposition. The reasons for deposition include developmental processes like epiphyseal fusion, as seen in the clavicle medial epiphysis and iliac crest, arthritic processes like joint lipping, enthesal processes like roughening of ligament/cartilage insertions, and reactions of the vertebral column to injury of its associated ligaments and intervertebral disks which causes the development of osteophytes (Maat *et al.*, 1995; Prescher, 1998; Snodgrass, 2004). The developmental component is expected due to the known positive correlation with age, much like well-documented and stable progression of juvenile development (Cunningham *et al.*, 2016), but the arthritic and enthesal components are surprising since it is generally understood that these processes are heavily influenced by activity (Shane Anderson and Loeser, 2010; Listi and Manhein, 2012; Milella *et al.*, 2012; Villotte and Knüsel, 2013; Becker, 2020). An explanation for this is that the link to activity, despite the correlation to age, is the result of long-term “microtraumatic stress” that entheses are subjected to in daily life movements of walking, carrying things, picking things up, and others (Milella *et al.*, 2012).

Villotte and Santos (2023) describe a moderate to strong association with age and state that enthesal changes are more common in older individuals, particularly males. This study saw significant differences in the trait-age correlation and degree of expression between males and females. The enthesal traits in this study that show significant differences between males and females are primarily located on the proximal femur, medial clavicle, distal humerus, and proximal radius and are primarily attachment points for important muscles controlling limb movement. Males tend to have more muscle mass and are more likely to do manual labour (Villotte *et al.*, 2010; Refai, 2019), which would increase the stress at each attachment site and thus increase enthesal activity. It is possible that the males in this study’s sample came from a background in manual labour, such as construction, farming, and mining. This combined with the male tendency to create bone is likely the reason for the difference between males and females for these traits and is consistent with the literature regarding enthesal changes, occupation estimation, and enthesopathy distribution within a population (Benjamin *et al.*, 2006; Villotte *et al.*, 2010; Ibáñez-Gimeno *et al.*, 2013; Lieverse *et al.*, 2013; Palmer *et al.*, 2016; Refai, 2019; Villotte and Santos, 2023). Different results may be seen in a more contemporary population due to changes in working conditions and sex distribution for different occupations.

The high prevalence of enthesal traits with a “poor” to “fair” correlation with age is consistent with the literature as well (Wilczak, 1998; Weiss, 2003, 2004; Milella *et al.*, 2012; Michopoulou *et al.*, 2015; Listi, 2016; Mazza and Silva, 2023). It should be noted that enthesal changes are typically correlated with increasing age but generally do not have a particularly high correlation with age (Listi, 2016; Villotte and Santos, 2023). Enthesal changes can happen due to repetitive movements or less frequent but high-force movements, such as in meal grinding or archery, and so are of great interest to archaeologists for their potential to inform the occupation of an individual (Villotte *et al.*, 2010; Yonemoto, 2016). Listi (2016) investigated enthesal changes to the femur and os coxa and found eight markers with significant but relatively weak correlations with age, including the ischial tuberosity and the iliac tubercle. However, they found that the most extreme presentation of these traits usually occurred in individuals over the age of 70 years and rarely occurred in individuals below the age of 60 years (Listi, 2016). This is consistent with this study’s findings that the most severe enthesal changes were recorded in elderly individuals. However, there were a few individuals who were relatively young and presented with considerable enthesal changes, which may have been related to activity or pathology (Belanger and Rowe, 2001; Villotte *et al.*, 2010; Milella *et al.*, 2012).

This study indicates that there are numerous enthesal traits observed throughout the skeleton that show significant, albeit relatively weak, correlations with age. This confirms that age is a significant factor in the development of enthesal changes. Furthermore, it highlights the importance of including these traits in the method despite their limited individual value. According to Lo *et al.* (2015), the significance of a predictor does not necessarily indicate its predictive capability. This supports the idea of using traits that may have a low correlation with age but have been found to be effective predictors of age when considered together, as suggested by Milner and Boldsen (2012).

There were significant differences between males and females in the weight and length of the tibia, humerus, and innominate. When the degree of expression of traits is considered, there are significant differences between males and females for all relative weights, actual weights, and lengths. The differences related to bony dimensions are consistent with known sexual dimorphism (Wilczak, 1998; Vance *et al.*, 2010; Weiss *et al.*, 2012).

Similarly to the enthesal traits, the vertebral differences are consistent with the findings of Snodgrass (2004), Van der Merwe *et al.* (2006), Watanabe and Terazawa (2006), and Praneatpolgrang *et al.* (2019). They are likely related to differences in the load-bearing capacity of the vertebral column, with males typically presenting with higher body/bone mass than females (Weiss, 2007). Differences in physical activity differences between the sexes are also possible causes (Villotte *et al.*, 2010). The above apply to sex-related differences for the sacral elbow, lipping of the cervical and lumbar vertebrae, and candlewax of the lumbar vertebrae. The difference in the degree of expression of the sacral elbow may be due to sex-specific differences in osteoporosis rates (Linstrom *et al.*, 2009; Richards *et al.*, 2010).

The correlation of age to the rib rim profile and ossification pattern was lower than expected compared to the literature (İşcan *et al.*, 1984, 1985; Hartnett, 2010b). When the male and female correlations are compared, the male correlations are consistent with the literature, while the female correlations are distinctly lower. The low correlation in the female sample is likely causing the reduction in the correlation for the pooled-sexes sample. The female sample's low correlation between age and the rib traits may be due to increased fragility due to a decrease in bone density post-menopause, resulting in breakage of ossifications at the sternal end of the rib. It is also possible that this is an artefact of the relatively small sample and an overabundance of damage to the sternal rib ends due to extensive use in research.

The correlations for rib two specifically were lower than those for ribs three to ten and are likely a result of differing load forces since rib two is an atypical rib and is subject to more superior-inferior movement than expansive movement as seen in ribs three to ten (White and Folkens, 2005). The low correlation observed for the first rib is likely due to differing load forces and articulations as an atypical rib, much like rib two (White and Folkens, 2005), even more so than the second rib. The trait tested for rib one is more related to cartilage ossification, like for ribs 2-10, than what has been reported in the literature for rib one, such as changes to the rib head and tubercle (Kunos *et al.*, 1999; Kurki, 2005; DiGangi *et al.*, 2009; Jooste and Steyn, 2023) so direct comparison is not possible.

The lack of correlation between age and shingle ribs may be due to various factors. Boldsen and colleagues (Boldsen *et al.*, 2021a) mention in the scoring manual for TA3

that shingle ribs are a rare trait, so the sample used in this study may not have been large enough to account for this rarity.

The lack of correlation between bone length and weight and age in adults is understandable because long bone growth is complete after epiphyseal fusion, which continues throughout adolescence (Cunningham *et al.*, 2016). Variation in bone weight may be due to severe osteoporosis as well as maceration damage. The lack of correlation for relative weight also makes sense since scoring this trait is highly subjective and impacted by over-maceration. Given that the study sample consisted of black South Africans, it is also likely that the lower frequency of osteoporosis and delayed age at onset influenced the correlation of bone weight with age (Botha *et al.*, 2019). This is especially pertinent given that the reference sample used in TA3 consists of very few black individuals compared to white individuals (Getz, 2017).

The lower strength but still significantly correlated traits such as lipping of the first cervical vertebra, DISH, iliac crest fusion, fifth lumbar vertebra epiphysis fusion, and first sacral vertebral body margin shape among others are consistent with being traits that are affected by age but are not highly informative on their own (Maat *et al.*, 1995; Prescher, 1998; Snodgrass, 2004; Cunningham *et al.*, 2016). The iliac crest and vertebral body symphyses fuse relatively early in life (between the ages of 14 and 27 years for the vertebral body (Cardoso and Ríos, 2011) and 17 to 24 years for the iliac crest (Webb and Suchey, 1985) when compared to many degenerative changes. The primary influence on such traits is age. However, the majority of the sample analysed is well past the threshold where such fusions occur, so their ability to inform an age estimate is limited. Similarly, the cervical vertebrae are not typically subjected to heavily loaded forces and are persistently affected by the head's weight, particularly in the lower cervical vertebrae (Prescher, 1998; Van der Merwe *et al.*, 2006). Diffuse Idiopathic Skeletal Hyperostosis (DISH) is a pathological condition that frequently "attacks" entheses as well as the anterior ligament of the vertebral column (Maat *et al.*, 1995; Burt, 2013; Cawley and Paine, 2015). As an individual with DISH ages, the disease will progress. However, the presence of DISH cannot make a meaningful contribution to an age estimate unless the progression rate of the disease is known and accounted for.

The spheno-occipital synchondrosis had a “poor” to “fair” ($p = 0.154-0.340$, $p = 0.026-0.329$) correlation with age, which is much lower than what has been reported in the literature (Akhlaghi *et al.*, 2010; Shirley and Jantz, 2011; Mahon *et al.*, 2017). This is likely because the study sample primarily consisted of adults, and other studies assessing the fusion of this synchondrosis included younger samples. In the context of adult age estimation, it is not particularly helpful beyond indicating that an individual is “young” or “old”.

The sacral elbow was poorly but significantly correlated with age, which is interesting. When collecting data, the principal investigator observed the sacral elbow in both young and old individuals, and a healed fracture always accompanied it. The probability of a sacral fracture increases with age due to decreased bone density (Linstrom *et al.*, 2009; Richards *et al.*, 2010). However, fractures associated with external trauma are most common in young adults, and the frequency of such fractures decreases with age (Yu *et al.*, 2023). No other studies use fractures of the sacrum as an age-predictive trait.

This study found that the most accurate age estimates were calculated when poorly correlated traits were removed from the analysis despite the decreased correlation between the true and estimated ages. This is supported by Getz’s (2016) indication that the best estimates are not made with a complete set of data points. However, no guidelines for choosing which traits to use were provided. Better age estimates might be attained if the analysis is performed only on the traits with “moderate” correlations with age. A low information trait may improve an estimate. However, including “irrelevant” traits can greatly reduce the accuracy and precision of age estimates by causing over-fitting of the model (Babyak, 2004; Hawkins, 2004; Valsecchi *et al.*, 2019). A potential solution to the problem of low information traits negatively impacting estimates is correctly weighting the contribution of each trait towards the estimate (Babyak, 2004; Hawkins, 2004).

The correlations between true ages and the point estimate ages are above 0.80, indicating that the relationship between true age and estimated age is mostly a product of age. This is a good correlation, and environmental factors and interpersonal variation can likely explain the remaining variation. The minor decrease in the correlation between true and estimated age when ages were re-estimated without poor

correlation traits is unexpected. This may be related to the fact that the samples did not all have the same poorly correlated traits, so some traits were better predictors in males than females and vice versa. The decrease may also be related to the over-fitting of the model due to the inclusion of many traits with low strengths of correlation with age (Babyak, 2004; Hawkins, 2004; Valsecchi *et al.*, 2019).

5.3 Differences between the left and right sides

The significant differences between the dimensions of the left and right humeri, tibiae, and calcanei may be related to handedness and the corresponding differences in robusticity (Krishan, 2011; Gutnik *et al.*, 2015; Krishan and Kanchan, 2016). It is important to note that the degree and direction of bony asymmetry are complex and multifactorial, with literature presenting many contrasting findings and opinions (Ruff and Jones, 1981; Glassman and Bass, 1986; Macho, 1991; Danforth and Thompson, 2008; Ubelaker and Zarenko, 2012). Theories for the presence of asymmetry include side-dominance in the limbs, causing more robusticity on the dominant side, as well as genetic differences and activity-based changes including enthesal and arthritic changes (Wilczak, 1998; Weiss, 2007; Villotte *et al.*, 2010; Krishan, 2011; Ubelaker and Zarenko, 2012; Gutnik *et al.*, 2015; Krishan and Kanchan, 2016; Palmer *et al.*, 2016).

In this study, there was a significant difference in the degree of expression of the lipping of the scapula's glenoid fossa ($Z = -2.846$, $p = 0.004$) between the left and right sides, as well as a higher correlation with age on the right side ($\rho = 0.676$, $p < 0.001$) than the left ($\rho = 0.611$, $p < 0.001$). This is consistent with the findings of Brennaman and colleagues (Brennaman, 2011; Brennaman *et al.*, 2017), who noted significant asymmetry in the degree of osteoarthritic traits for the hip, knee and shoulder in a North American white sample, particularly between the left and right composite shoulder scores, the right shoulder composite variable had a Spearman's rho of 0.7518 while the left had a rho of 0.7335.

During data collection, the researcher noticed that the left side often presented with more extensive lipping on the joint surfaces of the limbs, particularly the knees, shoulders, and elbows. The reason for this is unclear, but if it is related to activity, then it is consistent with the lower correlation with age on the left side due to the relative decrease of the influence of age on traits such as glenoid fossa lipping, humerus

greater tubercle pits, and acetabulum inferior joint lipping. A similar pattern was noted by Brennaman *et al.* (2017) and is supported by findings of slight left-side dominant asymmetry of osteoarthritis in the upper limb, potentially related to repetitive activity (Palmer *et al.*, 2016). Many other joint surfaces not included in the trait scoring manual exhibited the pattern of more pronounced lipping on the left side, including the trochlea of the ulna, the femoral condyles, and the tibia's epicondyles. It is possible that the sample population was more heavily involved in activities such as manual labour compared to today, promoting the development of "handed" enthesal changes. On the other hand, the same pattern of asymmetric expression of age-related traits is still observed in the forensic anthropology laboratory of the Human Variation and Identification Research Unit at the University of the Witwatersrand.

The asymmetry of the pubic symphysis is vital to take note of because of the lack of standards for dealing with asymmetry in this frequently used age estimation site (Acsádi and Neméskei, 1970; Brooks and Suchey, 1990; Hens *et al.*, 2008; Hartnett, 2010a; Garvin and Passalacqua, 2012; Godde and Hens, 2012; Lottering *et al.*, 2013). The researcher noticed a pattern of asymmetry during the data collection phase, with one side appearing older than the other overall. This has been supported by the significant difference in correlation with age between the left and right sides for the pubic symphyseal collar, relief, and dorsal margin. The right side had a higher correlation than the left for the pubic symphyseal relief (left $\rho = 0.583$, $p < 0.001$; right $\rho = 0.669$, $p < 0.001$) and dorsal margin (left $\rho = 0.469$, $p < 0.001$; right $\rho = 0.594$, $p < 0.001$), but the pubic symphyseal collar (left $\rho = 0.073$, $p = 0.517$; right $\rho = 0.071$, $p = 0.534$) had a slightly higher correlation for the left side than the right. Interestingly, there was no significant difference in the degree of the expression of these traits despite the differences in correlation between left and right. This may be related to the number of phases used for each trait, with the pubic symphyseal collar having two phases, the pubic symphyseal relief having three, and the pubic symphysis dorsal margin four.

With so few phases for each trait, it is possible that the asymmetry visually observed may still fit within the range of a single phase but may be the difference between an early presentation and a late presentation of the same phase. Overbury and colleagues (2009) noted an asymmetry in scoring pairs of pubic symphyses using the Suchey-Brooks (Brooks and Suchey, 1990) method with six phases for the full pubic

symphysis. They took note of both full-phase differences as well as partial-phase differences between the early and late stages of the same phase. The differences observed in this study may be similar but separated into components with fewer phases, which reduces the potential for observing full-phase development.

5.4 Accuracy, inaccuracy, and bias

When all traits were considered, the accuracy of the TA3 method performed the best in the female sample (63.6%). After the poorly correlated traits were removed from the analysis for all three samples, the highest accuracy (70.79%) was in the pooled-sex sample.

The average accuracy of approximately 60% is relatively low as compared to Merrit (2014) for the pubic symphysis, Colarusso *et al.* (2016) for the vertebrae, Meyer *et al.* (2021) for the Complex Method, and Hughes *et al.* (2021) for overall estimate accuracies in forensic cases irrespective of the method used. The overall accuracy was better than İşcan *et al.* (1992) for the pubic symphysis and ribs. However, accuracy was similar to Osborne *et al.*'s (2004) expanded intervals and Igarashi *et al.* (2005) for the auricular surface, Hartnett's (2010b) and Merrit (2014) for the pubic symphysis, and Navega *et al.* (2018) for the acetabulum.

The higher accuracy for more central age cohorts is consistent with the literature, which frequently finds higher accuracies for middle-aged individuals as opposed to young adults and particularly old adults (Godde and Hens, 2012; Jooste *et al.*, 2016; Getz and Galimany, 2022), as is the increase in bias for age cohorts closer to the extremes of the range. The method is highly accurate for ages 35-84, with wide interval widths of 29 to 43.7 years. The narrowest intervals were for individuals below 20 years of age, but their ages were also overestimated with a bias larger than the interval width. Overall, accuracy was relatively low with wide estimated interval widths. Age cohorts with high accuracy also had very wide intervals, while cohorts with narrow intervals had very low accuracy. This method appears to work acceptably for individuals aged 35-84 years but poorly for individuals below the age of 35 years and above the age of 84 years.

When all traits were included, the smallest overall bias was observed in the male sample (0.998 years), and the smallest overall inaccuracy (12.962 years) was found in the female sample. After the poorly correlated traits were removed, the smallest bias

(-0.448 years) and inaccuracy (11.439 years) were found in the female sample. The overall bias value is much smaller than any of the bias values for individual age cohorts, which is vital to take note of. One could be tempted to assess the error of a method based purely on the overall bias value provided, anticipating substantially better performance than is true to life.

It is essential to note the difference between the overall bias and inaccuracy values, which differ by at least a factor of 10, except for the re-estimated male sample. This is because the positive and negative bias values cancel each other out when summed. Therefore, the value calculated does not represent the bias pattern across the age spectrum. If a method yields high levels of bias on both ends of the age spectrum, the overall bias value will be deceptively small, providing a misleading idea of the error in an age estimate using said method. The overall inaccuracy is also not fully representative, but it is far less impacted by large errors at opposite ends of the age spectrum and is a more representative indicator of the average error. Proceeding through increasing age cohorts, the pattern of bias and inaccuracy values is a vital indicator of a method's ability to function within specific age ranges. As such, using only the overall values for bias and inaccuracy when reporting the error of a method is inadequate and misleading.

The average bias greatly improved for the pooled-sex (from 1.067 to 0.940 years) and female (from 1.139 to -0.448 years) samples, but male bias increased from 0.998 to 3.244 years when the poorly correlated traits were removed. The age cohort-specific biases also generally improved, except for the male cohort aged 35-44 years, which increased dramatically from 12.58 to 17 years. The increase in male bias and inaccuracy between the ages of 35 and 44 years may be related to the exclusion of a trait that erroneously appeared to be poorly correlated with age. It is possible that a trait that is informative for this age cohort appeared to be poorly correlated with age due to the study's small sample size. Further investigation of the poorly correlated traits in this study is recommended, particularly for traits close to the threshold for fair correlation.

Inaccuracy also improved but less so. The average accuracy improved by 5-10% overall and across all age cohorts. After the poorly correlated traits were removed the average bias was smaller than that of Meyer *et al.*'s (2021) test of the Complex Method

by Acsádi and Neméskei (1970), Baccino *et al.*'s (1999) test of seven age estimation methods, Colarusso's (2016) test of the Passalacqua (2009) method for the sacrum, San-Millán *et al.*'s (2019) method for the acetabulum, and Muñoz-Silva *et al.*'s (2020) test of Rissech *et al.*'s (2006) method for the acetabulum. However, the average inaccuracies are higher than those of Kunos *et al.* (1999) for the first rib, San-Millán *et al.* (2017) for the acetabulum, Muñoz-Silva *et al.*'s (2020) test of Rissech *et al.*'s (2006) method for the acetabulum, and the higher end of Jooste and Steyn's (2023) method for the first rib.

The pattern of overestimation of the ages of the young and underestimation of the ages of the old is consistent with literature for a wide variety of age estimation methods (İşcan *et al.*, 1992; Russell *et al.*, 1993; Kunos *et al.*, 1999; Rissech *et al.*, 2007; Colarusso, 2016; Jooste *et al.*, 2016; Blatt *et al.*, 2020; Muñoz-Silva *et al.*, 2020; Meyer *et al.*, 2021) however the problem of a missing middle seen by Getz (2016) in the St Brides sample, Navega *et al.* (2018b) in a Portuguese sample, as well as Kim and Algee-Hewitt (2022) in Japanese, Thai, and Filipino samples was not observed. The missing middle in the studies above is typically a result of the overestimation of ages for the middle-aged group. Over-estimation was more prevalent than under-estimation and primarily occurred in the younger cohorts. The bias turning point of age 55-64 is consistent with Igarashi *et al.*'s (2005) turning point of 30-49 and 50-99 years, Calce and Rogers' (2011) of 55 years, and San-Millán *et al.*'s (2019) for males at 50-59, but not with San-Millán *et al.*'s (2017, 2019) female samples at 70 years.

The study sample's estimates suffered from an attraction to the middle, which is a common problem in age estimation methods. Theoretically, the Bayesian approach and large uniform reference sample should avoid this problem (Boldsen *et al.*, 2002; Getz, 2017), but it was insufficient in this study. The difference in this study's attraction to the middle versus the missing middle in Getz's (2016) studies may be attributed to the over-fitting of the model to the data used in development, partially because of the number of traits used (Ousley and Jantz, 2012). It should also be noted that Getz (2016) did not include the pubic symphysis or auricular surface in the initial development of the model. That may have contributed to their observation of a missing middle due to those methods' diagnostic ability in younger individuals (Getz, 2017).

The lack of accounting for correlated variables may have contributed to the attraction to the middle, as seen in this study. Another factor potentially influencing the estimated age distribution is the uniform distribution, especially since the vast majority of the database is based on individuals of European origin (six out of eight sample populations). This introduces a bias towards more European ageing patterns instead of accounting for the variation in a South African black population and other non-European populations. An additional possibility is the reference sample for TA3 containing a few late adolescents, young adults, and old individuals, which would negatively impact the age-estimation ability of TA3 in those age groups.

The Bayesian analysis's inability to increase accuracy and precision in this study is unfortunate but not surprising, with the same reasoning as for the attraction to the middle in this study. The average inaccuracy and bias were slightly higher than Jooste *et al.*'s (2016) test of the previous version of Transition Analysis but were similar to Getz and Galimany's (2022) test of TA3. This study found a lower overall accuracy (70.79%) when poorly correlated traits were removed than Getz and Galimany (2022), who found an accuracy of 84.5% overall. This study's bias distribution is similar to that of Getz and Galimany (2022), who saw over-estimation below the age of 40 years and under-estimation above the age of 80 years, with a mix of positive and negative bias between 40 and 80 years and increasing under-estimation with increasing age.

5.5 Ability to work with incomplete information

No significant differences in age estimates were present between the complete ($p = 0.832$, $p < 0.001$) and incomplete sub-samples (Table 4.12). When a region was removed, the correlation between true age and estimated age decreased slightly, similar to when poorly correlated traits were removed from the original estimate calculations. Interestingly, when a region was removed from the estimate calculation, the sample size used by the TA3 software increased compared to that of a complete dataset. This is likely due to a lower number of individuals within the database with a complete dataset than the number of individuals missing certain regions of the data (Getz and Galimany, 2022).

Performance slightly decreased when the auricular surface ($p = 0.824$, $p < 0.001$) and other traits on the innominate ($p = 0.831$, $p < 0.001$) were removed, but a greater decrease in performance was seen when the pubic symphysis ($p = 0.808$, $p < 0.001$)

was removed. When the vertebrae ($p = 0.785$, $p < 0.001$) and upper limb ($p = 0.801$, $p < 0.001$) were removed, the applicable samples' correlation between true age and estimated age decreased similarly to when the pubic symphysis was removed. This indicates that the vertebrae, upper limb, and pubic symphysis are particularly informative for age estimation with TA3. This is partially consistent with the trait-age correlations in this study; the vertebrae are generally correlated with a strength of fair to moderate, as are the traits on the pubic symphysis apart from the pubic symphyseal collar, the upper limb traits are predominantly “poorly” to “fairly” correlated with age. However, the cumulative effect of many traits with relatively low predictive power may be sufficient to affect the age estimate.

When the ribs ($p = 0.834$, $p < 0.001$) were removed, the correlation between true and estimated age increased slightly, indicating an increase in performance, which is in contrast to most studies suggesting the reliability of ribs as an age indicator (İşcan *et al.*, 1992; Russell *et al.*, 1993; Oettlé and Steyn, 2000; Hartnett, 2010b; Merritt, 2014). A possible reason for this may be the use of all ribs instead of specifically rib four. If rib four is scored as a zero (absent) for the rim profile but rib eight is scored as a one (irregular), then the score recorded in the programme will be one, which frequently occurred in this study. Another possibility is that the inability to score both sides led to an accidental bias. As per convention, it was decided to use the left side when only one side could be scored in the program to be consistent. It is unknown what approach Getz and colleagues (Getz, 2017; Getz *et al.*, 2017; Getz and Galimany, 2022) used when considering bilateral traits because Getz's (2016) PhD thesis and the presentation by Getz and Byrnes (2021) utilised opposite sides of bilateral traits. Additionally, the test by Getz *et al.* (2017) uses a different reference sample to the test by Getz and Byrnes (2021), while the test by Getz and Galimany (2022) does not state the sample composition or how they handled bilateral elements. Given the uncertainty with how the rib statistics were carried out, there may be an error in the weighting of the bilateral rib traits with only one scoring input in the program (see Fig. 4.14).

This method mostly functioning even when an entire region was absent is consistent with Getz's (2016) testing with a reduced dataset. Given the difficulty in finding an individual with every single trait present out of the 141 traits, using a smaller number of traits for an estimate is preferable because of the larger sample size used for said estimation. Additionally, the relative functionality of the method when missing an entire

region is advantageous as it is not uncommon for forensic cases to be missing skeletal elements or severely damaged (Baliso *et al.*, 2023).

5.6 Problems identified during data collection and analysis

In the case of the ischial tuberosity bumps, the location description was physically impossible due to ligament attachment configuration, which necessitated a re-defining of the trait location before the scoring was possible (the new location definition and reference image can be found in Appendix C, Fig. C.31(a)). The trait was redefined after consultation with teaching textbooks, atlases, and the study supervisors. However, it may have influenced results since it is unclear if the area decided upon was consistent with the area used to develop the method.

There were two points of uncertainty for scoring traits on the sacrum. First is the inclusion of the sacral elbow, as it is a trauma-based trait. It is also very subjective, and determining whether there is a sharp interruption in the angle of the sacrum is very difficult given the position of the fracture compared to the auricular surfaces of the sacrum. Secondly, the fusion of sacral elements one and two was scored as open if there was an uninterrupted space of more than one centimetre, and fusing if the space was narrower. No indication was made for when there were multiple gaps, each of which was narrower than one centimetre, whose widths added up to one centimetre or more.

The scoring manual's definition of DISH is inconsistent with the current clinical and general forensic understanding of the pathology pattern, specifically the unilateral nature of DISH's ossification (Maat *et al.*, 1995; Cawley and Paine, 2015). In addition, some traits present in the scoring manual were rare/absent in the test sample, including the parietal depression, sternum ridge, and shingle ribs. This may be related to population differences or an artefact of this study's relatively small sample size. The humerus epicondyle trait primarily presented as lipping instead of exostoses, but the reason for this is unclear. The pubic symphyseal collar was also rare in the test sample, which may be related to population differences or simply the result of maceration and handling damage.

For the acetabulum articular surface extra bone, there was sometimes additional bone just outside the scoring area but none inside. This led to uncertainty about whether it should be scored, especially since it was difficult to identify the exact position of the

acetabulum rim in cases of extensive lipping. Similar difficulty in identifying the articular surface edge was observed in the scoring of occipital condyle lipping, especially when what could be considered lipping was smooth and not irregular as described for lipping. Traits involving the presence/absence of macroporosity were challenging due to difficulty distinguishing age-related porosity from developmental defects or pathological lesions. When scoring the auricular surface iliac exostoses, the manual mentioned that the user should not confuse developmental texture with exostoses. However, this was not easy to accomplish without a point of reference.

Problems with the software include problems with the graphic user interface (GUI) and statistical issues. Problems with the GUI include an inability to account for supernumerary vertebrae or for DISH that spans across four vertebrae between two regions of the vertebral column (Fig. 5.1). Traits like relative weights and most of the rib traits are bilateral, but the program only provides a single data entry point for scoring, as seen below in Figs. 5.2 and 5.3.

Cervical	Lipping abs >=3mm	Candlewax abs pres	Observable (Max: 10)
Thoracic	Lipping abs >=5mm	Candlewax abs pres	Observable (Max: 24)
Lumbar	Lipping abs >=5mm	Candlewax abs pres	Observable (Max: 10)
Cervical DISH	abs >=4 verts		
Thoracic DISH	abs >=4 verts		
Lumbar DISH	abs >=4 verts		

Fig. 5.1 The GUI for the scoring of vertebral lipping, candlewax, and DISH

Humerus	L & All	R
Humerus weight	norm light	

Fig. 5.2 The GUI for the scoring of the humerus relative weight, showing the lack of bilateral scoring

Sternum & Ribs	L & All	R & All	Scorable
R1 fusion	abs fused	abs fused	
Sternum central dorsal ridge	abs pres		
R2 rim profile	reg irreg	abs cone claw both	 (Max: 2)
R3-10 rim profile	reg irreg	abs cone claw both	 (Max: 16)
R2 shingle rib	norm shingle		 (Max: 2)
R3-10 shingle ribs	norm shingle		 (Max: 16)

Fig. 5.3 The GUI for traits associated with the ribs, showing the lack of bilateral scoring for all relevant rib traits except the fusion of rib one to the manubrium

Occasionally, the point estimate fell outside the 95% bounds, and both bounds were identical. This is presumed to be due to an error in the R code used for the estimation. However, it was interesting to note that these cases were most common when the limbs or poorly correlated traits were removed (Figs. 5.4-5.7).

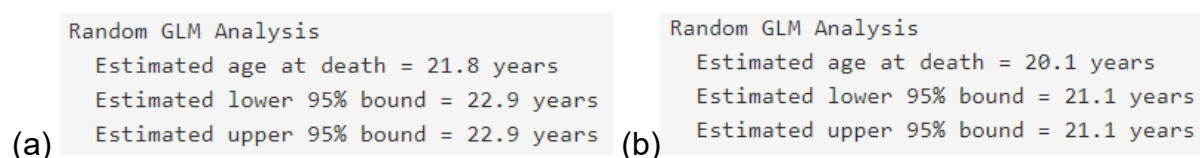


Fig. 5.4 The age estimates for a 16-year-old male when (a) all lower limb traits were removed from the analysis, and (b) all upper limb traits were removed from the analysis

```

Random GLM Analysis
Estimated age at death = 21.8 years
Estimated lower 95% bound = 23.4 years
Estimated upper 95% bound = 23.4 years

```

Fig. 5.5 The age estimate for a 15-year-old male for whom the male-specific poorly correlated traits were removed

```

Random GLM Analysis
Estimated age at death = 24.2 years
Estimated lower 95% bound = 24.9 years
Estimated upper 95% bound = 24.9 years

```

Fig. 5.6 The age estimate for a 15-year-old female for whom the female-specific poorly correlated traits were removed

```
Random GLM Analysis
Estimated age at death = 21.1 years
Estimated lower 95% bound = 22 years
Estimated upper 95% bound = 22 years
```

Fig. 5.7 The age estimate for a 17-year-old male for whom the pooled-sex sample's poorly correlated traits were removed

5.7 Limitations and future research directions

Multiple challenges were identified during both data collection and analysis. These challenges fell into a range of categories, including inadequate manual contents, inadequate maceration and accumulated handling damage, difficulty distinguishing between developmental defects and degradation in some instances, possible errors in the TA3 software, and problems with the reference sample and this study's sample.

Some traits were particularly challenging to score, such as the roughening of the femur's greater trochanter, ischial tuberosity bumps, S1-2 fusion, extra bone in the articular surface of the acetabulum, DISH, and occipital condyle lipping. This was commonly a result of a slightly different pattern of trait progression in the study sample compared to what was described in the manual (Boldsen *et al.*, 2021a). A revision of some trait definitions, such as the location for the ischial tuberosity bumps and the pattern of fusion for DISH, as well as more reference images of all phases of a trait instead of a focus on the extremes of the presence or absence of a trait, would aid the scoring process immensely. Photographs of the various trait phases seen in this study's sample can be found in Appendix C.

The output for the age estimation of a case consists of a point estimate (the type of which is unclear), upper and lower 95% bounds, an error value, the sample size used, and a correlation value. It is unclear where the error and correlation values are drawn from, and the error value provided is seemingly unrelated to the 95% maximum likelihood boundaries. There is also no way to determine if any trait scores for a case are atypical or outliers aside from manually checking the data by removing a single trait at a time. An indication of atypical traits and potential outliers would be beneficial for estimation since those variables can be removed, and age can be re-estimated.

The instances where the point estimate fell outside the 95% bounds while both bounds were identical occurred in young individuals (under 20 years old), especially when limb and poorly correlated traits were removed (see Figs. 4.15-4.18). This should not be

mathematically possible given the statistical model described by Getz (2016), so the R code must be reviewed to identify the point at which the model is failing and fix the error such that the model parameters are satisfied even in individuals who are at the extremely young end of the age spectrum.

An additional problem encountered when re-estimating the ages of the male sample was a potential over-weighting of the traits associated with the auricular surface. Even when most other traits were scored as absent, if the auricular surface's superior posterior and posterior exostoses were marked as present, the age estimate generated was considerably larger than the individual's true age. This strange pattern of weighting, in addition to the errors in estimated intervals compared to the point estimate described above, indicates difficulty in this method's ability to estimate the age of young individuals. A potential solution to the problem of correctly weighting traits is the performance of a principal component analysis test on the complete list of traits to assess the relative effect of age on each trait.

As stated by Getz (Getz, 2017; Getz and Byrnes, 2021), the TA3 age estimation statistics do not include any correction for correlated traits. This is concerning, given that their own testing showed significant correlations between several traits. Given the lack of correcting statistics, if an individual had identical scores for each side of a bilateral trait, both would be weighted equally, even if one side is more correlated with age than the other. In the cases where a single score in the software represents a bilateral trait, there are no guidelines as to which score should be entered in the event of asymmetrical scores. If an expanded statistical correction for correlated traits similar to that used by Boldsen *et al.* (2002) in the original version of Transition Analysis was applied to TA3, the accuracy of age estimates may improve.

Given the increase in accuracy when poorly correlated traits were removed from the analysis and the minimal decrease in correlation when entire regions were removed, it would be helpful if there was a guide for reducing the number of traits assessed per individual. Getz (2016) found that decreasing the number of traits increased the accuracy of estimates and interval widths but did not provide a way for others to attempt to repeat that elimination process. In Getz's (2016) study, the full range of traits in the scoring manual (Boldsen *et al.*, 2021a) were not assessed, and the testing for reduced numbers of traits did not include the full reference sample. It is uncertain

what the most effective elimination of traits would be at the current time since the scoring system utilises the full range of traits, the analysis uses the full reference sample, and there is no way to reliably repeat Getz's exclusion process.

It would be reasonable to expect that a uniform sample that is emphasised to be "from four continents" (Getz, 2017; Milner *et al.*, 2019; Getz and Byrnes, 2021) would have equal representation from each continent. This is not the case for TA3, and the emphasis on the number of continents is misleading. Out of the eight samples used, there is a single contribution from the African continent, specifically South Africa, a single contribution from Asia, specifically Thailand, and a single contribution from Europe, specifically Portugal. The rest of the reference samples (five out of eight) are from the United States of America, with only one American collection containing some black individuals. The result is that the reference sample for TA3 is incredibly biased towards European-American ageing patterns and cannot account for the variation seen in the different parts of Africa, Asia, and Europe. For the uniform distribution to be genuinely representative, more data must be collected in multiple countries and regions in Africa, Asia, and Europe so that there is an equivalent representation of each continent's variation. Data should also be collected for South American populations and populations in Oceania.

Future research should include larger sample sizes that are more representative of all the South African population groups. Another limitation of TA3 is the time it takes to assess a single individual. In the beginning, scoring a full skeleton took approximately five hours; later, the fastest scoring was approximately an hour. However, the average scoring time was an hour and a half. The time required to assess a single skeleton severely impacts its useability in a forensic context where time is of the essence. Another factor influencing the age distribution of the test sample itself includes the phenomenon of "age heaping", where many individuals in skeletal collections are lumped into ages rounded to the nearest five or ten (i.e. 20, 25, 30). This has been commented on by Konigsberg *et al.* (2008) and appears to be a common problem in collections without accurate and detailed records since their inception.

Chapter 6 Conclusion

This study aimed to assess the accuracy and reliability of the TA3 method in a black South African sample, to inform its validity in the black South African population.

Intra-observer repeatability was generally high, but the inter-observer repeatability was considerably lower. This indicates a high level of subjectivity in the trait-scoring methodology, partially concerning unclear trait definitions and an absence of reference photographs of sufficient detail for scoring multiple traits. This method also requires experience to apply since many of the traits used are uncommon in age estimation methods.

No traits had a strong correlation with age, but some traits had a moderate correlation with age, including many vertebral traits, the fusion of the medial clavicle, scapula glenoid fossa lipping, and most traits on the pubic symphysis. All of the cranial traits were poorly correlated with age, which was to be expected given the poor performance of the cranium in adult age estimation.

There were significant differences between males and females for the correlation of some traits to age and the degree of trait expression, with males typically presenting with larger bone weights and length and more extreme expression of enthesal traits. Significant asymmetry was observed in bone weights and lengths, consistent with handedness, but also in the lipping of the scapula glenoid fossa and multiple traits on the pubic symphysis.

The age estimates calculated by TA3 were strongly correlated with the true ages of the sample, both before and after the removal of poorly correlated traits. When entire regions were removed from the analysis, TA3's age estimates were still strongly correlated with age. The average accuracy of the pooled-sexes sample was approximately 70%, comparable to other age estimation methods but lower than the accuracy reported by Getz and Galimany (2022) in their test of TA3. The highest accuracies were found in the middle-aged cohorts, while accuracy decreased as age approached the extremes of the age spectrum assessed. The most accurate age estimates were made when poorly correlated traits were removed despite the decrease in the correlation between true and estimated age. The female sample was typically more accurate than the pooled-sex and male samples before and after the poorly correlated traits were removed.

The age estimates for this sample suffered from an attraction to the middle, with an overestimation of the ages of young individuals and an underestimation of the ages of older individuals. TA3 performed best in individuals aged 35 to 84 years.

In conclusion, TA3 is sufficiently accurate to estimate the age of black South African individuals between the ages of 35 and 84. However, it cannot accurately estimate age in particularly young or old cohorts. Due to the subjectivity of the scoring method, training is required before the use of TA3 in a practical forensic setting. The TA3 reference sample needs to be considerably expanded for it to be truly representative of a global population, including the South African population. TA3 must also be validated in other South African populations before it can be used in the South African forensic setting.

This study showed that TA3 is neither reliable nor accurate enough for use in a South African forensic context and that the correlations between trait score and age are moderate at best. This study also highlighted problems with the methodology, software, and statistics.

References

- Abrahams, P.H., Spratt, J.D., Loukas, M., Van Schoor, A.N. (2013). *McMinn's & Abrahams' clinical atlas of human anatomy*, 7th ed. Elsevier/Mosby, St. Louis.
- Ackermann, A., Steyn, M. (2014). A test of the Lamendin method of age estimation in South African canines. *Forensic Science International (Online)*. 236 (C), p.192.e1–6. <https://doi.org/10.1016/j.forsciint.2013.12.023>
- Acsádi, G., Neméskeri, J. (1970). *History of human life span and mortality*. Akadémiai Kiadó, Budapest.
- Agarwal, S.C., Grynpas, M.D. (2009). Measuring and interpreting age-related loss of vertebral bone mineral density in a medieval population. *American Journal of Physical Anthropology*. 139 (2), p.244–252. <https://doi.org/10.1002/ajpa.20977>
- Aggarwal, P., Saxena, S., Bansal, P. (2008). Incremental lines in root cementum of human teeth: An approach to their role in age estimation using polarizing microscopy. *Indian Journal of Dental Research: Official Publication of Indian Society for Dental Research*. 19 (4), p.326–30. <https://doi.org/10.4103/0970-9290.44536>
- Akhlaghi, M., Taghaddosinejad, F., Sheikhezadi, A., Valizadeh, B., Rezazadeh Shojaei, S.M. (2010). Age-at-death estimation based on the macroscopic examination of Spheno-occipital sutures. *Journal of Forensic and Legal Medicine*. 17 (6), p.304–308. <https://doi.org/10.1016/j.jflm.2010.04.009>
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*. 18 p.91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>
- Algee-Hewitt, B.F.B. (2017). Appendix B: Age estimation in modern forensic anthropology., in: *Forensic Anthropology: A Comprehensive Introduction*. CRC Press, Boca Raton, pp. 381–419.
- AlQahtani, S. j., Hector, M. p., Liversidge, H. m. (2010). Brief communication: The London atlas of human tooth development and eruption. *American Journal of Physical Anthropology*. 142 (3), p.481–490. <https://doi.org/10.1002/ajpa.21258>
- Alves Cardoso, F., Henderson, C. y. (2010). Enthesopathy formation in the humerus: Data from known age-at-death and known occupation skeletal collections. *American Journal of Physical Anthropology*. 141 (4), p.550–560. <https://doi.org/10.1002/ajpa.21171>
- Ameena, S., Heena, Z., Pramod, J., Deepa, H., Shaimaa (2021). Age estimation with cemental incremental lines in normal and periodontally diseased teeth using phase contrast microscope: an original research. *Egyptian Journal of Forensic Sciences*. 11 (1),. <https://doi.org/10.1186/s41935-021-00254-2>
- Anderson, M.C. (2003). A case of undifferentiated seronegative Spondyloarthropathy (Master of Arts). University of Cincinnati, Indiana.
- Babiyak, M.A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*. 66 (3), p.411–421. <https://doi.org/10.1097/01.psy.0000127692.23278.a9>
- Bacci, N., Nchabeleng, E.K., Billings, B.K. (2018). Forensic age-at-death estimation from the sternum in a black South African population. *Forensic Science International (Online)*. 282 p.233.e1-233.e7. <https://doi.org/10.1016/j.forsciint.2017.11.002>
- Baccino, E., Sinfield, L., Colomb, S., Baum, T.P., Martrille, L. (2014). Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. *Forensic Science International*. 244 p.247–251. <https://doi.org/10.1016/j.forsciint.2014.09.005>

- Baccino, E., Ubelaker, D.H., Hayek, L.-A., Zerilli, A. (1999). Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. *Journal of Forensic Sciences*. 44 (5), p.931–936. <https://doi.org/10.1520/JFS12019J>
- Baliso, A., Heathfield, L.J., Gibbon, V.E. (2023). Forensic human identification: retrospective investigation of anthropological assessments in the Western Cape, South Africa. *International Journal of Legal Medicine*. 137 (3), p.793–807. <https://doi.org/10.1007/s00414-022-02870-6>
- Bascou, A., Dubourg, O., Telmon, N., Dedouit, F., Saint-Martin, P., Savall, F. (2021). Age estimation based on computed tomography exploration: a combined method. *International Journal of Legal Medicine*. 135 (6), p.2447–2455. <https://doi.org/10.1007/s00414-021-02666-0>
- Bassed, R.B., Briggs, C., Drummer, O.H. (2010). Analysis of time of closure of the spheno-occipital synchondrosis using computed tomography. *Forensic Science International*. 200 (1–3), p.161–164. <https://doi.org/10.1016/j.forsciint.2010.04.009>
- Becker, J., Chan, C., Leeper, T.J. (2023). rio. A Swiss-Army Knife for Data I/O.
- Becker, S.K. (2020). Osteoarthritis, entheses, and long bone cross-sectional geometry in the Andes: Usage, history, and future directions. *International Journal of Paleopathology*. 29 p.45–53. <https://doi.org/10.1016/j.ijpp.2019.08.005>
- Belanger, T.A., Rowe, D.E. (2001). Diffuse Idiopathic Skeletal Hyperostosis: Musculoskeletal Manifestations: *Journal of the American Academy of Orthopaedic Surgeons*. 9 (4), p.258–267. <https://doi.org/10.5435/00124635-200107000-00006>
- Belcastro, M.G., Pietrobelli, A., Rastelli, E., Iannuzzi, V., Toselli, S., Mariotti, V. (2019). Variations in epiphyseal fusion and persistence of the epiphyseal line in the appendicular skeleton of two identified modern (19th–20th c.) adult Portuguese and Italian samples. *American Journal of Physical Anthropology*. 169 (3), p.448–463. <https://doi.org/10.1002/ajpa.23839>
- Benjamin, M., Toumi, H., Ralphs, J.R., Bydder, G., Best, T.M., Milz, S. (2006). Where tendons and ligaments meet bone: attachment sites ('entheses') in relation to exercise and/or mechanical load. *Journal of Anatomy*. 208 (4), p.471–490. <https://doi.org/10.1111/j.1469-7580.2006.00540.x>
- Bethard, J.D. (2005). A Test of the Transition Analysis Method for Estimation of Age-at-Death in Adult Human Skeletal Remains. University of Tennessee, Knoxville.
- Black, S., Scheuer, L. (1996). Age Changes in the Clavicle: from the Early Neonatal Period to Skeletal Maturity. *International Journal of Osteoarchaeology*. 6 (5), p.425–434. [https://doi.org/10.1002/\(SICI\)1099-1212\(199612\)6:5<425::AID-OA287>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1099-1212(199612)6:5<425::AID-OA287>3.0.CO;2-U)
- Blatt, S., Petersen, K., Michael, A. (2020). Estimating Age of Mature Adults from Degeneration of the Medial End of the Clavicle in an African American Sample. *Forensic Anthropology*. 3 (1), p.39–49. <https://doi.org/10.5744/fa.2020.1005>
- Bocquet-Appel, J.-P., Masset, C. (1982). Farewell to Paleodemography. *Journal of Human Evolution*. (11), p.321–333.
- Boldsen, J.L., Milner, G.R. (2016). Transition analysis age estimation: skeletal scoring manual, Fordisc version 1.02.
- Boldsen, J.L., Milner, G.R. (2012). Skeletal Age Estimation: Where we are and where we should go., in: A Companion to Forensic Anthropology, Blackwell Companions to Anthropology. Wiley-Blackwell, Malden, MA, pp. 224–238.

- Boldsen, J.L., Milner, G.R., Konigsberg, L.W., Wood, J.W. (2002). Transition analysis: a new method for estimating age from skeletons., in: Hoppa, R.D., Vaupel, J.W. (Eds.), *Paleodemography*. Cambridge University Press, pp. 73–106. <https://doi.org/10.1017/CBO9780511542428.005>
- Boldsen, J.L., Ousley, S.D., Milner, G.R., Getz, S.M., Weise, S., Tarp, P. (2021a). Transition Analysis 3 (TA3) Trait Manual Public Distribution Draft 1.0.
- Boldsen, J.L., Ousley, S.D., Milner, G.R., Getz, S.M., Weise, S., Tarp, P. (2021b). Transition Analysis 3.
- Botha, D., Bhagwandin, A., Lynnerup, N., Steyn, M. (2018). The use of stereological methods in the histomorphometric assessment of bone for age-at-death estimation. *Forensic Science International (Online)*. 290 p.353.e1-353.e7. <https://doi.org/10.1016/j.forsciint.2018.06.025>
- Botha, D., Lynnerup, N., Steyn, M. (2019). Age estimation using bone mineral density in South Africans. *Forensic Science International*. 297 p.307–314. <https://doi.org/10.1016/j.forsciint.2019.02.020>
- Botha, D., Pretorius, S., Myburgh, J., Steyn, M. (2016). Age estimation from the acetabulum in South African black males. *International Journal of Legal Medicine*. 130 (3), p.809–817. <https://doi.org/10.1007/s00414-015-1299-7>
- Brennaman, A.L. (2011). Examination of osteoarthritis for age-at-death estimation in a modern population (Master of Science). Boston University, Massachusetts.
- Brennaman, A.L., Love, K.R., Bethard, J.D., Pokines, J.T. (2017). A Bayesian Approach to Age-at-Death Estimation from Osteoarthritis of the Shoulder in Modern North Americans. *Journal of Forensic Sciences*. 62 (3), p.573–584. <https://doi.org/10.1111/1556-4029.13327>
- Brickley, M., Ives, R. (2008). *The bioarchaeology of metabolic bone disease*. Elsevier Academic Press, Oxford.
- Brooks, S., Suchey, J.M. (1990). Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*. 5 (3), p.227–238. <https://doi.org/10.1007/BF02437238>
- Brownlee, J. (2017). What is the Difference Between Test and Validation Datasets? Machine Learning Mastery. URL <https://machinelearningmastery.com/difference-test-validation-datasets/> (accessed 1.24.22).
- Buckberry, J. (2015). The (mis)use of adult age estimates in osteology. *Annals of Human Biology*. 42 (4), p.321–331. <https://doi.org/10.3109/03014460.2015.1046926>
- Buckberry, J.L., Chamberlain, A.T. (2002). Age estimation from the auricular surface of the ilium: A revised method. *American Journal of Physical Anthropology*. 119 (3), p.231–239. <https://doi.org/10.1002/ajpa.10130>
- Bullock, M., Márquez, L., Hernández, P., Ruíz, F. (2013). Paleodemographic age-at-death distributions of two Mexican skeletal collections: A comparison of transition analysis and traditional aging methods. *American Journal of Physical Anthropology*. 152 (1), p.67–78. <https://doi.org/10.1002/ajpa.22329>
- Burr, D.B., Martin, R.B., Schaffler, M.B., Jurmain, R.D., Harner, E.J., Radin, E.L. (1983). Osteoarthritis: Sex-specific relationship to osteoporosis. *American Journal of Physical Anthropology*. 61 (3), p.299–303. <https://doi.org/10.1002/ajpa.1330610304>
- Burt, N.M. (2013). *Identification and interpretation of joint disease in paleopathology and forensic anthropology*. Charles C. Thomas, Publisher, Ltd, Springfield, Illinois.

- Calce, S.E. (2012). A new method to estimate adult age-at-death using the acetabulum. *American Journal of Physical Anthropology*. 148 (1), p.11–23. <https://doi.org/10.1002/ajpa.22026>
- Calce, S.E., Rogers, T.L. (2011). Evaluation of Age Estimation Technique: Testing Traits of the Acetabulum To Estimate Age at Death in Adult Males*. *Journal of Forensic Sciences*. 56 (2), p.302–311. <https://doi.org/10.1111/j.1556-4029.2011.01700.x>
- Campanacho, V., Santos, A.L. (2013). Comparison of the Entheseal Changes of the os coxae of Portuguese Males (19th–20th centuries) with Known Occupation. *International Journal of Osteoarchaeology*. 23 (2), p.229–236. <https://doi.org/10.1002/oa.2297>
- Cardoso, H.F.V. (2008). Age estimation of adolescent and young adult male and female skeletons II, epiphyseal union at the upper limb and scapular girdle in a modern Portuguese skeletal sample. *American Journal of Physical Anthropology*. 137 (1), p.97–105. <https://doi.org/10.1002/ajpa.20850>
- Cardoso, H.F.V., Abrantes, J., Humphrey, L.T. (2014). Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period. *International Journal of Legal Medicine*. 128 (5), p.809–824. <https://doi.org/10.1007/s00414-013-0925-5>
- Cardoso, H.F.V., Gomes, J., Campanacho, V., Marinho, L. (2013). Age estimation of immature human skeletal remains using the post-natal development of the occipital bone. *International Journal of Legal Medicine*. 127 (5), p.997–1004. <https://doi.org/10.1007/s00414-013-0818-7>
- Cardoso, H.F.V., Ríos, L. (2011). Age estimation from stages of epiphyseal union in the presacral vertebrae. *American Journal of Physical Anthropology*. 144 (2), p.238–247. <https://doi.org/10.1002/ajpa.21394>
- Cardoso, H.F.V., Spake, L., Humphrey, L.T. (2017). Age estimation of immature human skeletal remains from the dimensions of the girdle bones in the postnatal period. *American Journal of Physical Anthropology*. 163 (4), p.772–783. <https://doi.org/10.1002/ajpa.23248>
- Cawley, W.D., Paine, R.R. (2015). Skeletal indicators of reactive arthritis: A case study comparison to other skeletal conditions, such as rheumatoid arthritis, ankylosing spondylitis, ankylosing sero-negative SpA, and DISH. *International Journal of Paleopathology*. 11 p.70–74. <https://doi.org/10.1016/j.ijpp.2015.10.001>
- Chan, A.H.W., Crowder, C.M., Rogers, T.L. (2007). Variation in cortical bone histology within the human femur and its impact on estimating age at death. *American Journal of Physical Anthropology*. 132 (1), p.80–88. <https://doi.org/10.1002/ajpa.20465>
- Chan, Y.H. (2003). Biostatistics 104: Correlational Analysis. *Singapore Medical Journal*. 44 (12), p.614–619.
- Chandrakanth, H.V., Kanchan, T., Krishan, K., Arun, M., Pramod Kumar, G.N. (2012). Estimation of age from human sternum: an autopsy study on a sample from South India. *International Journal of Legal Medicine*. 126 (6), p.863–868. <https://doi.org/10.1007/s00414-012-0752-0>
- Christensen, A.M., Crowder, C.M. (2009). Evidentiary Standards for Forensic Anthropology. *Journal of Forensic Sciences*. 54 (6), p.1211–1216. <https://doi.org/10.1111/j.1556-4029.2009.01176.x>

- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 20 (1), p.37–46. <https://doi.org/10.1177/001316446002000104>
- Colarusso, T. (2016). A Test of the Passalacqua Age at Death Estimation Method Using the Sacrum. *Journal of Forensic Sciences*. 61 (S1), p.S22–S29. <https://doi.org/10.1111/1556-4029.12967>
- Collins, D.H. (1949). *The pathology of articular and spinal diseases*. Edward Arnold, London.
- Crowder, C., Rosella, L. (2007). Assessment of Intra- and Intercostal Variation in Rib Histomorphometry: Its Impact on Evidentiary Examination*. *Journal of Forensic Sciences*. 52 (2), p.271–276. <https://doi.org/10.1111/j.1556-4029.2007.00388.x>
- Cunningham, C., Scheuer, L., Black, S.M. (2016). *Developmental juvenile osteology*, 2nd ed. Elsevier Academic Press, Amsterdam.
- Curate, F., Navega, D., Cunha, E., Coelho, J. d'Oliveira (2022). DXAGE 2.0 — adult age at death estimation using bone loss in the proximal femur and the second metacarpal. *International Journal of Legal Medicine*. 136 (5), p.1483–1494. <https://doi.org/10.1007/s00414-022-02840-y>
- Curate, F., Perinha, A., Silva, A.M., Cunha, E., Umbelino, C., Nogueira, C. (2019). Metacarpal cortical bone loss and osteoporotic fractures in the Coimbra Identified Skeletal Collection. *International Journal of Osteoarchaeology*. 29 (1), p.73–81. <https://doi.org/10.1002/oa.2717>
- Danforth, M.E., Thompson, A. (2008). An evaluation of determination of handedness using standard osteological measurements. *Journal of Forensic Sciences*. 53 (4), p.777–781. <https://doi.org/10.1111/j.1556-4029.2008.00741.x>
- Dayal, M.R., Kegley, A.D.T., Štrkalj, G., Bidmos, M.A., Kuykendall, K.L. (2009). The history and composition of the Raymond A. Dart Collection of Human Skeletons at the University of the Witwatersrand, Johannesburg, South Africa. *American Journal of Physical Anthropology*. 140 (2), p.324–335. <https://doi.org/10.1002/ajpa.21072>
- De Luca, S., Pacifici, A., Pacifici, L., Polimeni, A., Fischetto, S.G., Velandia Palacio, L.A., Vanin, S., Cameriere, R. (2016). Third molar development by measurements of open apices in an Italian sample of living subjects. *Journal of Forensic and Legal Medicine*. 38 p.36–42. <https://doi.org/10.1016/j.jflm.2015.11.007>
- Dengsong, Y., Shoumin, Y., Gonghao, C. (1988). Study on the depression and thinning in parietal bone. *Acta Anthropologica Sinica*. 7 (03), p.226–230.
- Diedenhofen, B. (2022). cocor. Comparing Correlations.
- DiGangi, E.A., Bethard, J.D., Kimmerle, E.H., Konigsberg, L.W. (2009). A new method for estimating age-at-death from the first rib. *American Journal of Physical Anthropology*. 138 p.164–176.
- Dirkmaat, D.C., Cabo, L.L., Ousley, S.D., Symes, S.A. (2008). New perspectives in forensic anthropology. *American Journal of Physical Anthropology*. 137 (S47), p.33–52. <https://doi.org/10.1002/ajpa.20948>
- Dominguez, V.M., Harden, A.L., Wascher, M., Agnew, A.M. (2020). Rib Variation at Multiple Locations and Implications for Histological Age Estimation*. *Journal of Forensic Sciences*. 65 (6), p.2108–2111. <https://doi.org/10.1111/1556-4029.14520>
- Duangto, P., Iamaroon, A., Prasitwattanaseree, S., Mahakkanukrauh, P., Janhom, A. (2017). New models for age estimation and assessment of their accuracy using

- developing mandibular third molar teeth in a Thai population. *International Journal of Legal Medicine*. 131 (2), p.559–568. <https://doi.org/10.1007/s00414-016-1467-4>
- Dubourg, O., Faruch-Bilfeld, M., Telmon, N., Savall, F., Saint-Martin, P. (2020). Technical note: age estimation by using pubic bone densitometry according to a twofold mode of CT measurement. *International Journal of Legal Medicine*. 134 (6), p.2275–2281. <https://doi.org/10.1007/s00414-020-02349-2>
- Esan, T.A., Schepartz, L.A. (2019). The timing of permanent tooth development in a Black Southern African population using the Demirjian method. *International Journal of Legal Medicine*. 133 (1), p.257–268. <https://doi.org/10.1007/s00414-018-1968-4>
- Evert, L. (2011). Unidentified Bodies in Forensic Pathology Practice in South Africa Demographic and Medico-Legal Perspectives (Master of Science). University of Pretoria, Pretoria.
- Facchini, F., Veschi, S. (2004). Age determination on long bones in a skeletal subadults sample (b-12 years). *Collegium Antropologicum*. 28 (1), p.89–98.
- Falys, C.G., Prangle, D. (2015). Estimating age of mature adults from the degeneration of the sternal end of the clavicle: Aging Using the Sternal End of the Clavicle. *American Journal of Physical Anthropology*. 156 (2), p.203–214. <https://doi.org/10.1002/ajpa.22639>
- Fisher, R.A. (1925). *Statistical methods for research workers*. Oliver & Boyd, Edinburgh.
- Foss, M.V., Byers, P.D. (1972). Bone density, osteoarthritis of the hip, and fracture of the upper end of the femur. *Annals of the Rheumatic Diseases*. 31 (4), p.259–264. <https://doi.org/10.1136/ard.31.4.259>
- Gamer, M., Lemon, J., Fellows, I., Singh, P. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. Various Coefficients of Interrater Reliability and Agreement.
- García-Donas, J.G., Dyke, J., Paine, R.R., Nathana, D., Kranioti, E.F. (2016). Accuracy and sampling error of two age estimation techniques using rib histomorphometry on a modern sample. *Journal of Forensic and Legal Medicine*. 38 p.28–35. <https://doi.org/10.1016/j.jflm.2015.11.012>
- Garvin, H., Passalacqua, N., Uhl, N., Gipson, D., Overbury, R., Caho, L. (2012). Developments in Forensic Anthropology: Age-at-death estimation., in: A Companion to Forensic Anthropology, Blackwell Companions to Anthropology. Wiley-Blackwell, Malden, MA, pp. 202–223.
- Garvin, H.M., Passalacqua, N.V. (2012). Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation: Age Estimation Practices. *Journal of Forensic Sciences*. 57 (2), p.427–433. <https://doi.org/10.1111/j.1556-4029.2011.01979.x>
- Getz, S.M. (2020). The use of transition analysis in skeletal age estimation. *WIREs Forensic Science*. 2 (6), p.e1378. <https://doi.org/10.1002/wfs2.1378>
- Getz, S.M. (2017). Improved Skeletal Age-at-Death Estimation and its Impact on Archaeological Analyses (Doctor of Philosophy). Pennsylvania State University, Pennsylvania.
- Getz, S.M., Byrnes, P.D. (2021). A28 Testing the Traits of TA3: Setting a Baseline for Method Development and Performance., in: Proceedings of the American Academy of Forensic Sciences. Presented at the 73rd Annual Scientific Meeting, American Academy of Forensic Sciences, Virtual, p. 56.

- Getz, S.M., Galimany, J. (2022). The Evaluation of the Transition Analysis 3 (TA3) (Beta Ver. 0.8.0) Age-Estimation Program Using Two International Samples., in: Proceedings of the American Academy of Forensic Sciences. Presented at the A responsive academy, meeting and surpassing the challenges of a modern forensic science world, American Academy of Forensic Sciences, Seattle, Washington, p. 82.
- Getz, S.M., Milner, G.R., Boldsen, J.L. (2017). A89 Rethinking Adult Skeletal Age Estimation: An Expanded Approach to Transition Analysis Using Binary Traits., in: Proceedings of the American Academy of Forensic Sciences. Presented at the 69th Annual Scientific Meeting, American Academy of Forensic Sciences, New Orleans, LS, p. 192.
- Gilbert, B.M., McKern, T.W. (1973). A method for aging the female Os pubis. *American Journal of Physical Anthropology*. 38 (1), p.31–38. <https://doi.org/10.1002/ajpa.1330380109>
- Glassman, D.M., Bass, W.M. (1986). Bilateral asymmetry of long arm bones and jugular foramen: implications for handedness. *Journal of Forensic Sciences*. 31 (2), p.589–595.
- Godde, K., Hens, S.M. (2012). Age-at-death estimation in an Italian historical sample: A test of the Suchey-Brooks and transition analysis methods. *American Journal of Physical Anthropology*. 149 (2), p.259–265. <https://doi.org/10.1002/ajpa.22126>
- Goliath, J.R., Stewart, M.C., Stout, S.D. (2016). Variation in osteon histomorphometrics and their impact on age-at-death estimation in older individuals. *Forensic Science International (Online)*. 262 p.282.e1-282. <https://doi.org/10.1016/j.forsciint.2016.02.053>
- González-Colmenares, G., Botella-López, M.C., Moreno-Rueda, G., Fernández-Cardenete, J.R. (2007). Age Estimation by a Dental Method: A Comparison of Lamendin's and Prince & Ubelaker's Technique. *Journal of Forensic Sciences*. 52 (5), p.1156–1160. <https://doi.org/10.1111/j.1556-4029.2007.00508.x>
- Gustafson, G., Malmö, D.O. (1950). Age Determinations on Teeth. *The Journal of the American Dental Association*. 41 (1), p.45–54. <https://doi.org/10.14219/jada.archive.1950.0132>
- Gutnik, B., Skurvydas, A., Zuoza, A., Zuoziene, I., Mickevičienė, D., Alekrinskis, A., Nash, D. (2015). Evaluation of Bilateral Asymmetry Between Upper Limb Masses in Right-Handed Young Adults of Both Sexes. *Perceptual and motor skills*. 120 (3), p.804–815. <https://doi.org/10.2466/25.10.PMS.120v16x3>
- Hagelthorn, C.L., Alblas, A., Greyling, L. (2019). The accuracy of the Transition Analysis of aging on a heterogenic South African population. *Forensic Science International*. 297 p.370.e1-370.e5. <https://doi.org/10.1016/j.forsciint.2019.02.012>
- Hartnett, K.M. (2010a). Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample—Part I: Pubic Bone*,†. *Journal of Forensic Sciences*. 55 (5), p.1145–1151. <https://doi.org/10.1111/j.1556-4029.2010.01399.x>
- Hartnett, K.M. (2010b). Analysis of Age-at-Death Estimation Using Data from a New, Modern Autopsy Sample—Part II: Sternal End of the Fourth Rib*,†. *Journal of Forensic Sciences*. 55 (5), p.1152–1156. <https://doi.org/10.1111/j.1556-4029.2010.01415.x>
- Hawkins, D.M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*. 44 (1), p.1–12. <https://doi.org/10.1021/ci0342472>

- Henderson, C.Y., Mariotti, V., Pany-Kucera, D., Villotte, S., Wilczak, C. (2016). The New 'Coimbra Method': A Biologically Appropriate Method for Recording Specific Features of Fibrocartilaginous Entheseal Changes. *International Journal of Osteoarchaeology*. 26 (5), p.925–932. <https://doi.org/10.1002/oa.2477>
- Hens, S.M., Godde, K. (2016). Auricular Surface Aging: Comparing Two Methods that Assess Morphological Change in the Ilium with Bayesian Analyses. *Journal of Forensic Sciences*. 61 (S1), p.S30–S38. <https://doi.org/10.1111/1556-4029.12982>
- Hens, S.M., Rastelli, E., Belcastro, G. (2008). Age Estimation from the Human Os Coxa: A Test on a Documented Italian Collection*. *Journal of Forensic Sciences*. 53 (5), p.1040–1043. <https://doi.org/10.1111/j.1556-4029.2008.00818.x>
- Hoppa, R.D., Vaupel, J.W. (2002). *Paleodemography*. Cambridge University Press, New York.
- Hothorn, T., Winell, H., van de Wiel, M., Zeileis, A. (2023). coin. Conditional Inference Procedures in a Permutation Test Framework.
- Hou, Y., Li, M., Olze, A., Schmidt, S., Schulz, R., Pfeiffer, H., Chen, T., Guo, Y., Schmeling, A. (2019). Dental age estimation using degenerative changes in lower premolars in a northern Chinese population. *Rechtsmedizin*. 29 (5), p.407–414. <https://doi.org/10.1007/s00194-019-00346-3>
- Hughes, C.E., Juarez, C., Yim, A.-D. (2021). Forensic anthropology casework performance: Assessing accuracy and trends for biological profile estimates on a comprehensive sample of identified descendent cases. *Journal of Forensic Sciences*. 66 p.1602–1616. <https://doi.org/10.1111/1556-4029.14782>
- Ibáñez-Gimeno, P., Galtés, I., Jordana, X., Fiorin, E., Manyosa, J., Malgosa, A. (2013). Entheseal Changes and Functional Implications of the Humeral Medial Epicondyle. *International Journal of Osteoarchaeology*. 23 (2), p.211–220. <https://doi.org/10.1002/oa.2299>
- Igarashi, Y., Uesu, K., Wakebe, T., Kanazawa, E. (2005). New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology*. 128 (2), p.324–339. <https://doi.org/10.1002/ajpa.20081>
- İşcan, M.Y., Loth, S.R., Scheuerman, E.H. (1992). Age Assessment from the Sternal End of the Rib and Pubic Symphysis: A Systematic Comparison. *Anthropologie (1962-)*. 30 (1), p.41–44.
- İşcan, M.Y., Loth, S.R., Wright, R.K. (1985). Age Estimation from the Rib by Phase Analysis: White Females. *Journal of Forensic Sciences*. 30 (3), p.853–863. <https://doi.org/10.1520/JFS11018J>
- İşcan, M.Y., Loth, S.R., Wright, R.K. (1984). Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *American Journal of Physical Anthropology*. 65 (2), p.147–156. <https://doi.org/10.1002/ajpa.1330650206>
- İşcan, M.Y., Steyn, M. (2013a). *The human skeleton in forensic medicine*, 3rd ed. Charles C. Thomas, Springfield, Illinois.
- İşcan, M.Y., Steyn, M. (2013b). Skeletal Age., in: *The Human Skeleton in Forensic Medicine*. Charles C. Thomas, Springfield, pp. 59–142.
- Johannesdottir, F., Putman, M.S., Burnett-Bowie, S.-A.M., Finkelstein, J.S., Yu, E.W., Bouxsein, M.L. (2022). Age-Related Changes in Bone Density, Microarchitecture, and Strength in Postmenopausal Black and White Women:

- The SWAN Longitudinal HR-pQCT Study. *Journal of Bone and Mineral Research*. 37 (1), p.41–51. <https://doi.org/10.1002/jbmr.4460>
- Jooste, N., L'Abbé, E.N., Pretorius, S., Steyn, M. (2016). Validation of transition analysis as a method of adult age estimation in a modern South African sample. *Forensic Science International*. 266 p.580.e1-580.e7. <https://doi.org/10.1016/j.forsciint.2016.05.020>
- Jooste, N., Steyn, M. (2023). The first rib as a method of adult age-at-death estimation in a modern South African sample. *International Journal of Legal Medicine*. 137 (3), p.743–752. <https://doi.org/10.1007/s00414-023-02978-3>
- Jurmain, R.D. (1980). The pattern of involvement of appendicular degenerative joint disease. *American Journal of Physical Anthropology*. 53 (1), p.143–150. <https://doi.org/10.1002/ajpa.1330530119>
- Jurmain, R.D. (1977). Stress and the etiology of osteoarthritis. *American Journal of Physical Anthropology*. 46 (2), p.353–365. <https://doi.org/10.1002/ajpa.1330460214>
- Karkhanis, S., Mack, P., Franklin, D. (2014). Age estimation standards for a Western Australian population using the dental age estimation technique developed by Kvaal et al. *Forensic Science International*. 235 p.104.e1-104.e6. <https://doi.org/10.1016/j.forsciint.2013.12.008>
- Kassambara, A. (2023). ggpubr. “ggplot2” Based Publication Ready Plots.
- Katz, D., Suchey, J.M. (1986). Age determination of the male Os pubis. *American Journal of Physical Anthropology*. 69 (4), p.427–435. <https://doi.org/10.1002/ajpa.1330690402>
- Keough, N., L'Abbé, E.N., Steyn, M. (2009). The evaluation of age-related histomorphometric variables in a cadaver sample of lower socioeconomic status: implications for estimating age at death. *Forensic Science International (Online)*. 191 (1), p.114.e1-114.e6. <https://doi.org/10.1016/j.forsciint.2009.07.012>
- Kim, J., Algee-Hewitt, B.F.B. (2022). Age-at-death patterns and transition analysis trends for three Asian populations: Implications for [paleo]demography. *American Journal of Biological Anthropology*. 177 (2), p.207–222. <https://doi.org/10.1002/ajpa.24419>
- Kim, Y.-S., Kim, D.-I., Park, D.-K., Lee, J.-H., Chung, N.-E., Lee, W.-T., Han, S.-H. (2007). Assessment of Histomorphological Features of the Sternal End of the Fourth Rib for Age Estimation in Koreans*. *Journal of Forensic Sciences*. 52 (6), p.1237–1242. <https://doi.org/10.1111/j.1556-4029.2007.00566.x>
- Kimmerle, E.H., Prince, D.A., Berg, G.E. (2008). Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth*. *Journal of Forensic Sciences*. 53 (3), p.594–600. <https://doi.org/10.1111/j.1556-4029.2008.00715.x>
- Klales, A.R. (Ed.) (2020). *Sex Estimation of the Human Skeleton: History, Methods, and Emerging Techniques*. Elsevier Academic Press.
- Konigsberg, L.W., Herrmann, N.P., Wescott, D.J., Kimmerle, E.H. (2008). Estimation and Evidence in Forensic Anthropology: Age-at-Death. *Journal of Forensic Sciences*. 53 (3), p.541–557. <https://doi.org/10.1111/j.1556-4029.2008.00710.x>
- Kranioti, E.F., Bonicelli, A., García-Donas, J.G. (2019). Bone-mineral density: clinical significance, methods of quantification and forensic applications. *Research and Reports in Forensic Medical Science*. 9 p.9–21. <https://doi.org/10.2147/RRFMS.S164933>

- Kranioti, E.F., Michopoulou, E., Tsiminikaki, K., Bonicelli, A., Kalochristianakis, M., Xhemali, B., Paine, R.R., García-Donas, J.G. (2020). Bone histomorphometry of the clavicle in a forensic sample from Albania. *Forensic Science International*. 313 p.110335. <https://doi.org/10.1016/j.forsciint.2020.110335>
- Krishan, K. (2011). Marked limb bilateral asymmetry in an agricultural endogamous population of North India. *American Journal of Human Biology*. 23 (5), p.674–685. <https://doi.org/10.1002/ajhb.21196>
- Krishan, K., Kanchan, T. (2016). Forensic and Anthropological Application of Body Asymmetry: A Comment on Gutnik et al. (2015). *Perceptual and motor skills*. 122 (2), p.578–582. <https://doi.org/10.1177/0031512516640667>
- Kumar, A., Shree, B., Gandhi, P., Joshi, R. (2015). Assessment Of Accuracy Of Gustafson's Age Estimation Criteria In A Population Of Known Age: A Forensic Study. *Journal of Advanced Medical and Dental Sciences Research*. 3 (5), p.S1–S3.
- Kunos, C.A., Simpson, S.W., Russell, K.F., HersHKovitz, I. (1999). First rib metamorphosis: Its possible utility for human age-at-death estimation. *American Journal of Physical Anthropology*. 110 (3), p.303–323. [https://doi.org/10.1002/\(SICI\)1096-8644\(199911\)110:3<303::AID-AJPA4>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1096-8644(199911)110:3<303::AID-AJPA4>3.0.CO;2-O)
- Kurki, H. (2005). Use of the first rib for adult age estimation: a test of one method. *International Journal of Osteoarchaeology*. 15 (5), p.342–350. <https://doi.org/10.1002/oa.788>
- Kvaal, S.I., Kolltveit, K.M., Thomsen, I.O., Solheim, T. (1995). Age estimation of adults from dental radiographs. *Forensic Science International*. 74 (3), p.175–185. [https://doi.org/10.1016/0379-0738\(95\)01760-G](https://doi.org/10.1016/0379-0738(95)01760-G)
- Lamendin, H., Baccino, E., Humbert, J.F., Tavernier, J.C., Nossintchouk, R.M., Zerilli, A. (1992). A Simple Technique for Age Estimation in Adult Corpses: The Two Criteria Dental Method. *Journal of Forensic Sciences*. 37 (5), p.13327J. <https://doi.org/10.1520/JFS13327J>
- Landis, J.R., Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 33 (1), p.159. <https://doi.org/10.2307/2529310>
- Langley, N.R., Gooding, A.F., Tersigni-Tarrant, M.A. (2017). Age estimation methods., in: *Forensic Anthropology: A Comprehensive Introduction*. CRC Press, Boca Raton, pp. 175–194.
- Langley-Shirley, N., Jantz, R.L. (2010). A Bayesian Approach to Age Estimation in Modern Americans from the Clavicle. *Journal of Forensic Sciences*. 55 (3), p.571–583. <https://doi.org/10.1111/j.1556-4029.2010.01089.x>
- Le Cabec, A., Tang, N.K., Ruano Rubio, V., Hillson, S. (2019). Nondestructive adult age at death estimation: Visualizing cementum annulations in a known age historical human assemblage using synchrotron X-ray microtomography. *American Journal of Physical Anthropology*. 168 (1), p.25–44. <https://doi.org/10.1002/ajpa.23702>
- Liao, E.-Y., Wu, X.-P., Deng, X.-G., Huang, G., Zhu, X.-P., Long, Z.-F., Wang, W.-B., Tang, W.-L., Zhang, H. (2002). Age-Related Bone Mineral Density, Accumulated Bone Loss Rate and Prevalence of Osteoporosis at Multiple Skeletal Sites in Chinese Women. *Osteoporosis International*. 13 (8), p.669–676. <https://doi.org/10.1007/s001980200091>
- Lieverse, A.R., Bazaliiskii, V.I., Goriunova, O.I., Weber, A.W. (2013). Lower limb activity in the Cis-Baikal: Entheseal changes among middle holocene siberian

- foragers. *American Journal of Physical Anthropology*. 150 (3), p.421–432. <https://doi.org/10.1002/ajpa.22217>
- Lillie, E.M., Urban, J.E., Lynch, S.K., Weaver, A.A., Stitzel, J.D. (2016). Evaluation of Skull Cortical Thickness Changes With Age and Sex From Computed Tomography Scans. *Journal of Bone and Mineral Research*. 31 (2), p.299–307. <https://doi.org/10.1002/jbmr.2613>
- Lin, L.I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*. 45 (1), p.255. <https://doi.org/10.2307/2532051>
- Linstrom, N.J., Heiserman, J.E., Kortman, K.E., Crawford, N.R., Baek, S., Anderson, R.L., Pitt, A.M., Karis, J., Ross, J.S., Lekovic, G.P., Dean, B.L. (2009). Anatomical and Biomechanical Analyses of the Unique and Consistent Locations of Sacral Insufficiency Fractures. *Spine*. 34 (4), p.309–315. <https://doi.org/10.1097/BRS.0b013e318191ea01>
- Listi, G.A. (2016). The Use of Entheseal Changes in the Femur and Os Coxa for Age Assessment. *Journal of Forensic Sciences*. 61 (1), p.12–18. <https://doi.org/10.1111/1556-4029.12905>
- Listi, G.A., Manhein, M.H. (2012). The Use of Vertebral Osteoarthritis and Osteophytosis in Age Estimation. *Journal of Forensic Sciences*. 57 (6), p.1537–1540. <https://doi.org/10.1111/j.1556-4029.2012.02152.x>
- Lo, A., Chernoff, H., Zheng, T., Lo, S.-H. (2015). Why significant variables aren't automatically good predictors. *Proceedings of the National Academy of Sciences*. 112 (45), p.13892–13897. <https://doi.org/10.1073/pnas.1518285112>
- Lopez-Cerquera, M.A., Casallas, D.A. (2019). Age Estimation in a Colombian Modern Skeletal Sample: A Test of the Transition Analysis Method. *Forensic Anthropology*. 2 (1), p.22–28. <https://doi.org/10.5744/fa.2018.1030>
- Lottering, N., MacGregor, D.M., Meredith, M., Alston, C.L., Gregory, L.S. (2013). Evaluation of the suchey–brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. *American Journal of Physical Anthropology*. 150 (3), p.386–399. <https://doi.org/10.1002/ajpa.22213>
- Lovejoy, C.O. (1985). Dental wear in the Libben population: Its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology*. 68 (1), p.47–56. <https://doi.org/10.1002/ajpa.1330680105>
- Lovejoy, C.O., Meindl, R.S., Pryzbeck, T.R., Mensforth, R.P. (1985). Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*. 68 (1), p.15–28. <https://doi.org/10.1002/ajpa.1330680103>
- Lucy, D., Pollard, A.M. (1995). Further Comments on the Estimation of Error Associated with the Gustafson Dental Age Estimation Method. *Journal of Forensic Sciences*. 40 (2), p.222–227. <https://doi.org/10.1520/JFS15346J>
- Luo, S., Fan, F., Zhang, X., Liu, A., Lin, Y., Cheng, Z., Song, C., Wang, J., Deng, Z., Zhan, M. (2023). Forensic age estimation in adults by pubic bone mineral density using multidetector computed tomography. *International Journal of Legal Medicine*. 137 (5), p.1527–1533. <https://doi.org/10.1007/s00414-023-03067-1>
- Lynnerup, N., Frohlich, B., Thomsen, J.L. (2006). Assessment of age at death by microscopy: Unbiased quantification of secondary osteons in femoral cross sections. *Forensic Science International (Online)*. 159 p.S100–S103. <https://doi.org/10.1016/j.forsciint.2006.02.023>

- Maat, G.J.R., Mastwijk, R.W., van der Velde, E.A. (1995). Skeletal distribution of degenerative changes in vertebral osteophytosis, vertebral osteoarthritis and DISH. *International Journal of Osteoarchaeology*. 5 (3), p.289–298. <https://doi.org/10.1002/oa.1390050308>
- MacDonald, K.V., Sanmartin, C., Langlois, K., Marshall, D.A. (2014). Symptom onset, diagnosis and management of osteoarthritis. *Health Reports*. 25 (9), p.10–7.
- Macho, G.A. (1991). Anthropological evaluation of left-right differences in the femur of southern African populations. *Anthropologischer Anzeiger; Bericht Über Die Biologisch-Anthropologische Literatur*. 49 (3), p.207–216.
- Maggio, A., Franklin, D. (2019). Histomorphometric age estimation from the femoral cortex: A test of three methods in an Australian population. *Forensic Science International (Online)*. 303. <https://doi.org/10.1016/j.forsciint.2019.109950>
- Mahon, T.-J., Friedling, L.J., Gordon, G.M. (2018a). The use of ventral fusion between sacral elements S1 and S2 as an additional age-at-death indicator in a black South African skeletal sample. *Forensic Science International (Online)*. 286 p.267.e1-267.e6. <https://doi.org/10.1016/j.forsciint.2018.02.011>
- Mahon, T.-J., Friedling, L.J., Gordon, G.M. (2018b). Union of the medial clavicular epiphysis in a South African Black skeletal sample. *HOMO*. 69 (5), p.259–265. <https://doi.org/10.1016/j.jchb.2018.09.005>
- Mahon, T.-J., Friedling, L.J., Gordon, G.M. (2017). Spheno-occipital synchondrosis: Examining the degree of fusion in a South African Black skeletal sample. *Forensic Science International (Online)*. 278 p.408.e1-408. <https://doi.org/10.1016/j.forsciint.2017.06.010>
- Mann, H.B., Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*. 18 (1), p.50–60. <https://doi.org/10.1214/aoms/1177730491>
- Maples, W.R., Rice, P.M. (1979). Some Difficulties in the Gustafson Dental Age Estimations. *Journal of Forensic Sciences*. 24 (1), p.168–172. <https://doi.org/10.1520/JFS10804J>
- Marera, D.O., Satyapal, K.S. (2018). Fusion of the Medial Clavicular Epiphysis in the South African and Kenyan Populations: Fusión de la Epífisis Clavicular Medial en las Poblaciones de Sudáfrica y Kenia. *International Journal of Morphology*. 36 (3), p.1101–1107. <https://doi.org/10.4067/S0717-95022018000301101>
- Márquez-Grant, N. (2015). An overview of age estimation in forensic anthropology: perspectives and practical considerations. *Annals of Human Biology*. 42 (4), p.308–322. <https://doi.org/10.3109/03014460.2015.1048288>
- Martrille, L., Irinopoulou, T., Bruneval, P., Baccino, E., Fornes, P. (2009). Age at Death Estimation in Adults by Computer-Assisted Histomorphometry of Decalcified Femur Cortex. *Journal of Forensic Sciences*. 54 (6), p.1231–1237. <https://doi.org/10.1111/j.1556-4029.2009.01178.x>
- Mays, S. (2014). Resorption of mandibular alveolar bone following loss of molar teeth and its relationship to age at death in a human skeletal population. *American Journal of Physical Anthropology*. 153 (4), p.643–652. <https://doi.org/10.1002/ajpa.22465>
- Mazza, B., Silva, A.M. (2023). Analyzing enthesal changes in commingled human remains from Mesolithic and Neolithic periods in Portugal. *International Journal of Osteoarchaeology*. n/a (n/a), p.e3273. <https://doi.org/10.1002/oa.3273>
- McBride, G.B. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. National Institute of Water and Atmospheric Research LTD, Hamilton, New Zealand.

- McDonald, S.W., Miller, J. (2022). When does the spheno-occipital synchondrosis close? *Clinical Anatomy*. 35 (4), p.512–525. <https://doi.org/10.1002/ca.23847>
- McKern, T.W., Stewart, T.D. (1957). Skeletal age changes in young American males analysed from the standpoint of age identification (No. EP-45). Quartermaster research & development command, Massachusetts.
- Meindl, R.S., Lovejoy, C.O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology*. 68 (1), p.57–66. <https://doi.org/10.1002/ajpa.1330680106>
- Meinl, A., Huber, C.D., Tangl, S., Gruber, G.M., Teschler-Nicola, M., Watzek, G. (2008). Comparison of the validity of three dental methods for the estimation of age at death. *Forensic Science International (Online)*. 178 (2), p.96–105. <https://doi.org/10.1016/j.forsciint.2008.02.008>
- Merritt, C.E. (2014). A Test of Hartnett's Revisions to the Pubic Symphysis and Fourth Rib Methods on a Modern Sample. *Journal of Forensic Sciences*. 59 (3), p.703–711. <https://doi.org/10.1111/1556-4029.12380>
- Metsäniitty, M., Waltimo-Sirén, J., Ranta, H., Fieuws, S., Thevissen, P. (2019). Dental age estimation in Somali children and sub-adults combining permanent teeth and third molar development. *International Journal of Legal Medicine*. 133 (4), p.1207–1215. <https://doi.org/10.1007/s00414-019-02053-w>
- Meulendyk, T., Dewar, G., Schillaci, M.A. (2011). The Meul Index: A non-destructive method for estimating the cortical index in skeletal samples. *International Journal of Osteoarchaeology*. 21 (2), p.243–246. <https://doi.org/10.1002/oa.1121>
- Meyer, A., Merwe, A. van der, Steyn, M. (2021). An evaluation of the Acsádi and Nemeskéri Complex Method of adult age estimation in a modern South African skeletal sample. *Forensic Science International (Online)*. 321. <http://dx.doi.org/10.1016/j.forsciint.2021.110740>
- Michener, S., Bell, L.S., Schuurman, N.C., Swanlund, D. (2020). A Method to Interpolate Osteon Volume Designed for Histological Age Estimation Research. *Journal of Forensic Sciences*. 65 (4), p.1247–1259. <https://doi.org/10.1111/1556-4029.14305>
- Michopoulou, E., Nikita, E., Valakos, E.D. (2015). Evaluating the efficiency of different recording protocols for enthesal changes in regards to expressing activity patterns using archival data and cross-sectional geometric properties. *American Journal of Physical Anthropology*. 158 (4), p.557–568. <https://doi.org/10.1002/ajpa.22822>
- Microsoft Corporation (2023). Microsoft Excel.
- Milella, M., Giovanna Belcastro, M., Zollikofer, C.P.E., Mariotti, V. (2012). The effect of age, sex, and physical activity on enthesal morphology in a contemporary Italian skeletal collection. *American Journal of Physical Anthropology*. 148 (3), p.379–388. <https://doi.org/10.1002/ajpa.22060>
- Milenkovic, P., Djukic, K., Djonic, D., Milovanovic, P., Djuric, M. (2013). Skeletal age estimation based on medial clavicle—a test of the method reliability. *International Journal of Legal Medicine*. 127 (3), p.667–676. <https://doi.org/10.1007/s00414-012-0791-6>
- Milner, G.R., Boldsen, J.L. (2012). Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology*. 148 (1), p.98–110. <https://doi.org/10.1002/ajpa.22047>

- Milner, G.R., Ousley, S.D., Boldsen, J.L. (2019). Adult Age Estimated From New Skeletal Traits and Enhanced Computer-Based Transition Analysis (Forensic science R&D Program No. 300659). United States Department of Justice.
- Monum, T., Makino, Y., Prasitwattanaseree, S., Yajima, D., Chiba, F., Torimitsu, S., Hoshioka, Y., Yoshida, M., Urabe, S., Oya, Y., Iwase, H. (2020). Age estimation from ossification of sternum and true ribs using 3D post-mortem CT images in a Japanese population. *Legal Medicine*. 43 p.101663. <https://doi.org/10.1016/j.legalmed.2019.101663>
- Muñoz-Silva, V., Sanabria-Medina, C., Rissech, C. (2020). Application and analysis of the Rissech acetabular adult aging method in a Colombian sample. *International Journal of Legal Medicine*. 134 (6), p.2261–2273. <https://doi.org/10.1007/s00414-020-02422-w>
- Navega, D., Coelho, J. d'Oliveira, Cunha, E., Curate, F. (2018a). DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. *Journal of Forensic Sciences*. 63 (2), p.497–503. <https://doi.org/10.1111/1556-4029.13582>
- Navega, D., Godinho, M., Cunha, E., Ferreira, M.T. (2018b). A test and analysis of Calce (2012) method for skeletal age-at-death estimation using the acetabulum in a modern skeletal sample. *International Journal of Legal Medicine*. 132 (5), p.1447–1455. <https://doi.org/10.1007/s00414-018-1902-9>
- Nawrocki, S.P. (1998). Chapter 13 REGRESSION FORMULAE FOR ESTIMATING AGE AT DEATH FROM CRANIAL SUTURE CLOSURE., in: Reichs, K. (Ed.), *Forensic Osteology: Advances in the Identification of Human Remains*. Charles C. Thomas, Springfield, Illinois, pp. 276–292.
- Nayyar, A., Gayitri, H., Naphade, M., Naphade, U., Qureshi, A., Siddiqui, F., Anwar, S., Shroff, P. (2016). Reliability of Kvaal's and Cameriere's methods of age estimation in a specific populace in central India. *Indian Journal of Oral Health and Research*. 2 (2),. <https://doi.org/10.4103/2393-8692.196121>
- Oettlé, A.C., Steyn, M. (2000). Age Estimation from Sternal Ends of Ribs by Phase Analysis in South African Blacks. *Journal of Forensic Sciences*. 45 (5), p.14831J. <https://doi.org/10.1520/JFS14831J>
- Ohtani, S., Yamamoto, T. (2010). Age Estimation by Amino Acid Racemization in Human Teeth. *Journal of Forensic Sciences*. 55 (6), p.1630–1633. <https://doi.org/10.1111/j.1556-4029.2010.01472.x>
- Osborne, D.L., Simmons, T.L., Nawrocki, S.P. (2004). Reconsidering the Auricular Surface as an Indicator of Age at Death. *Journal of Forensic Sciences*. 49 (5), p.1–7. <https://doi.org/10.1520/JFS2003348>
- Ousley, S.D., Jantz, R.L. (2012). Fordisc 3 and statistical methods., in: Dirkmaat, D.C. (Ed.), *A Companion to Forensic Anthropology*. Wiley-Blackwell, Malden, MA, pp. 311–327.
- Overbury, R.S., Cabo, L.L., Dirkmaat, D.C., Symes, S.A. (2009). Asymmetry of the os pubis: Implications for the Suchey-Brooks method. *American Journal of Physical Anthropology*. 139 (2), p.261–268. <https://doi.org/10.1002/ajpa.20999>
- Palmer, J.L.A., Hoogland, M.H.L., Waters-Rist, A.L. (2016). Activity Reconstruction of Post-Medieval Dutch Rural Villagers from Upper Limb Osteoarthritis and Enthesal Changes. *International Journal of Osteoarchaeology*. 26 (1), p.78–92. <https://doi.org/10.1002/oa.2397>
- Passalacqua, N. (2009). Forensic age-at-death estimation from the human sacrum. *Journal of Forensic Sciences*. 54 (2), p.255–262.
- Pedersen, T.L. (2023). patchwork: The Composer of Plots. The Composer of Plots.

- Plate, T., Heiberger, R. (2016). abind: Combine Multidimensional Arrays.
- Plomp, K.A., Roberts, C.A., Strand Viðarsdóttir, U. (2015). Morphological Characteristics of Healthy and Osteoarthritic Joint Surfaces in Archaeological Skeletons. *International Journal of Osteoarchaeology*. 25 (4), p.515–527. <https://doi.org/10.1002/oa.2319>
- Pradeep, L., Kokila, G., Gopinathan, P.A., Guruswamy, S., Nazir, S.H., Chatterjee, A. (2021). Age Estimation with Cemental Annulation Using Light, Phase Contrast and Polarized Microscopy. *Journal of Microscopy and Ultrastructure*. 9 (2), p.55–60. https://doi.org/10.4103/JMAU.JMAU_3_20
- Praneatpolgrang, S., Prasitwattanaseree, S., Mahakkanukrauh, P. (2019). Age estimation equations using vertebral osteophyte formation in a Thai population: comparison and modified osteophyte scoring method. *Anatomy & Cell Biology*. 52 (2), p.149. <https://doi.org/10.5115/acb.2019.52.2.149>
- Prescher, A. (1998). Anatomy and pathology of the aging spine. *European Journal of Radiology*. 27 (3), p.181–195. [https://doi.org/10.1016/S0720-048X\(97\)00165-4](https://doi.org/10.1016/S0720-048X(97)00165-4)
- Prince, D.A., Kimmerle, E.H., Konigsberg, L.W. (2008). A Bayesian Approach to Estimate Skeletal Age-at-Death Utilizing Dental Wear. *Journal of Forensic Sciences*. 53 (3), p.588–593. <https://doi.org/10.1111/j.1556-4029.2008.00714.x>
- Prince, D.A., Konigsberg, L.W. (2008). New Formulae for Estimating Age-at-Death in the Balkans Utilizing Lamendin's Dental Technique and Bayesian Analysis*. *Journal of Forensic Sciences*. 53 (3), p.578–587. <https://doi.org/10.1111/j.1556-4029.2008.00713.x>
- Prince, D.A., Ubelaker, D.H. (2002). Application of Lamendin's Adult Dental Aging Technique to a Diverse Skeletal Sample. *Journal of Forensic Sciences*. 47 (1), p.107–116. <https://doi.org/10.1520/JFS15209J>
- Prodhan, R., Ubelaker, D.H., Prince, D.A. (2006). H2O Evaluation of Three Methods of Age Estimation From Human Skeletal Remains (Suchey-Brooks, Lamendin, and Two-Step Strategy)., in: Proceedings of the American Academy of Forensic Sciences. Presented at the 58th Annual Scientific Meeting, American Academy of Forensic Sciences, Seattle, Washington.
- R Core Team (2022a). R: A language and environment for statistical computing.
- R Core Team (2022b). stats. The R Stats Package.
- Ramsey, P.H. (1989). Critical Values for Spearman's Rank Order Correlation. *Journal of Educational Statistics*. 14 (3), p.245–253. <https://doi.org/10.3102/10769986014003245>
- Redfield, A. (1970). A new aid to aging immature skeletons: Development of the occipital bone. *American Journal of Physical Anthropology*. 33 (2), p.207–220. <https://doi.org/10.1002/ajpa.1330330206>
- Refai, O. (2019). Entheseal changes in ancient Egyptians from the pyramid builders of Giza—Old Kingdom. *International Journal of Osteoarchaeology*. 29 (4), p.513–524. <https://doi.org/10.1002/oa.2748>
- Richards, A.M., Coleman, N.W., Knight, T.A., Belkoff, S.M., Mears, S.C. (2010). Bone Density and Cortical Thickness in Normal, Osteopenic, and Osteoporotic Sacra. *Journal of Osteoporosis*. 2010 p.504078. <https://doi.org/10.4061/2010/504078>
- Riggs, B.L., Wahner, H.W., Dunn, W.L., Mazess, R.B., Offord, K.P., Melton, L.J. (1981). Differential changes in bone mineral density of the appendicular and axial skeleton with aging: relationship to spinal osteoporosis. *The Journal of Clinical Investigation*. 67 (2), p.328–335. <https://doi.org/10.1172/JCI110039>

- Rinker, T., Kurkiewicz, D. (2019). pacman: Package Management for R.
- Ríos, L., Weisensee, K., Rissech, C. (2008). Sacral fusion as an aid in age estimation. *Forensic Science International (Online)*. 180 (2), p.111.e1-111.e7. <https://doi.org/10.1016/j.forsciint.2008.06.010>
- Rissech, C., Estabrook, G., Cunha, E., Malgosa, A. (2006). Using the Acetabulum to Estimate Age at Death of Adult Males*. *Journal of forensic sciences*. 51 p.213–29. <https://doi.org/10.1111/j.1556-4029.2006.00060.x>
- Rissech, C., Estabrook, G.F., Cunha, E., Malgosa, A. (2007). Estimation of age-at-death for adult males using the acetabulum, applied to four Western European populations. *Journal of Forensic Sciences*. 52 (4), p.774–778. <https://doi.org/10.1111/j.1556-4029.2007.00486.x>
- 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 (1), p.201–12. <https://doi.org/10.1007/s00414-012-0713-7>
- Rissech, C., Marquez-Grant, N., Turbon, D. (2014). New formulae for age estimation of immature skeletal remains. Recomendations for forensic anthropology and osteoarheology. A preliminary contribution. *Estudis d'evolució, etologia i cognició en primats*. p.79–83.
- Rodriguez-Fontenla, C., López-Golán, Y., Calaza, M., Pombo-Suarez, M., Gómez-Reino, J.J., González, A. (2012). Genetic risk load and age at symptom onset of knee osteoarthritis. *Journal of Orthopaedic Research*. 30 (6), p.905–909. <https://doi.org/10.1002/jor.22018>
- Ruff, C.B., Jones, H.H. (1981). Bilateral asymmetry in cortical bone of the humerus and tibia—sex and age factors. *Human Biology*. 53 (1), p.69–86.
- Russell, K.F., Simpson, S.W., Genovese, J., Kinkel, M.D., Meindl, R.S., Lovejoy, C.O. (1993). Independent test of the fourth rib aging technique. *American Journal of Physical Anthropology*. 92 (1), p.53–62. <https://doi.org/10.1002/ajpa.1330920105>
- San-Millán, M., Rissech, C. (2023). Inter-population differences in acetabular senescence: relevance in age-at-death estimation. *International Journal of Legal Medicine*. 137 (3), p.701–719. <https://doi.org/10.1007/s00414-023-02954-x>
- San-Millán, M., Rissech, C., Turbón, D. (2019). Application of the recent SanMillán–Rissech acetabular adult aging method in a North American sample. *International Journal of Legal Medicine*. 133 (3), p.909–920. <https://doi.org/10.1007/s00414-019-02005-4>
- San-Millán, M., Rissech, C., Turbón, D. (2017). New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *International Journal of Legal Medicine*. 131 (2), p.501–525. <https://doi.org/10.1007/s00414-016-1406-4>
- Schaefer, M., Black, S. (2005). Comparison of Ages of Epiphyseal Union in North American and Bosnian Skeletal Material. *Journal of forensic sciences*. 50 p.777–84. <https://doi.org/10.1520/JFS2004497>
- Schauberger, P., Walker, A. (2023). openxlsx: Read, Write and Edit xlsx Files.
- Shane Anderson, A., Loeser, R.F. (2010). Why is osteoarthritis an age-related disease? *Best Practice & Research Clinical Rheumatology, Osteoarthritis*. 24 (1), p.15–26. <https://doi.org/10.1016/j.berh.2009.08.006>
- Shaw, H.M., Benjamin, M. (2007). Structure–function relationships of entheses in relation to mechanical load and exercise. *Scandinavian Journal of Medicine &*

- Science in Sports*. 17 (4), p.303–315. <https://doi.org/10.1111/j.1600-0838.2007.00689.x>
- Shirley, N.R., Jantz, R.L. (2011). Spheno-Occipital Synchondrosis Fusion in Modern Americans*,†. *Journal of Forensic Sciences*. 56 (3), p.580–585. <https://doi.org/10.1111/j.1556-4029.2011.01705.x>
- Shirley, N.R., Ramirez Montes, P.A. (2015). Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component-Based Methods. *Journal of Forensic Sciences*. 60 (1), p.107–111. <https://doi.org/10.1111/1556-4029.12617>
- Signorell, A. (2023). DescTools: Tools for Descriptive Statistics.
- Smith, B.H. (1984). Patterns of molar wear in hunter–gatherers and agriculturalists. *American Journal of Physical Anthropology*. 63 (1), p.39–56. <https://doi.org/10.1002/ajpa.1330630107>
- Snodgrass, J.J. (2004). Sex Differences and Aging of the Vertebral Column. *Journal of Forensic Sciences*. 49 (3), p.1–6. <https://doi.org/10.1520/JFS2003198>
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*. 15 (1), p.72–101. <https://doi.org/10.2307/1412159>
- Statistics South Africa (2021). Mid-year population estimates 2021 (Statistical release No. P0302). Statistics South Africa, Pretoria.
- Statistics South Africa (2019). Victims of crime survey 2018/2019 (Statistical release No. P0341). Statistics South Africa, Pretoria.
- Stevenson, P.H. (1924). Age order of epiphyseal union in man. *American Journal of Physical Anthropology*. 7 (1), p.53–93. <https://doi.org/10.1002/ajpa.1330070115>
- Stewart, T.D. (1958). Rate of Development of Vertebral Osteoarthritis in American Whites and Its Significance In Skeletal Age Identification. *Leech*. 28 (3–5), p.144–151.
- Stott, G.G., Sis, R.F., Levy, B.M. (1982). Cemental Annulation as an Age Criterion in Forensic Dentistry. *Journal of Dental Research*. 61 (6), p.814–817. <https://doi.org/10.1177/00220345820610063401>
- Stout, S.D., Paine, R.R. (1992). Histological age estimation using rib and clavicle. *American Journal of Physical Anthropology*. 87 (1), p.111–115. <https://doi.org/10.1002/ajpa.1330870110>
- Todd, T.W. (1920). Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology*. 3 (3), p.285–334. <https://doi.org/10.1002/ajpa.1330030301>
- Todd, T.W., Lyon Jr., D.W. (1925a). Suture closure—its progress and age relationship. Part IV.—Ectocranial closure in adult males of Negro stock. *American Journal of Physical Anthropology*. 8 (2), p.149–168. <https://doi.org/10.1002/ajpa.1330080203>
- Todd, T.W., Lyon Jr., D.W. (1925b). Cranial suture closure. Its progress and age relationship. Part II.—Ectocranial closure in adult males of white stock. *American Journal of Physical Anthropology*. 8 (1), p.23–45. <https://doi.org/10.1002/ajpa.1330080103>
- Todd, T.W., Lyon Jr., D.W. (1925c). Cranial suture closure. Its progress and age relationship. Part III.—Endocranial closure in adult males of Negro stock. *American Journal of Physical Anthropology*. 8 (1), p.47–71. <https://doi.org/10.1002/ajpa.1330080104>

- Todd, T.W., Lyon Jr., D.W. (1924). Endocranial suture closure. Its progress and age relationship. Part I.—Adult males of white stock. *American Journal of Physical Anthropology*. 7 (3), p.325–384. <https://doi.org/10.1002/ajpa.1330070320>
- Toutin, R., Bilfeld, M.F., Raspaud, C., Bec, C., Telmon, N., Savall, F., Dedouit, F. (2022). Contribution of the use of clavicle bone density in age estimation. *International Journal of Legal Medicine*. 136 (4), p.1017–1025. <https://doi.org/10.1007/s00414-021-02741-6>
- Ubelaker, D.H., Zarenko, K.M. (2012). Can handedness be determined from skeletal remains? A chronological review of the literature. *Journal of forensic sciences*. 57 (6), p.1421–1426. <https://doi.org/10.1111/j.1556-4029.2012.02173.x>
- Usher, B.M. (2002). Reference samples: the first step in linking biology and age in the human skeleton., in: Hoppa, R.D., Vaupel, J.W. (Eds.), *Paleodemography*. Cambridge University Press, New York, pp. 29–47.
- Valsecchi, A., Irurita Olivares, J., Mesejo, P. (2019). Age estimation in forensic anthropology: methodological considerations about the validation studies of prediction models. *International Journal of Legal Medicine*. 133 (6), p.1915–1924. <https://doi.org/10.1007/s00414-019-02064-7>
- Van der Merwe, A.E., Işcan, M.Y., L'Abbè, E.N. (2006). The pattern of vertebral osteophyte development in a South African population. *International Journal of Osteoarchaeology*. 16 (5), p.459–464. <https://doi.org/10.1002/oa.841>
- Vanbelle, S. (2016). A New Interpretation of the Weighted Kappa Coefficients. *Psychometrika*. 81 (2), p.399–410. <https://doi.org/10.1007/s11336-014-9439-4>
- Vance, V.L., Steyn, M., L'Abbé, E.N., Becker, P.J. (2010). A cross-sectional analysis of age related changes in the osteometric dimensions of long bones in modern South Africans of European and African descent. *Forensic Science International*. 199 (1), p.110.e1-110.e9. <https://doi.org/10.1016/j.forsciint.2010.02.036>
- Villotte, S., Assis, S., Cardoso, F.A., Henderson, C.Y., Mariotti, V., Milella, M., Pany-Kucera, D., Speith, N., Wilczak, C.A., Jurmain, R. (2016). In search of consensus: Terminology for enthesal changes (EC). *International Journal of Paleopathology*. 13 p.49–55. <https://doi.org/10.1016/j.ijpp.2016.01.003>
- Villotte, S., Castex, D., Couallier, V., Dutour, O., Knüsel, C.J., Henry-Gambier, D. (2010). Enthesopathies as occupational stress markers: Evidence from the upper limb. *American Journal of Physical Anthropology*. 142 (2), p.224–234. <https://doi.org/10.1002/ajpa.21217>
- Villotte, S., Knüsel, C.J. (2013). Understanding Enthesal Changes: Definition and Life Course Changes. *International Journal of Osteoarchaeology*. 23 (2), p.135–146. <https://doi.org/10.1002/oa.2289>
- Villotte, S., Santos, F. (2023). The effect of age on enthesal changes: A study of modifications at appendicular attachment sites in a large sample of identified human skeletons. *International Journal of Osteoarchaeology*. 33 (3), p.389–401. <https://doi.org/10.1002/oa.3197>
- Watanabe, S., Terazawa, K. (2006). Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. *Legal Medicine*. 8 (3), p.156–160. <https://doi.org/10.1016/j.legalmed.2006.01.001>
- Webb, P.A.O., Suchey, J.M. (1985). Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *American Journal of Physical Anthropology*. 68 (4), p.457–466. <https://doi.org/10.1002/ajpa.1330680402>

- Weiss, E. (2007). Muscle markers revisited: Activity pattern reconstruction with controls in a central California Amerind population. *American Journal of Physical Anthropology*. 133 (3), p.931–940. <https://doi.org/10.1002/ajpa.20607>
- Weiss, E. (2004). Understanding muscle markers: Lower limbs. *American Journal of Physical Anthropology*. 125 (3), p.232–238. <https://doi.org/10.1002/ajpa.10397>
- Weiss, E. (2003). Understanding muscle markers: Aggregation and construct validity. *American Journal of Physical Anthropology*. 121 (3), p.230–240. <https://doi.org/10.1002/ajpa.10226>
- Weiss, E., Corona, L., Schultz, B. (2012). Sex differences in musculoskeletal stress markers: Problems with activity pattern reconstructions. *International Journal of Osteoarchaeology*. 22 (1), p.70–80. <https://doi.org/10.1002/oa.1183>
- White, T., Folkens, P. (2005). *The Human Bone Manual*. Elsevier Academic Press, San Diego.
- Wickham, H. (2020). reshape2. Flexibly Reshape Data: A Reboot of the Reshape Package.
- Wickham, H., Bryan, J. (2023). readxl: Read Excel Files.
- Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. (2023a). ggplot2: Elegant Graphics for Data Analysis.
- Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023b). dplyr: A Grammar of Data Manipulation.
- Wilcox, R.R. (2016). Comparing dependent robust correlations. *British Journal of Mathematical and Statistical Psychology*. 69 (3), p.215–224. <https://doi.org/10.1111/bmsp.12069>
- Wilcox, R.R., Schönbrodt, F.D. (2023a). WRS. A Package of R.R. Wilcox' Robust Statistics Functions.
- Wilcox, R.R., Schönbrodt, F.D. (2023b). WRSedit.
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics bulletin*. 1 (6), p.80–83.
- Wilczak, C.A. (1998). Consideration of sexual dimorphism, age, and asymmetry in quantitative measurements of muscle insertion sites. *International Journal of Osteoarchaeology*. 8 (5), p.311–325. [https://doi.org/10.1002/\(SICI\)1099-1212\(1998090\)8:5<311::AID-OA443>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1212(1998090)8:5<311::AID-OA443>3.0.CO;2-E)
- Willems, G., Lee, S.-S., Uys, A., Bernitz, H., Llano-Pérula, M.C. de, Fieuws, S., Thevissen, P. (2018). Age estimation based on Willems method versus new country-specific method in South African black children. *International Journal of Legal Medicine*. 132 (2), p.599–607. <https://doi.org/10.1007/s00414-017-1686-3>
- Willems, G., Olmen, A.V., Carels, C., Spiessens, B. (2001). Dental Age Estimation in Belgian Children: Demirjian's Technique Revisited. *Journal of Forensic Sciences*. 46 (4), p.893–895. <https://doi.org/10.1520/JFS15064J>
- Winburn, A.P. (2019). Validation of the Acetabulum As a Skeletal Indicator of Age at Death in Modern European-Americans. *Journal of Forensic Sciences*. 64 (4), p.989–1003. <https://doi.org/10.1111/1556-4029.13972>
- Wittschieber, D., Vieth, V., Domnick, C., Pfeiffer, H., Schmeling, A. (2013). The iliac crest in forensic age diagnostics: evaluation of the apophyseal ossification in conventional radiography. *International Journal of Legal Medicine*. 127 (2), p.473–479. <https://doi.org/10.1007/s00414-012-0763-x>
- Wittwer-Backofen, U., Gampe, J., Vaupel, J.W. (2004). Tooth cementum annulation for age estimation: Results from a large known-age validation study. *American*

- Journal of Physical Anthropology*. 123 (2), p.119–129.
<https://doi.org/10.1002/ajpa.10303>
- Wochna, K., Bonikowski, R., Śmigielski, J., Berent, J. (2018). Aspartic acid racemization of root dentin used for dental age estimation in a Polish population sample. *Forensic Science, Medicine and Pathology*. 14 (3), p.285–294.
<https://doi.org/10.1007/s12024-018-9984-8>
- Xanthopoulou, P., Valakos, E., Youlatos, D., Nikita, E. (May). Assessing the accuracy of cranial and pelvic ageing methods on human skeletal remains from a modern Greek assemblage. *Forensic Science International (Online)*. 286 p.266.e1-266.e8. <https://doi.org/10.1016/j.forsciint.2018.03.005>
- Yiu Luk, S., Fai Shum, J.S., Wai Chan, J.K., San Khoo, J.L. (2010). Bilateral thinning of the parietal bones: a case report and review of radiological features. *The Pan African Medical Journal*. 4 p.7.
- Yonemoto, S. (2016). Differences in the effects of age on the development of enthesal changes among historical Japanese populations. *American Journal of Physical Anthropology*. 159 (2), p.267–283.
<https://doi.org/10.1002/ajpa.22870>
- Yoshino, M., Imaizumi, K., Miyasaka, S., Seta, S. (1994). Histological estimation of age at death using microradiographs of humeral compact bone. *Forensic Science International*. 64 (2), p.191–198. [https://doi.org/10.1016/0379-0738\(94\)90231-3](https://doi.org/10.1016/0379-0738(94)90231-3)
- Yu, Y.-H., Tsai, P.-J., Liu, C.-H., Chen, I.-J., Hsu, Y.-H., Chou, Y.-C., Tseng, I.-C. (2023). Causes of Increased Use of Closed Reduction and Internal Fixation for High-Energy-Related Traumatic Sacral Fractures. *World Journal of Surgery*. 47 (4), p.903–911. <https://doi.org/10.1007/s00268-022-06876-4>
- Zampetti, S., Mariotti, V., Radi, N., Belcastro, M.G. (2016). Variation of skeletal degenerative joint disease features in an identified Italian modern skeletal collection. *American Journal of Physical Anthropology*. 160 (4), p.683–693.
<https://doi.org/10.1002/ajpa.22998>

Appendix A Ethical clearance

 UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
---	--

Office of the Deputy Vice-Chancellor (Research & Innovation)

TO: Professor AO Ihunwo
School: Anatomical Sciences
Department:
Medical School
University

E-mail: Amadi.Ihunwo@wits.ac.za

CC: Supervisor: <>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 04/05/2022

REF: R14/49

PROTOCOL NO: W-CBP-220504-01 (This is your ethics application study reference number.
Please quote this reference number in all correspondence relating to this study)

PROJECT TITLE: *Research on cadaveric material*
Research conducted on donated bodies and their tissues
under sections 62-64 under the National Health Act No.61 of 2003
Formerly W-CJ-140604-1

Please find attached the Ethics Waiver Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.



MSWorks2000/Iain0007/ClearScanWaiver.wps

SCHOOL OF ANATOMICAL SCIENCES ETHICS WAIVER CLEARANCE LETTER

Faculty of Health Sciences
School of Anatomical Sciences
University of the Witwatersrand
Johannesburg

Re: In terms of Chapter 8, sections 62-64 of the National Health Act No 61 of 2003 donated bodies and their tissues may be used for, among other purposes, health, and research. Use of such Material is subject only to permission from the responsible person in the School of Anatomical Sciences – the Head or person designed by the Head.

Human Research Ethics Committee (Medical) Clearance Certificate:

W-CBP-220504-01

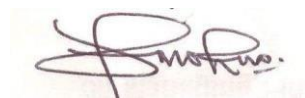
This letter serves to confirm that the Head of School, based in the School of Anatomical Sciences, Faculty of Health Sciences, has reviewed the research proposal entitled:

Assessing the accuracy of Transition Analysis 3.0 for estimating age in a Black South African population

by principal investigator: Ms Jennifer Tempest

and supervisor/collaborators: Dr Anja Meyer and Prof Desiré Brits

and has granted approval for ethics waiver to conduct the abovementioned research study.



Professor Amadi Ihunwo
School of Anatomical Sciences
Faculty of Health Sciences

19 August 2022

Dated

Appendix B Scoring manual

All the traits described in this chapter are derived from the scoring manual by Boldsen *et al.* (Boldsen *et al.*, 2021a) and interpreted in the researcher's own words. All photographs were taken by the researcher unless indicated otherwise

Table B.1 Cranial traits descriptions and scoring

Cranium (Images in Figs B.1-B.3)		
Trait	Location/definition	Scoring
Parietal depression	On parietal bone between superior temporal line & sagittal suture, view posteriorly and superiorly	0 (absent): Parietal contour is typical (outwardly rounded to flat) 1 (present): Concave depression visible on the parietal bone - Often rounded/oval depression, sometimes a long 2-3 cm wide groove superior to the temporal line
Spheno-occipital synchondrosis	At meeting point between pars basilaris and sphenoid bone	0 (open): Space present between the sphenoid and occipital base 1 (closed): Groove separating sphenoid and occipital base completely obliterated
Occipital condyle lipping ¹	Edges of articular surface of occipital condyles	0 (absent): No lipping present or lipping around >50% of the surface edge 1 (present ≥50%): Lip ≥1 mm surrounding ≥50% of the surface but <75%, must be irregular 2 (present ≥75%): Lip ≥1 mm surrounding ≥75% of the surface, must be irregular

¹Mechanical sliding calipers

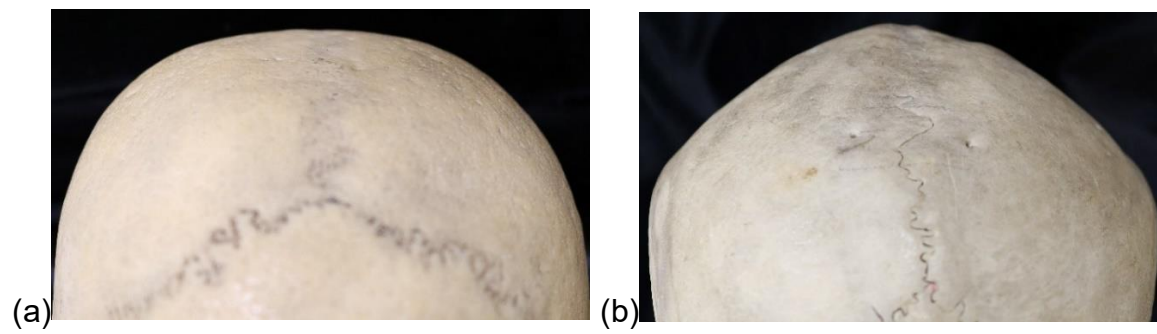


Fig. B.1 Parietal depression (a) 0, (b) 1

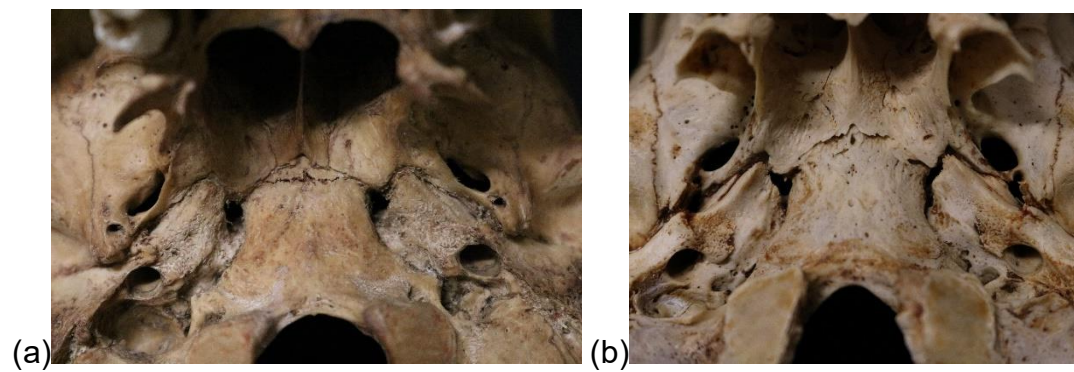


Fig. B.2 Spheno-occipital synchondrosis (a) 0, (b) 1

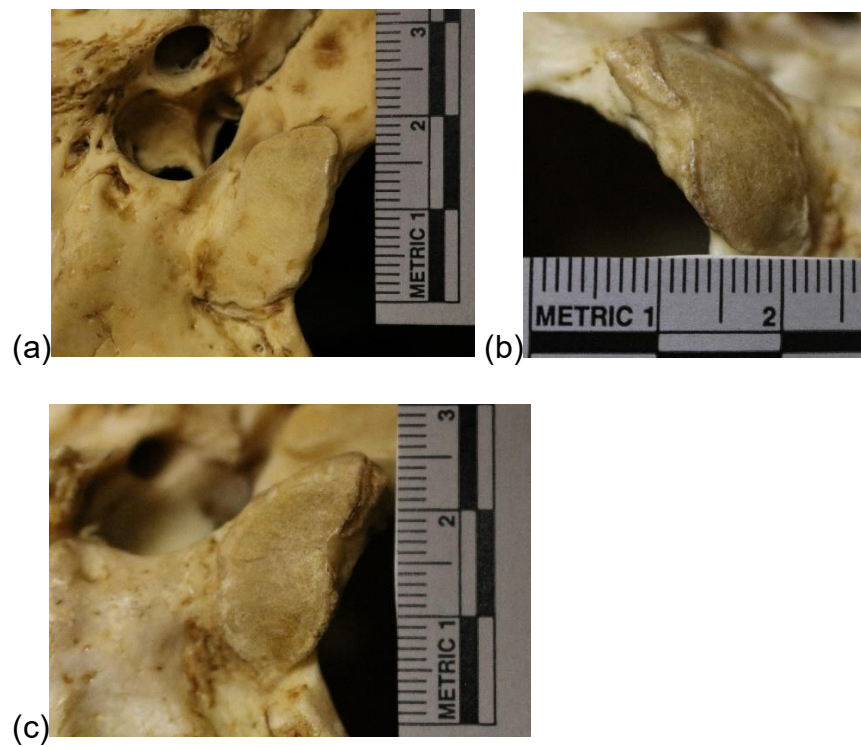


Fig. B.3 Occipital condyle lipping (a) 0, (b) 1, (c) 2

Table B.2 Lower limb traits descriptions and scoring

Lower limb (Images in Figs B.4-B.9)		
Trait	Location	Scoring
Femur fovea margin lipping ¹	Margin of femoral fovea [ligamentum teres attachment]	0 (absent): Margin smooth, rounded, or sharp but still regular 1 (lipped): ≥ 10 mm continuous irregular bone growth along the margin
Femur head surface extra bone ¹	Femoral head except for articular margin	0 (absent): Smooth surface or < 5 mm extra bone present 1 (extra bone 5-9.9 mm): 5-9.9 mm extra bone present 2 (extra bone ≥ 10 mm): ≥ 10 mm extra bone present
Femur greater trochanter roughening ¹	Lateral surface & lateral inferior edge of greater trochanter	0 (absent): Smooth or $< 10 \times 20$ mm triangle roughened 1 (rough triangle): 10×20 mm triangle roughened 2 (rough rectangle): 10×20 mm rectangle roughened 3 (overhang): Irregular growth extends ≥ 1 mm for ≥ 10 mm along margin, overhang present
Femur trochanteric fossa exostoses	Inside trochanteric fossa [obturator externus insertion]	0 (absent): Smooth or maximum of 1 exostosis 1 (present): 2+ exostoses present
Femur trochanteric medial exostoses	Proximal to trochanteric fossa on medial side [obturator internus & gemelli insertion]	0 (absent): Smooth or a maximum of 1 exostosis 1 (present): 2+ exostoses present
Tibia weight relative	Entire tibia	0 (normal): Feels as heavy as expected when first picked up 1 (light): Feels unexpectedly light when first picked up
Tibia weight metric ^{2 3}	Entire tibia	Maximum length measured to nearest millimetre with osteometric board & mass measured on digital scale to nearest gram, maximum length and mass used to approximate bone density
Fibula wings ¹	Lateral surface & anterior edge of lateral malleolus [anterior talo- & tibio-fibular ligaments insertions]	0 (absent): No bony projection present or < 1 mm projection for < 3 mm proximal-distal 1 (present): ≥ 1 mm bony projection ≥ 3 mm long proximal-distal

Lower limb (Images in Figs B.4-B.9)		
Trait	Location	Scoring
Calcaneus weight relative	Entire calcaneus	0 (normal): Feels as heavy as expected when first picked up 1 (light): Feels unexpectedly light when first picked up
Calcaneus weight metric ^{2 3}	Entire calcaneus	Maximum length measured to nearest millimetre with osteometric board & mass measured on digital scale to nearest gram, maximum length and mass used to approximate bone density

¹Mechanical sliding calipers

²Paleo-Tech field osteometric board

³Standard baking scale with sensitivity to 0.1 g

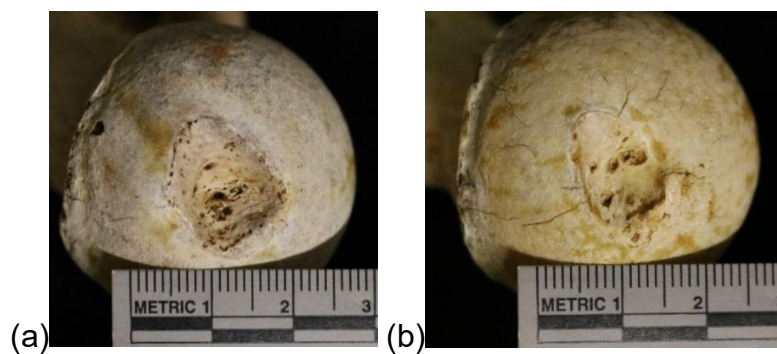


Fig. B.4 Femur fovea margin lipping (a) 0, (b) 1



Fig. B.5 Femur head surface extra bone (a) 0, (b) 1, (c) 2

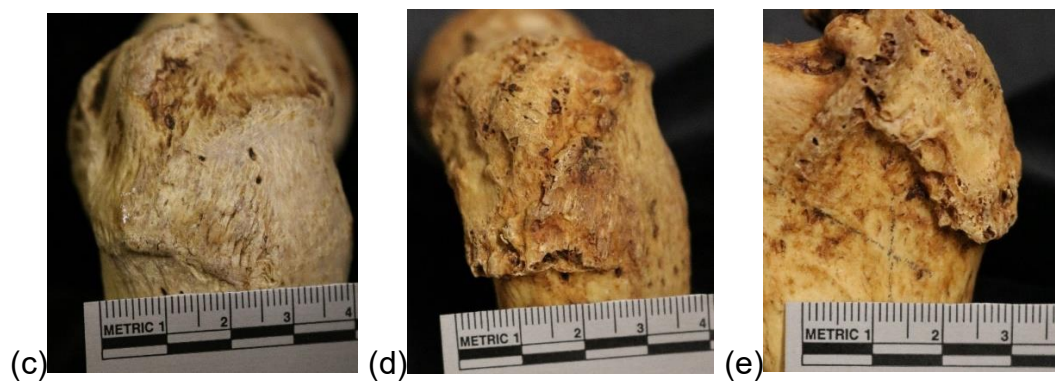
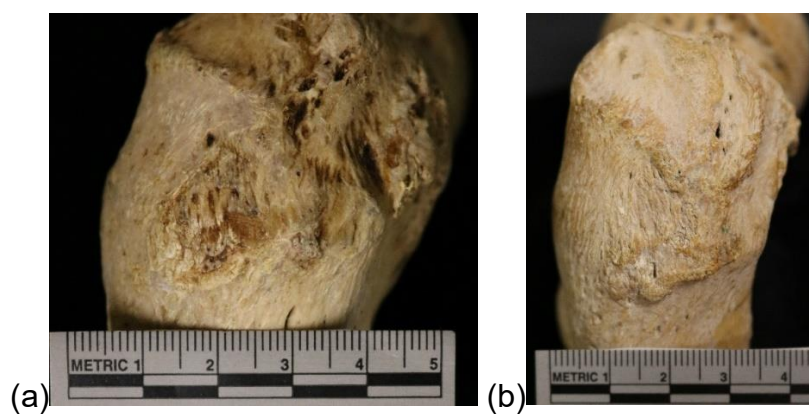


Fig. B.6 Femur greater trochanter roughening (a) 0, (b) 1, (c) 2, (d) 3 lateral view, (e) 3 posterior view

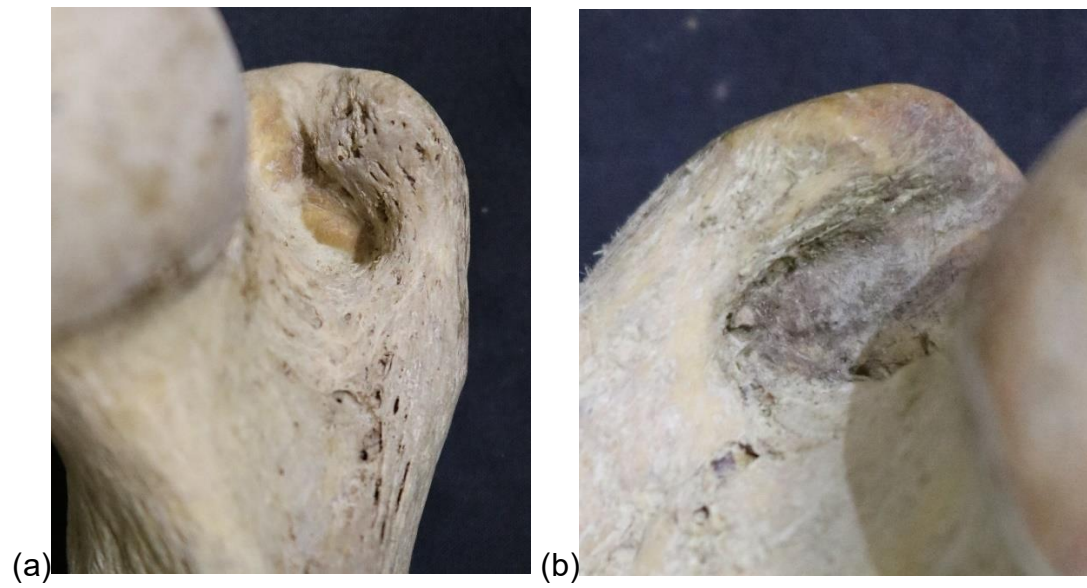


Fig. B.7 Femur trochanteric fossa exostoses (a) 0, (b) 1

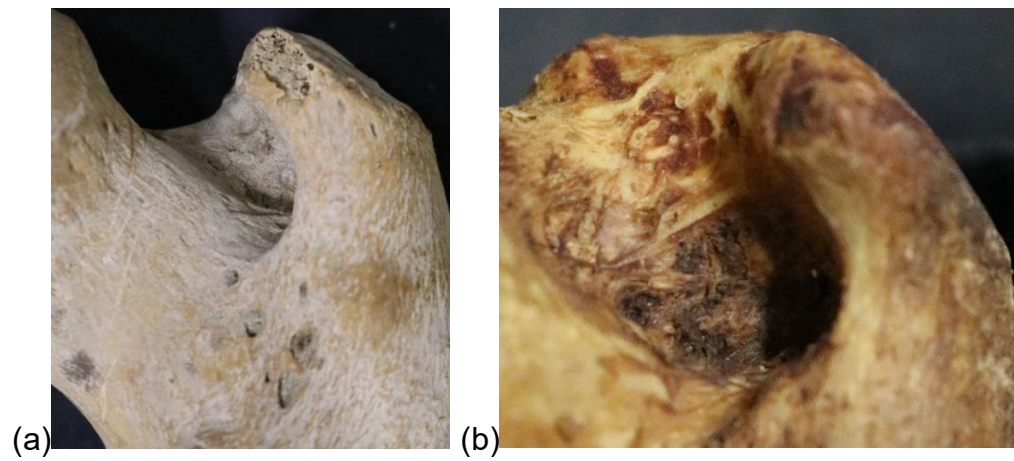


Fig. B.8 Femur trochanteric medial exostoses (a) 0, (b) 1

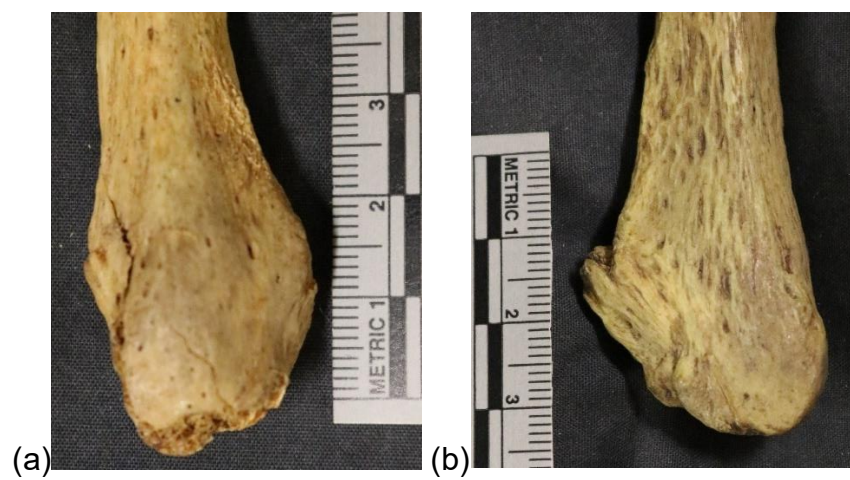


Fig. B.9 Fibula wings (a) 0, (b) 1

Table B.3 Upper limb traits descriptions and scoring

Upper limb (Images in Figs B.10-B.22)		
Trait	Location	Scoring
Clavicle medial epiphysis fusion ¹	Medial surface & edges of clavicle	0 (not fused): Ridged surface, no epiphyseal fusion 1 (partially fused): Smooth epiphysis present on part of surface with marked space between epiphyseal edge and initial bony surface 2 (remnant line): Fusion complete but shallow groove present at point of fusion ≥ 5 mm long 3 (fully fused): Point of fusion not evident
Clavicle medial epiphysis gravel ¹	Medial surface clavicle*	0 (absent): Surface smooth, or < 5 mm irregular bone present 1 (present): ≥ 5 mm irregular bone growth present on surface
Clavicle medial macroporosity ¹	Medial surface clavicle*	0 (absent): Surface smooth, has only microporosity, or < 3 macropores (also if epiphysis is unfused) 1 (macroporosity ≥ 3): ≥ 3 macropores present, minimum size of 1mm 2 (macroporosity $\geq 75\%$): $\geq 75\%$ of the surface covered in macroporosities
Clavicle lateral macroporosity ¹	Lateral articular surface for acromion	0 (absent): Surface smooth, has only microporosity, or < 3 macropores 1 (macroporosity): 3+ macropores present, minimum size of 1 mm
Scapula glenoid fossa lipping ¹	Margins of the glenoid fossa	0 (absent): All edges smooth & rounded 1 (present $< 50\%$): Lipping present on $< 50\%$ of margin 2 (present $\geq 50\%$): Lipping present on $\geq 50\%$ of margin
Humerus weight subjective	Entire humerus	0 (normal): Feels as heavy as expected when first picked up 1 (light): Feels unexpectedly light when first picked up
Humerus weight metric ^{2 3}	Entire humerus	Maximum length measured to nearest millimetre with osteometric board & mass measured on digital scale to nearest gram, maximum length and mass used to approximate bone density
Humerus lesser tubercle bumps	The anterior surface of lesser tubercle of the humerus	0 (absent): Smooth surface or $< 1/3$ of surface bumpy 1 (present): $\geq 1/3$ surface bumpy
Humerus lesser tubercle margin shape	Lateral margin of lesser tubercle of the humerus	0 (round/flat): Smooth transition between lateral margin & anterior surface, no elevation of lateral margin

Upper limb (Images in Figs B.10-B.22)		
Trait	Location	Scoring
		1 (raised): $\geq 50\%$ margin elevated above anterior surface 2 (lipped): Irregular & rough margin distinctly elevated
Humerus greater tubercle pits	The superior surface of greater tubercle of the humerus	0 (absent): No pitting present 1 (present): 2+ macropores on superior surface
Humerus medial epicondyle ¹	Anterior surface & posterior edge of medial epicondyle of humerus	0 (absent): < 5 mm rim and $< 1/2$ surface covered in exostoses 1 (rough): $\geq 1/2$ surface covered in exostoses or ≥ 5 mm irregular rim on posterior edge
Humerus lateral epicondyle ¹	Anterior surface & posterior edge of lateral epicondyle of humerus	0 (absent): < 5 mm rim and $< 1/2$ surface covered in exostoses 1 (rough): $\geq 1/2$ surface covered in exostoses or ≥ 5 mm irregular rim on posterior edge (often more pronounced than medial epicondyle)
Radius tuberosity medial crest ¹	Medial aspect of radial tuberosity [biceps brachii insertion point]	0 (absent): Rounded or sharp edge, minimal lipping present (< 1 mm high or < 10 mm long) 1 (present): Narrow & high protuberance at tendon insertion, ≥ 1 mm high & ≥ 10 mm long
Ulna olecranon spur ¹	Posterior-proximal margin of olecranon process [triceps brachii insertion]	0 (absent): Smooth edge or < 2 mm extra bone 1 (present): ≥ 2 mm spurs present
Trapezium lipping ¹	Border of articular surface for 1 st metacarpal	0 (absent): No lipping or < 2 mm lipping (slight lipping only) 1 (present): ≥ 2 mm lipping extending from 2+ points on margin, also if fused to 1 st metacarpal

*If epiphysis unfused, score as 0

¹Mechanical sliding calipers

²Paleo-Tech field osteometric board

³Standard baking scale with sensitivity to 0.1g



Fig. B.10 Clavicle medial epiphysis fusion (a) 0, (b) 1, (c) 2 (Boldsen *et al.*, 2021a), (d) 3



Fig. B.11 Clavicle medial epiphysis gravel (a) 0, (b) 1

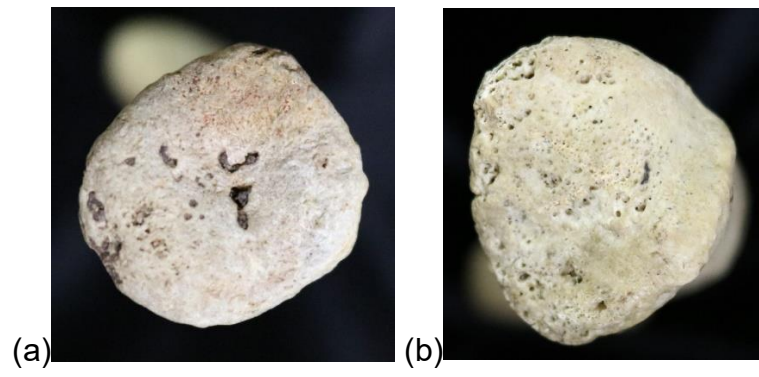


Fig. B.12 Clavicle medial macroporosity (a) 1, (b) 2

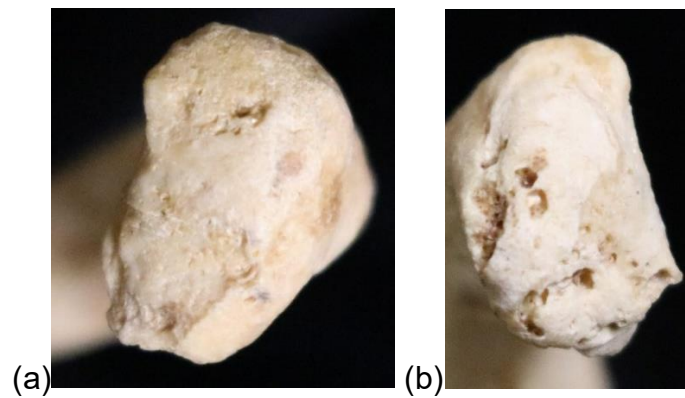


Fig. B.13 Clavicle lateral macroporosity (a) 0, (b) 1



Fig. B.14 Scapula margin lipping (a) 0, (b) 1, (c) 2

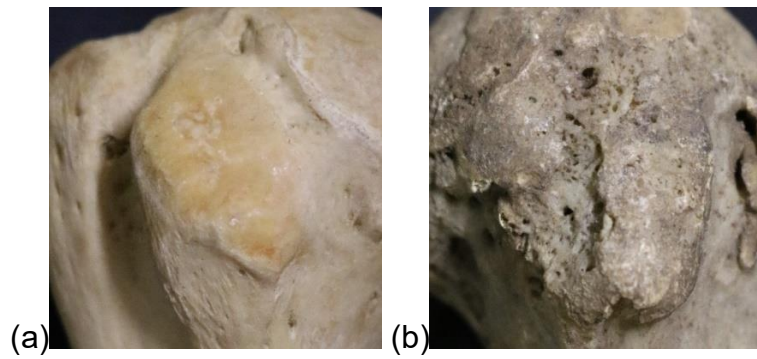


Fig. B.15 Humerus lesser tubercle bumps (a) 0, (b) 1

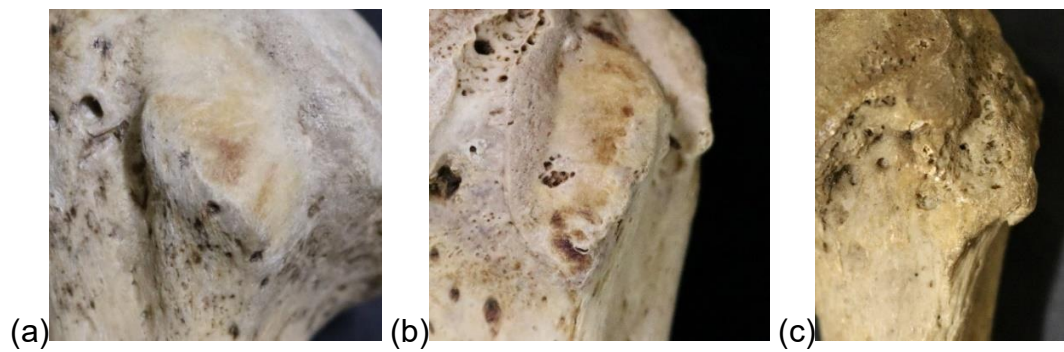


Fig. B.16 Humerus lesser tubercle margin lipping (a) 0, (b) 1, (c) 2

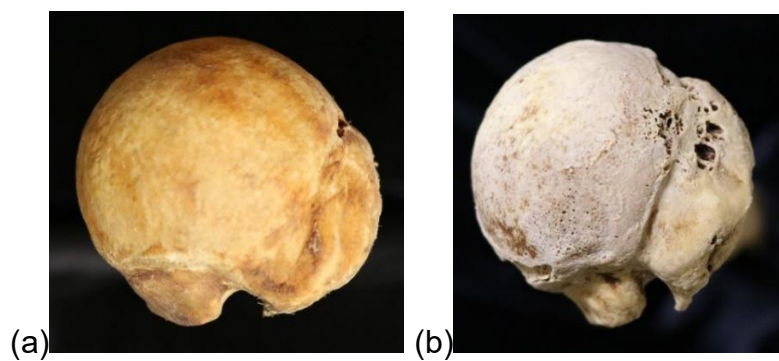


Fig. B.17 Humerus greater tubercle pits (a) 0, (b) 1

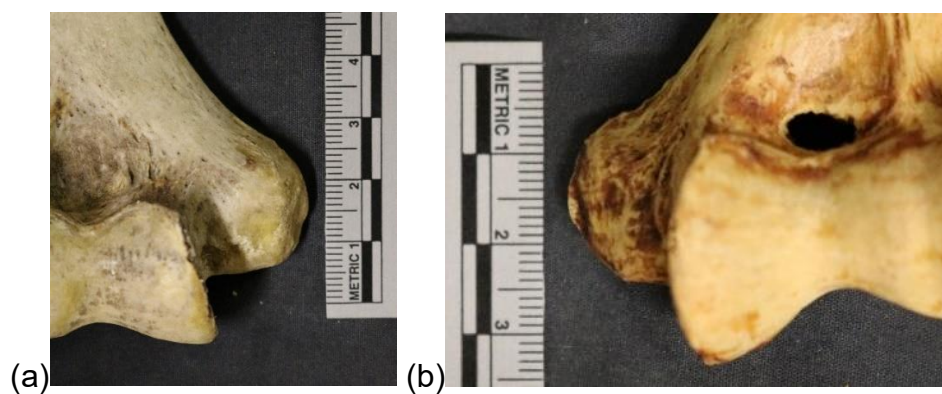


Fig. B.18 Humerus medial epicondyle (a) 0, (b) 1

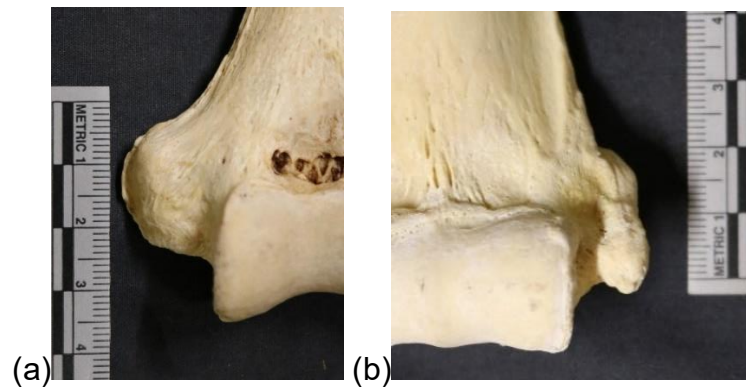


Fig. B.19 Humerus lateral epicondyle (a) 0, (b) 1

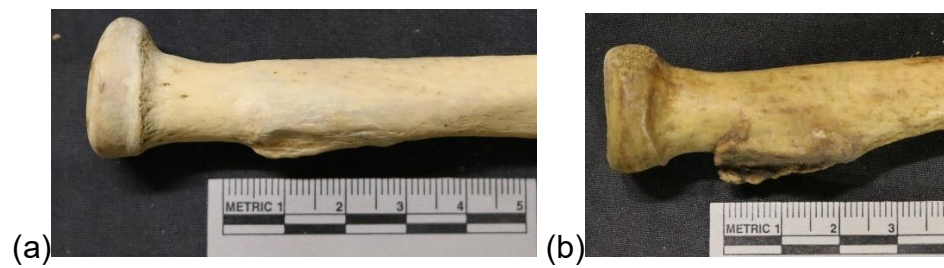


Fig. B.20 Radius tuberosity medial crest (a) 0, (b) 1

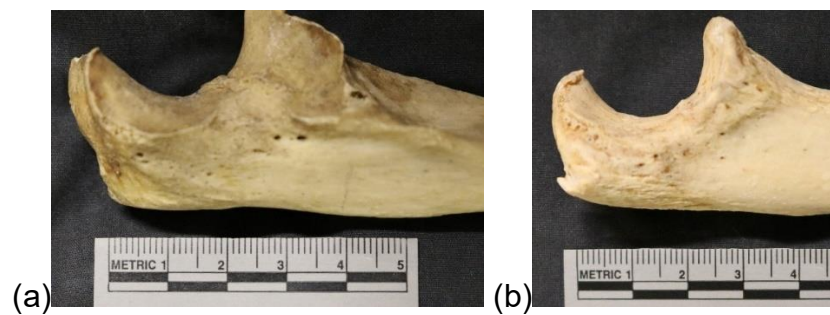


Fig. B.21 Ulna olecranon spur (a) 0, (b) 1



Fig. B.22 Trapezium lippling (a) 0, (b) 1

Table B.4 Innominate traits descriptions & scoring

Innominate* (Images in Figs B.23-B.31)		
Trait	Location	Scoring
Innominate weight relative	Entire innominate	0 (normal): Feels as heavy as expected when first picked up 1 (light): Feels unexpectedly light when first picked up
Innominate weight metric ^{2 3}	Entire innominate	Maximum height measured to nearest millimetre with osteometric board & mass measured on digital scale to nearest gram, maximum length and mass used to approximate bone density
Sacroiliac joint fusion	Sacroiliac joint	0 (normal): No fusion present 1 (fused): Any fusion evident at any point (take note of location) 2 (superior-anterior): Fusion along anterior sacroiliac ligament
Iliac crest fusion ¹	Superior iliac blade anterior to & including iliac tubercle	0 (not fused): Epiphysis not present 1 (partially fused): Only partial attachment of epiphysis to iliac blade, open & sharp edges 2 (remnant line): Completely fused but shallow groove ≥ 10 mm long at point of fusion 3 (fully fused): Point of fusion not evident
Iliac crest tuberculum ossification ¹	Along iliac crest from iliac tubercle to halfway to anterior superior iliac spine	0 (absent): No ossification or < 3 mm ossification for < 10 mm along crest 1 (ossification): Ossified fascia extending outwards ≥ 3 mm from crest for ≥ 10 mm 2 (double ossification): Two distinct rows of ossification
Ilium AHS exostoses	Surface anterior to inferior iliac spine	0 (absent): Smooth or $< 75\%$ covered in exostoses 1 (exostoses): $\geq 75\%$ surface covered in exostoses
Acetabulum posterior margin lipping ¹	30 mm forming posterior part of acetabulum outer margin	0 (absent): Smooth, round, or sharp edge 1 (present): Distinct & irregular lipping along ≥ 10 mm margin
Acetabulum articular surface extra bone ¹	Lateral-posterior articular surface of the acetabulum	0 (absent): No or < 5 mm extra bone 1 (present): ≥ 5 mm extra bone on surface
Acetabulum inferior joint lipping ¹	Inferior margin of acetabulum articular surface	0 (absent): No lipping or < 3 mm 1 (present): ≥ 3 mm lipping

Innominate* (Images in Figs B.23-B.31)		
Ischial tuberosity superior margin spur ¹	Superior margin of ischial tuberosity [attachment of semimembranosus]	0 (absent): No spur present 1 (present): Narrow, irregular spur ≥ 3 mm high for ≥ 10 mm
Ischial tuberosity medial spur ¹	Inferior 3 rd of medial border of ischial tuberosity [sacrotuberous ligament attachment]	0 (absent): No ossification or < 10 mm ossification 1 (present): Narrow ossification extends ≥ 10 mm medially
Ischial tuberosity bumps	Superior part of ischial tuberosity, superior to ridge between adductor magnus attachment and attachments for semimembranosus and semitendinosus & biceps femoris	0 (absent): Smooth surface 1 (present $< 50\%$): 1+ bumps but $< 50\%$ surface covered 2 (present $\geq 50\%$): $\geq 50\%$ surface bumpy

*included under lower limb in TA3 trait manual (Boldsen *et al.*, 2021a)

¹Mechanical sliding calipers

²Paleo-Tech field osteometric board

³Standard baking scale with sensitivity to 0.1g

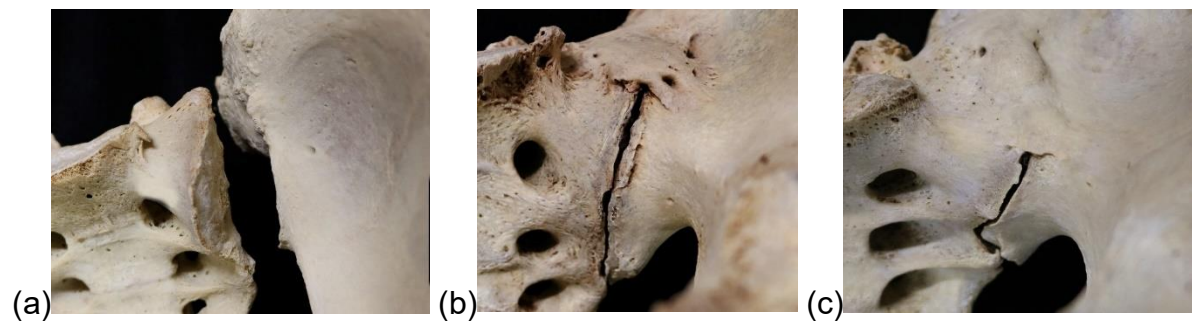


Fig. B.23 Sacro-iliac joint fusion (a) 0, (b) 1, (c) 2

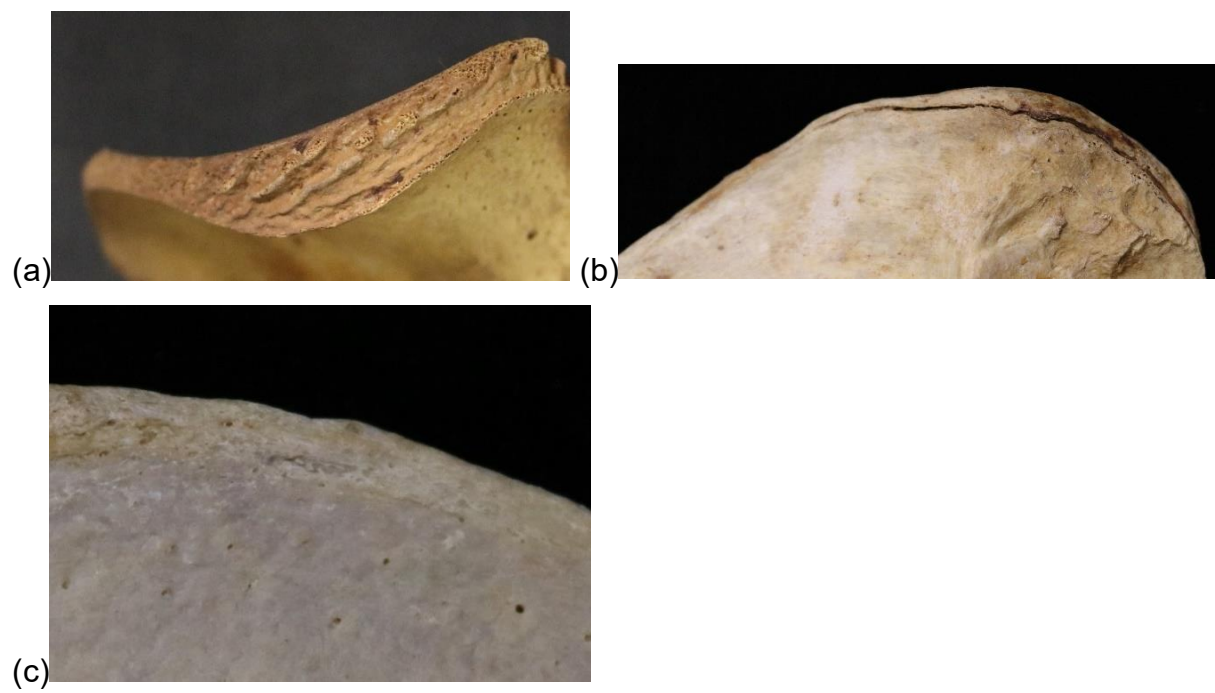


Fig. B.24 Iliac crest fusion (a) 0, (b) 1, (c) 2

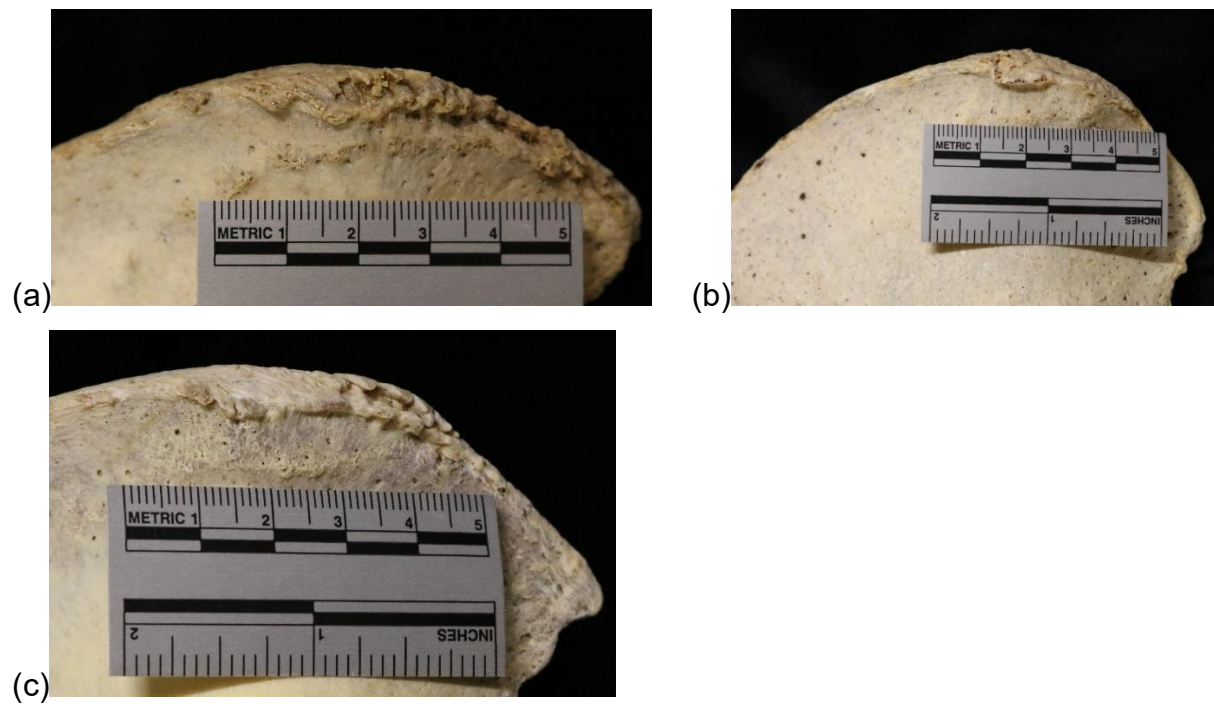


Fig. B.25 Iliac crest tuberculum ossification (a) 0 (insufficient size), (b) 1, (c) 2

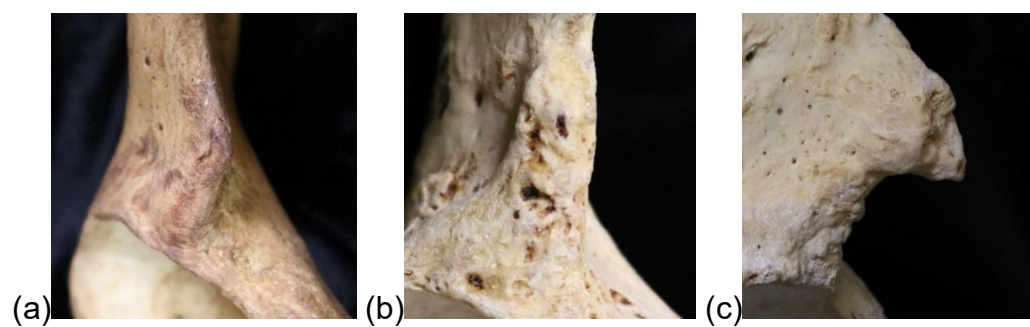


Fig. B.26 Ilium anterior inferior iliac crest exostoses (a) 0, (b) 1 anterior view, (c) 1 lateral view

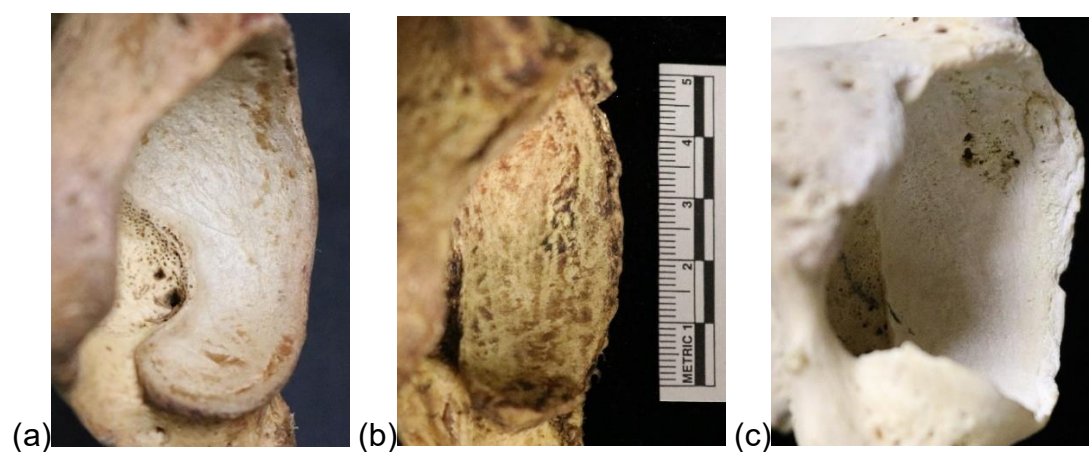


Fig. B.27 (a) Acetabulum posterior margin lipping & articular surface extra bone 0, (b) lipping 1, (c) extra bone 1

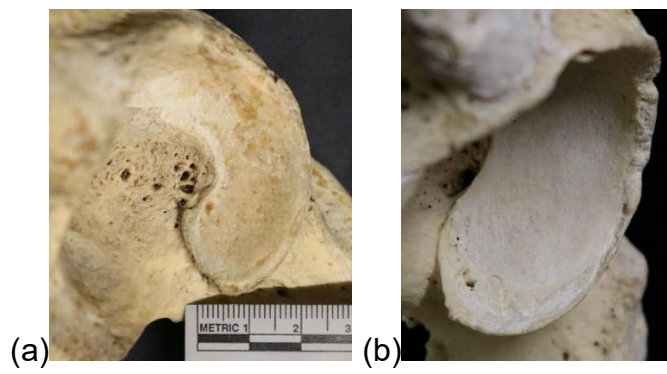


Fig. B.28 Acetabulum inferior joint lipping (a) 0, (b) 1

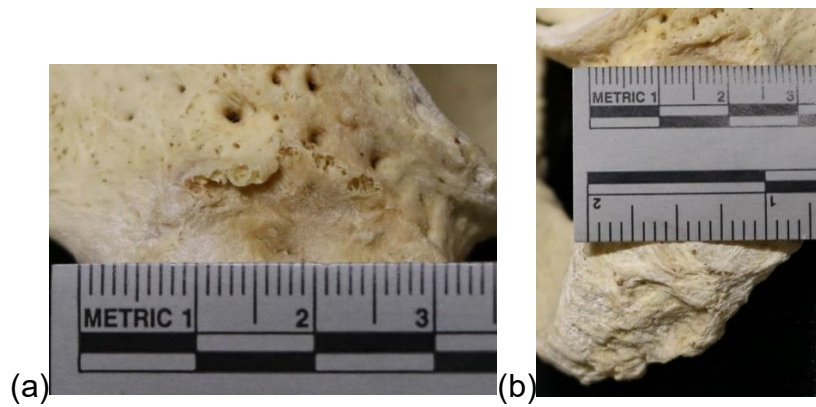


Fig. B.29 Ischial tuberosity superior margin spur (a) 0, (b) 1

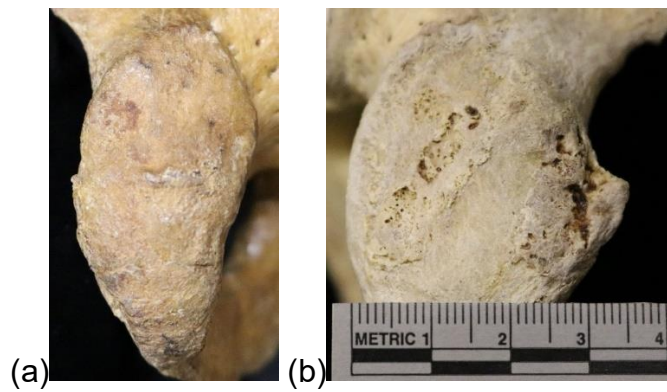


Fig. B.30 Ischial tuberosity medial spur (a) 0, (b) 1

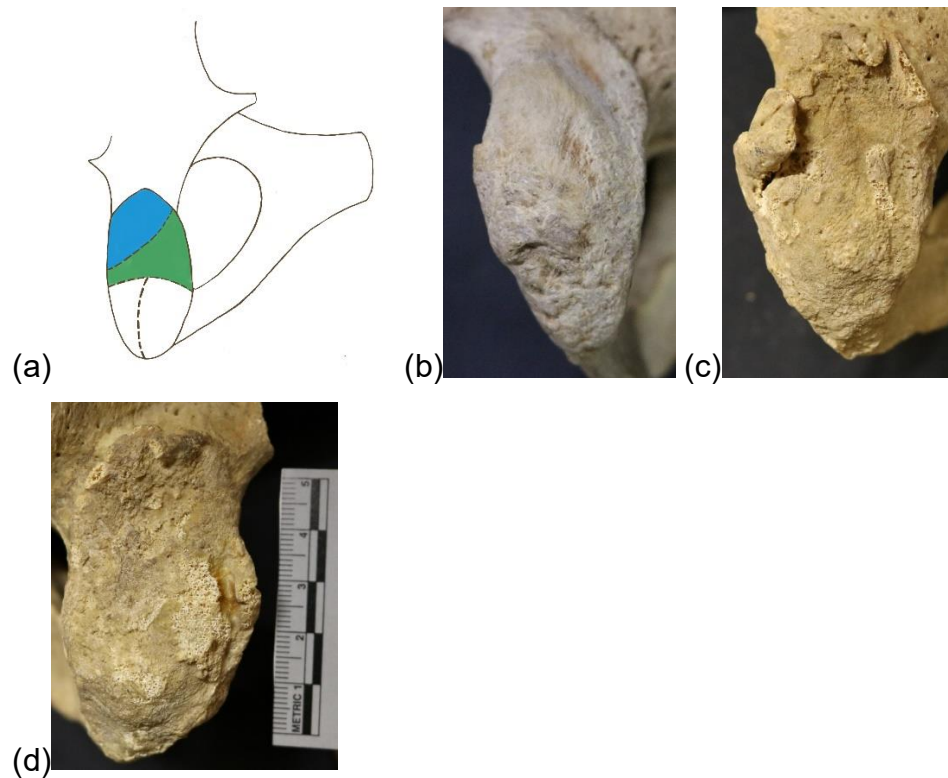


Fig. B.31 Ischial tuberosity bumps (a) The new scoring area highlighted in blue and green (Semimembranosus attachment in blue and semitendinosus & biceps femoris attachments in green), (b) 0, (c) 1, (d) 2

Table B.5 Auricular surface traits descriptions & scoring

Auricular surface of the Ilium (Images in Figs B.32 & B.33)		
Trait	Location	Scoring
Inferior surface porosity ¹	Lowermost 10mm of inferior demi face of auricular surface of the ilium	0 (smooth): Little to no porosity 1 (porosity): Micro- or macro-porosity over ≥50% area [DO NOT include any lipping]
Superior posterior iliac exostoses	Medial surface of posterior ilium [ventral sacroiliac ligament attachment]	0 (smooth): Smooth or maximum of 1 exostosis 1 (exostoses): 2+ exostoses present
Inferior posterior iliac exostoses	Medial surface of posterior ilium [interosseous sacroiliac ligament attachment]	0 (smooth): Smooth or maximum of 1 exostosis 1 (exostoses): 2+ exostoses present
Posterior exostoses	Between superior & inferior iliac regions of auricular surface of ilium	0 (smooth): Completely smooth surface 1 (discontinuous exostoses): 1+ exostoses present but not completely covering the area 2 (continuous exostoses): Area completely covered in exostoses

¹Mechanical sliding calipers

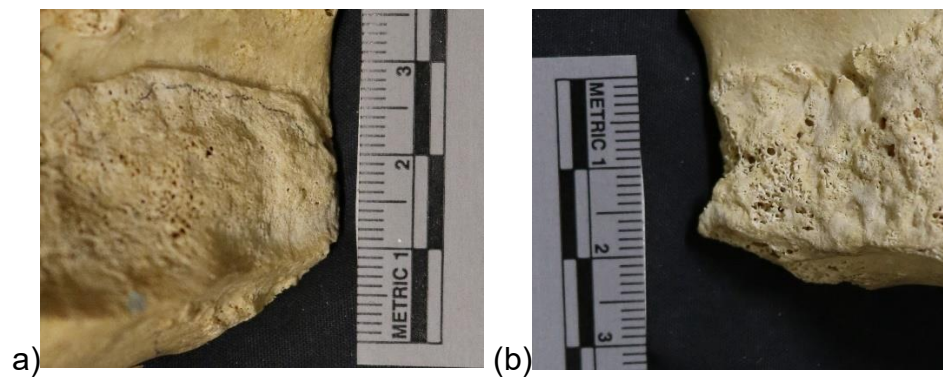


Fig. B.32 Auricular surface inferior surface porosity (a) 0, (b) 1

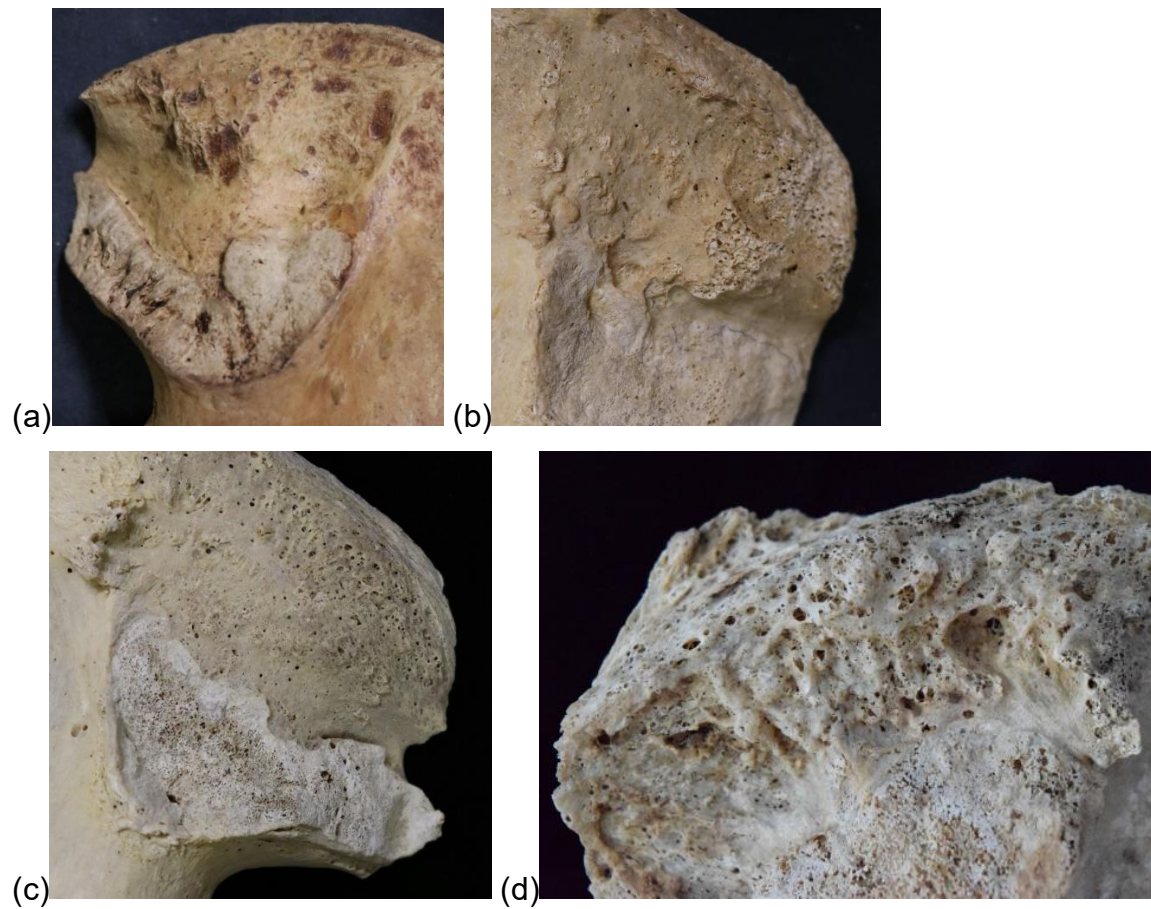


Fig. B.33 (a) Auricular surface superior posterior exostoses, inferior posterior exostoses, & posterior exostoses 0, (b) Auricular surface superior posterior exostoses 1, (c) Auricular surface inferior posterior exostoses & posterior exostoses 1, (d) Auricular surface posterior exostoses 2 (Boldsen *et al.*, 2021a)

Table B.6 Pubic symphysis traits descriptions & scoring

Pubic symphysis (Images in Figs B.34-B.36)		
Trait	Location	Scoring
Pubic symphyseal collar ¹	Anterior to ventral margin of pubic symphysis	0 (absent): No extra bone or <10 mm 1 (collar): ≥10 mm extra bone parallel to anterior border
Symphyseal relief	Pubic symphyseal face	0 (billows): Most of area covered in clearly defined billows 1 (residual): Looks mostly flat but 2+ consecutive billows present 2 (flat/irregular): Completely flat, maximum of 1 billow
Superior protuberance	Superior region of pubic symphyseal face	0 (serrated): No knob, just billows 1 (knob): Distinctive smooth knob projecting 2 (flat): Little/no elevation + no billows
Ventral symphyseal margin ¹	Ventral margin of pubic symphysis	0 (serrated/bevelled): Typical billows, no rampart forming 1 (rampart): Rampart forming/formed but no rim 2 (rim): Narrow & sharp irregular rim ≥10 mm long, borders depressed symphyseal face 3 (breakdown): Pitting & erosion of rim for ≥10 mm
Dorsal symphyseal margin ¹	Dorsal margin of pubic symphysis	0 (serrated): Typical billows 1 (flattened): ≥10 mm meeting point between face & margin but no rim 2 (rim): Elevated rim ≥10 mm around the flattened face 3 (breakdown): Pitting & erosion of rim for ≥10 mm

¹Mechanical sliding calipers

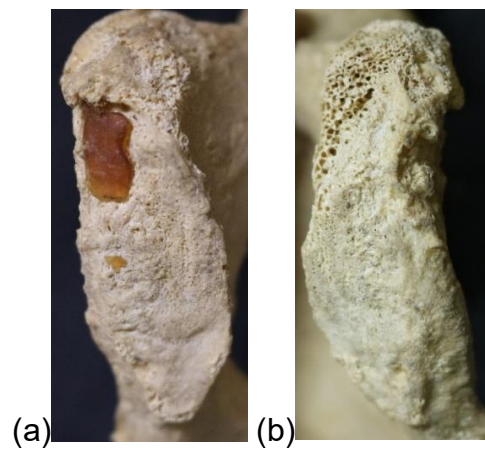


Fig. B.34 Pubic symphyseal collar (a) 0, (b) 1



Fig. B.35 (a) Pubic symphyseal relief 0, Pubic superior protuberance 0, (b) Pubic symphyseal relief 1, (c) Pubic symphyseal relief 2, (d) Pubic superior protuberance 1, (e) Pubic superior protuberance 2

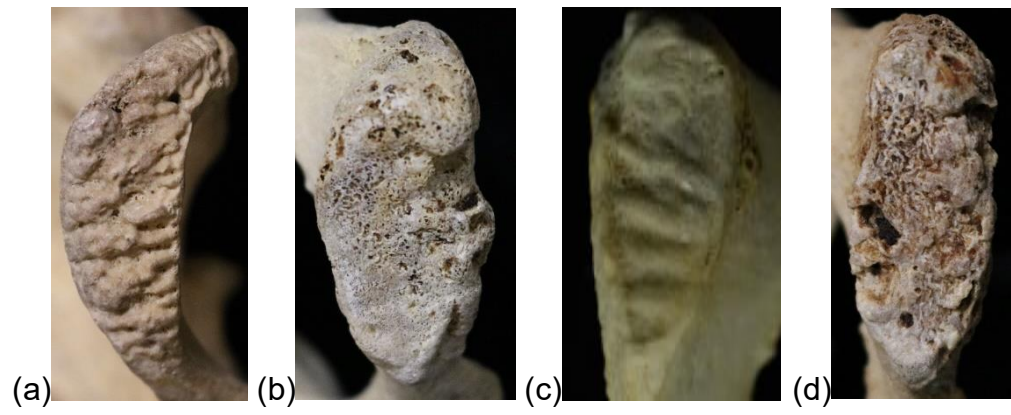


Fig. B.36 Pubic symphysis ventral margin & dorsal margin (a) 0, (b) 1, (c) 2 (Boldsen *et al.*, 2021a), (d)
3

Table B.7 Vertebral traits descriptions and scoring

Vertebrae (Images in Figs B.37-B.46)		
Trait	Location	Scoring
C1 lipping ¹	Edges of articular surface for dens on C1 on posterior surface of anterior arch	0 (absent): No lipping present or around <50% of margin 1 (present): Irregular lipping extending ≥ 1.5 mm around ≥ 50 of margin
C1 eburnation	Articular surface for dens on C1 on posterior surface of anterior arch	0 (absent): No eburnation present 1 (present): 1+ spot of eburnation
L1 spinous process flattening ¹	Dorsal end of spinous process of L1	0 (smooth convex/round): No ligament ossification, normal contour 1 (smooth flat): Flattened in lateral view, no ligament ossification 2 (ligament ossification): ≥ 1 mm ligament ossification present
L1 epiphyseal ring fusion	Superior & inferior surfaces of body of L1	0 (not fused): No fusion 1 (partially fused): Some fusion but not entire epiphyseal ring 2 (remnant line): Epiphyseal ring completely fused but shallow groove present at fusion point 3 (fully fused): Point of fusion not evident
L5 epiphyseal ring fusion	Superior & inferior surfaces of body of L5	0 (not fused): No fusion 1 (partially fused): Some fusion but not entire epiphyseal ring 2 (remnant line): Epiphyseal ring completely fused but shallow groove present at fusion point 3 (fully fused): Point of fusion not evident
L1 surface morphology ¹	The anterior portion of superior & inferior surfaces of body of L1, within 1 cm of vertebral border, includes an epiphyseal ring	0 (billows): $\geq 50\%$ scoring area covered with distinct billows, must be well defined for ≥ 10 mm anteriorly 1 (flat): The body looks flat, residual billows may be present 2 (porous): $\geq 1/3$ scoring area covered with macropores or concentrations of micropores
L5 surface morphology ¹	Anterior portion of superior & inferior surfaces of L5, within 1 cm of vertebral border, includes an epiphyseal ring	0 (billows): $\geq 50\%$ scoring area covered with distinct billows, must be well defined for ≥ 10 mm anteriorly 1 (flat): The body looks flat, residual billows may be present

Vertebrae (Images in Figs B.37-B.46)		
Trait	Location	Scoring
		2 (porous): $\geq 1/3$ scoring area covered with macropores or concentrations of micropores
L5 margin shape ¹	Anterior 20 mm (across midline) of superior & inferior margins of body of L5	0 (round): Smooth & rounded edge, also if epiphyseal ring is unfused 1 (sharp): Edge feels sharp 2 (lipped): Any lipping present
S1 margin shape ¹	Anterior 20 mm (across midline) of superior margin of body of S1	0 (round): Smooth & rounded edge, also if epiphyseal ring is unfused 1 (sharp): Edge feels sharp 2 (lipped): Any lipping present
S1-S2 fusion ¹	The anterior surface of sacrum where first two sacral bodies meet	0 (gap ≥ 10 mm): Uninterrupted gap ≥ 10 mm present 1 (gap < 10 mm): Uninterrupted gap < 10 mm present 2 (closed): No gap evident
Sacral “elbow”	Angle between S2 and S4 on anterior surface of the sacrum	0 (absent): Normal smooth arc 1 (present): Sharp interruption of arc (90°-110° angle)
Vertebral lipping ¹	Anterior segments of superior & inferior edges of bodies of C3-C7, T1-T12, L1-L5	<u>Cervical</u> 0 (absent): < 3 mm lipping extending from margin of the body 1 (present): ≥ 3 mm lipping extending from margin, includes fusion between vertebrae <u>Thoracic & Lumbar</u> 0 (absent): < 5 mm lipping extending from margin 1 (present): ≥ 5 mm lipping extending from margin, includes fusion
Vertebral candlewax	Superior & inferior edges & anterior surfaces of bodies of C3-C7, T1-T12, L1-L6	0 (absent): No candlewax present 1 (present): Smooth & sclerotic bone extends across intervertebral joints on any vertebral edges [record #scorable edges] [score C,T, L separately]
Vertebral DISH	Superior & inferior edges & anterior surfaces of bodies of C3-L5	0 (absent): No candlewax or candlewax doesn't span 3+ joints 1 (present): Candlewax spans across 4+ adjacent vertebrae [take note whether C, T, L]

¹Mechanical sliding calipers

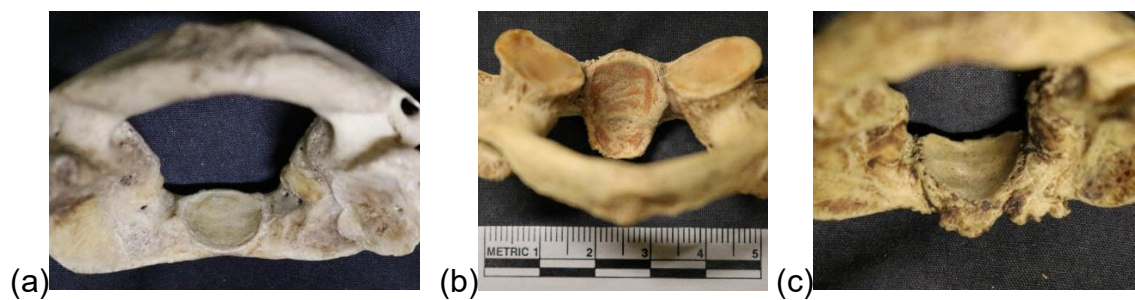


Fig. B.37 (a) C1 lipping & eburnation 0, (b) C1 lipping 1, (c) C1 eburnation 1



Fig. B.38 L1 spinous process flattening (a) 0, (b) 1, (c) 2

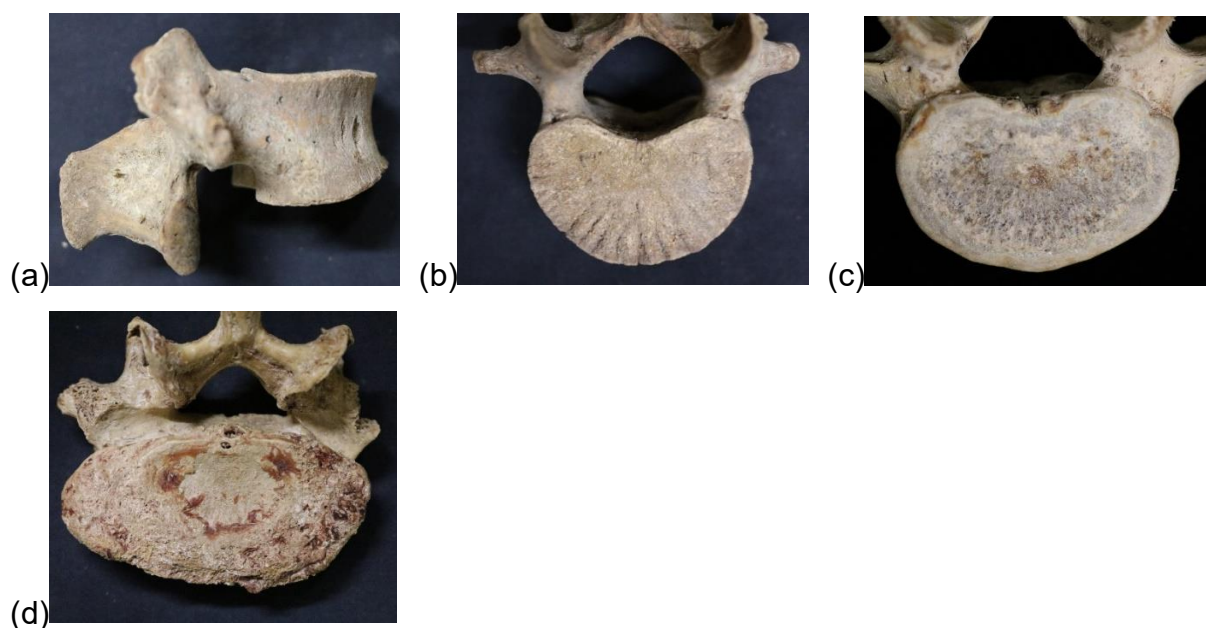


Fig. B.39 L1-5 epiphyseal ring fusion (a) 0 lateral view, (b) 0 superior view, (c) 2, (d) 3

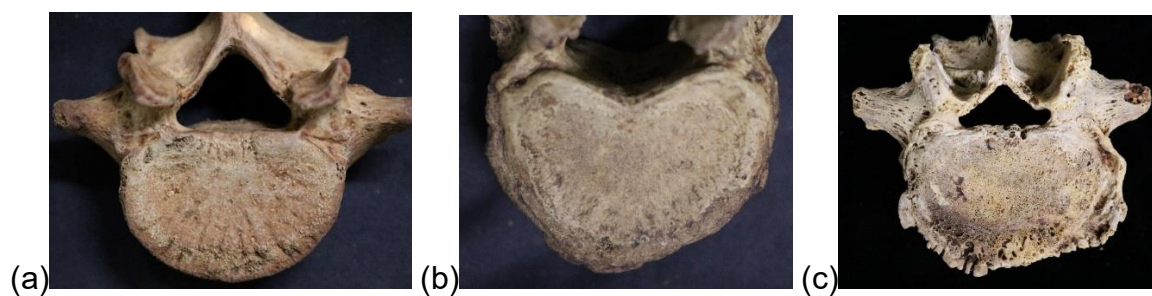


Fig. B.40 L1-5 surface morphology (a) 0, (b) 1, (c) 2



Fig. B.41 L5 margin shape (a) 0, (b) 1, (c) 2

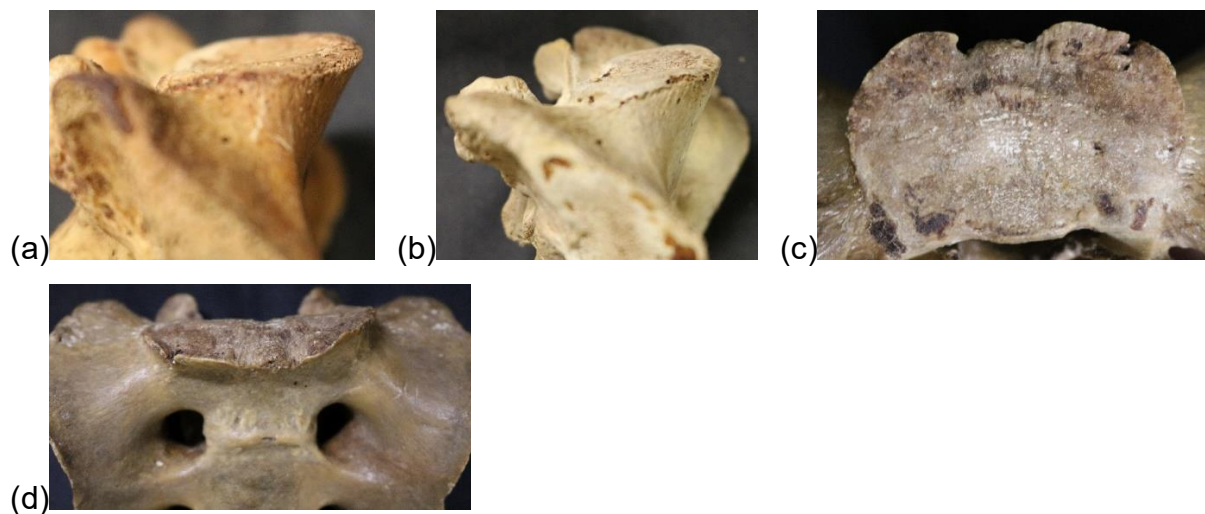


Fig. B.42 S1 margin shape (a) 0, (b) 1, (c) 2 superior view, (d) 2 anterior view



Fig. B.43 S1-2 fusion (a) 0, (b) 1, (c) 2

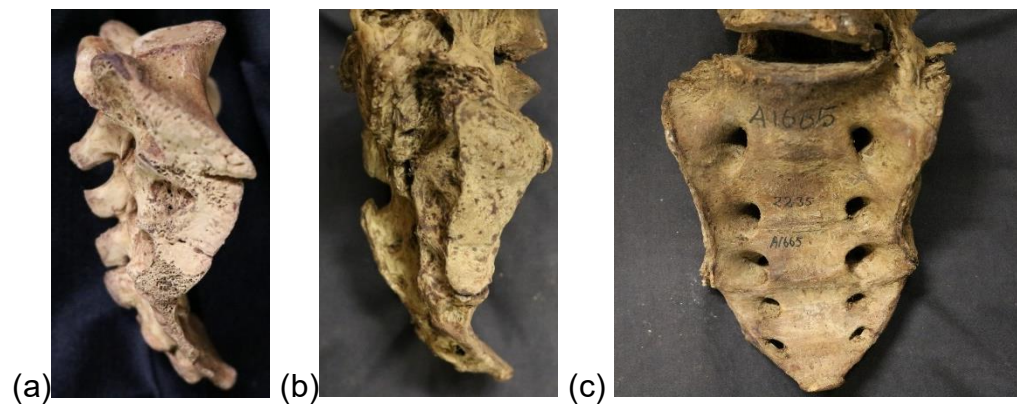


Fig. B.44 Sacral elbow (a) 0, (b) 1, (c) anterior view of the sacral elbow and healed fracture

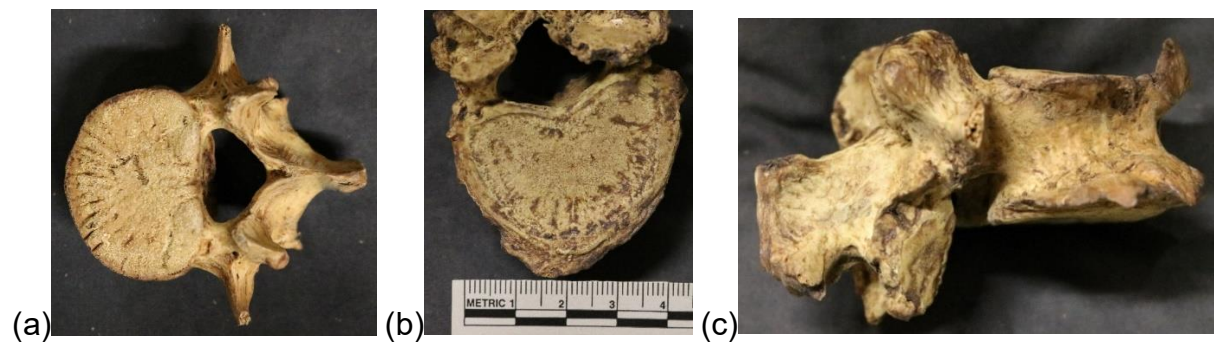


Fig. B.45 (a) Vertebral lipping & candlewax 0, (b) Vertebral lipping 1, (c) Vertebral candlewax 1



Fig. B.46 DISH 1

Table B.8 Ribs & sternum traits descriptions & scoring

Ribs & sternum (Images in Figs B.47-B.51)		
Trait	Location	Scoring
R1 fusion	Articulation of manubrium and 1st rib	0 (absent): No fusion present 1 (fused): 1+ bony bridges between 1 st rib & manubrium
Sternum dorsal ridge ¹	Dorsal surface of completely fused sternal body	0 (absent): No ridge present 1 (present): Single ridge in midline at least 15 mm long and 1.5 mm wide at 1+ points in superior-inferior direction
R2 -10 sternal end rim profiles	Sternal ends of R2-R10 [specifically the border]	0 (regular): Edge not lipped 1 (irregular): Irregular bony lipping present, cartilage ossification present
Rib ossification pattern ¹	Sternal ends of R2-R10 [specifically the border]	0 (absent): Neither claw nor cone present 1 (cone): Bony growth extending ≥ 5 mm from centre of sternal end 2 (claw): Irregular bony growth extending ≥ 5 mm from periphery of sternal end 3 (both): Both claw and cone present
R2 -10 shingle ribs	Anterior 3 rd of R2-10	0 (normal): Thick robust bone, oval cross-section with sharp inferior edge 1 (shingle): Narrow cross-section with sharp superior & inferior edges, potential flattening in distal 3 rd

¹Mechanical sliding calipers

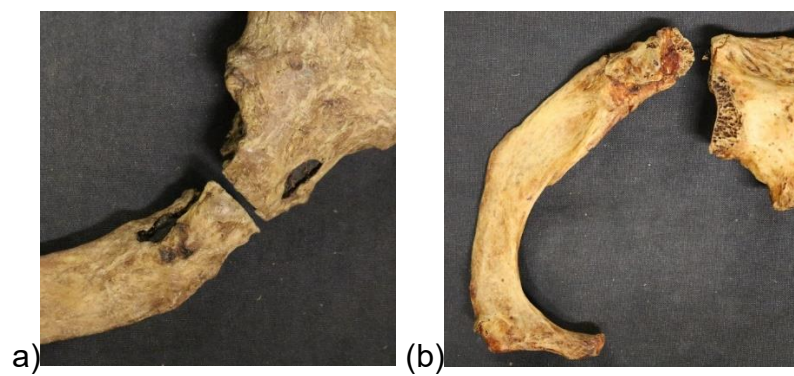


Fig. B.47 Rib 1 fusion (a) 0, (b) 1

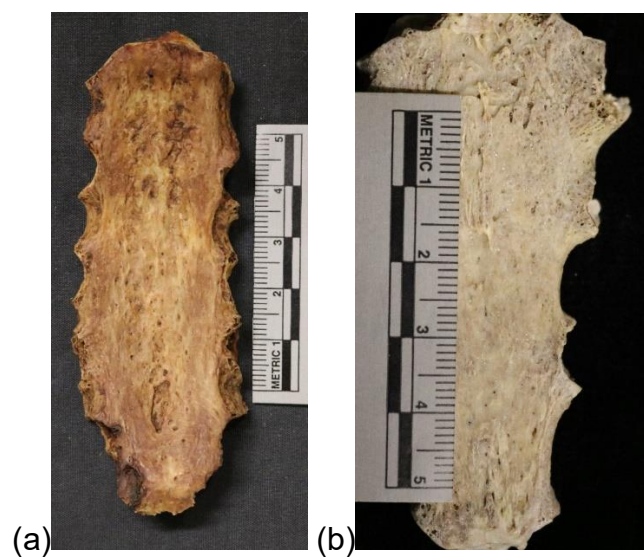


Fig. B.48 Sternum dorsal ridge (a) 0, (b) 1

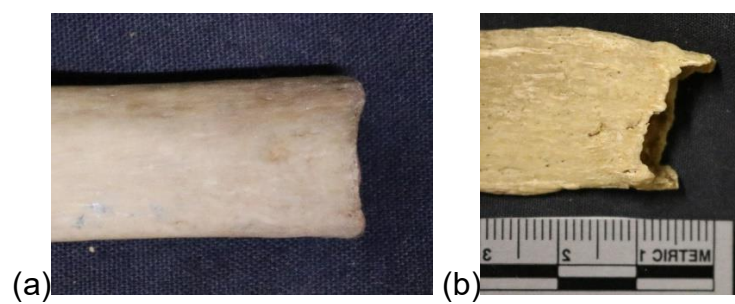


Fig. B.49 Rib sternal end rim profile (a) 0, (b) 1

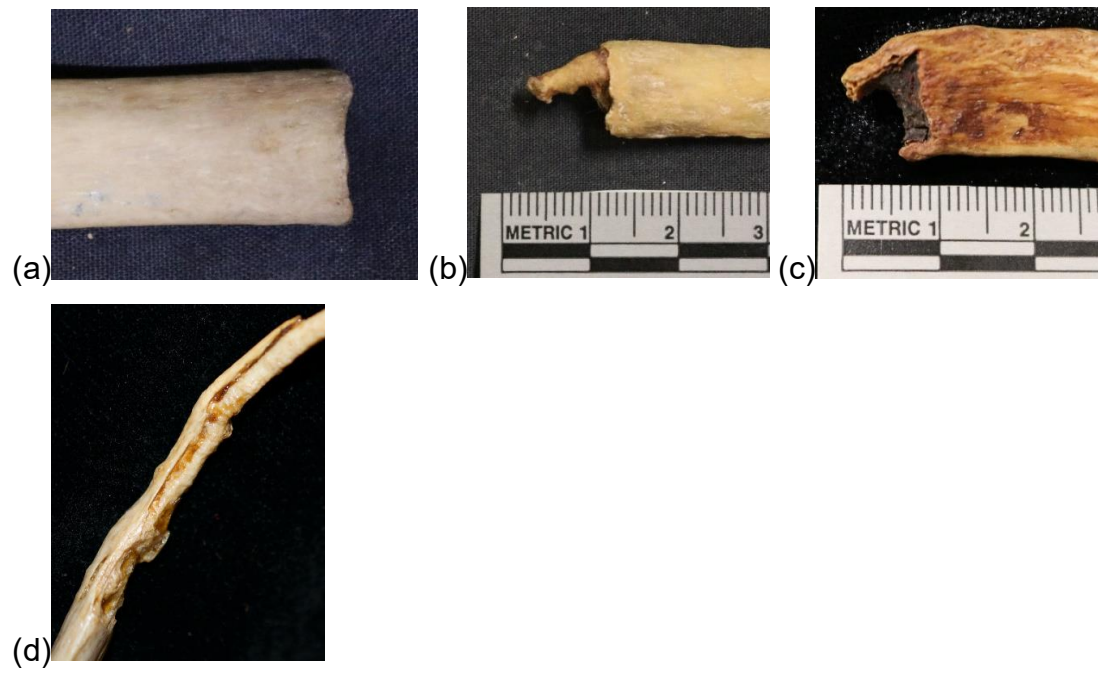


Fig. B.50 Rib ossification pattern (a) 0, (b) 1, (c) 2, (d) 3

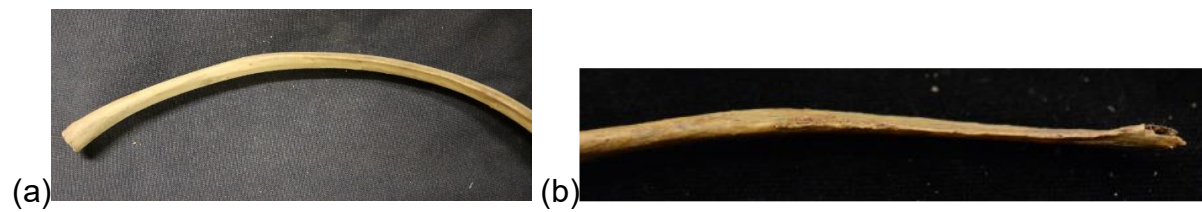


Fig. B.51 Shingle ribs (a) 0, (b) 1 (Boldsen *et al.*, 2021a)

Appendix C Expanded results tables & diagrams

C.1 Repeatability

Table C.1 Cohen's Weighted Kappa for intra-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$)

Trait	Cohen's Kappa	Strength of Agreement
Femur fovea margin lipping	0.474	Moderate
Femur head surface extra bone	0.434	Moderate
Femur greater trochanter roughening	0.464	Moderate
Femur trochanteric fossa exostoses	0.878	Almost perfect
Femur trochanteric medial exostoses	0.571	Moderate
Tibia relative weight	0.769	Substantial
Fibula wings	1.000	Almost perfect
Calcaneus relative weight	0.737	Substantial
Clavicle medial epiphysis fusion	0.825	Almost perfect
Clavicle medial epiphysis gravel	1.000	Almost perfect
Clavicle medial macroporosity	0.857	Almost perfect
Clavicle lateral macroporosity	1.000	Almost perfect
Scapula glenoid fossa lipping	0.919	Almost perfect
Humerus relative weight	0.412	Moderate
Humerus lesser tubercle anterior surface bumps	1.000	Almost perfect
Humerus lesser tubercle margin shape	0.605	Moderate
Humerus greater tubercle pits	0.615	Substantial
Humerus medial epicondyle	1.000	Almost perfect
Humerus lateral epicondyle	0.375	Fair
Radius tuberosity medial crest	1.000	Almost perfect
Ulna olecranon spur	0.821	Almost perfect
Trapezium lipping	0.583	Moderate
Innominate relative weight	0.615	Substantial
Sacroiliac joint fusion	1.000	Almost perfect
Iliac crest fusion	0.844	Almost perfect
Iliac crest tuberculum ossification	1.000	Almost perfect
Ilium anterior inferior iliac spine exostoses	<0.001	Slight
Acetabulum posterior margin lipping	0.400	Fair
Acetabulum posterior articular surface extra bone	0.412	Moderate
Acetabulum inferior joint lipping	0.643	Substantial
Ischial tuberosity superior margin spur	1.000	Almost perfect
Ischial tuberosity medial spur	NA	NA
Ischial tuberosity bumps	0.857	Almost perfect
Auricular surface inferior surface porosity	0.643	Substantial
Auricular surface superior posterior iliac exostoses	0.615	Substantial
Auricular surface inferior posterior iliac exostoses	0.474	Moderate
Auricular surface posterior exostoses	0.778	Substantial
Pubic symphyseal collar	<0.001	Slight

Table C.1 continued

Trait	Cohen's Kappa	Strength of Agreement
Pubic symphyseal relief	0.690	Substantial
Pubic superior protuberance/apex	0.407	Moderate
Pubic symphysis ventral margin	0.583	Moderate
Pubic symphysis dorsal margin	0.597	Moderate
C1 lipping	1.000	Almost perfect
C1 eburnation	1.000	Almost perfect
L1 spinous process flattening	0.865	Almost perfect
L1 superior epiphyseal ring fusion	1.000	Almost perfect
L1 inferior epiphyseal ring fusion	1.000	Almost perfect
L1 superior surface morphology	0.615	Substantial
L1 inferior surface morphology	0.318	Fair
L5 superior epiphyseal fusion	1.000	Almost perfect
L5 inferior epiphyseal fusion	1.000	Almost perfect
L5 superior surface morphology	0.643	Substantial
L5 inferior surface morphology	1.000	Almost perfect
L5 superior margin shape	1.000	Almost perfect
L5 inferior margin shape	0.865	Almost perfect
S1 margin shape	0.857	Almost perfect
S1-2 fusion	0.875	Almost perfect
Sacral elbow	1.000	Almost perfect
C lipping	0.444	Moderate
C candlewax	0.600	Moderate
T lipping	0.583	Moderate
T candlewax	0.800	Substantial
L lipping	1.000	Almost perfect
L candlewax	1.000	Almost perfect
DISH	1.000	Almost perfect
R1 fusion	1.000	Almost perfect
Sternum central dorsal ridge	0.270	Fair
R2 rim profile	0.800	Substantial
R2 ossification pattern	0.800	Substantial
R3-10 rim profile	0.737	Substantial
R3-10 ossification pattern	0.661	Substantial
R2-10 shingle ribs	0.483	Moderate
Parietal depression	-0.190	Poor
Spheno-occipital synchondrosis	1.000	Almost perfect
Occipital condyle lipping	<0.001	Slight

Table C.2 Lin's Concordance Correlation Coefficient for intra-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$) *density calculated by dividing actual weight by length as recommended in the trait scoring guide

Trait	Lin's CCC	Strength of Agreement
Tibia relative density*	0.9994	Almost perfect
Tibia actual weight	0.9999	Almost perfect
Tibia length	0.9992	Almost perfect
Calcaneus relative density*	0.9966	Almost perfect
Calcaneus actual weight	0.9956	Almost perfect
Calcaneus length	0.9826	Substantial
Humerus relative density*	0.9997	Almost perfect
Humerus actual weight	0.9999	Almost perfect
Humerus length	0.9997	Almost perfect
Innominate relative density*	0.9997	Almost perfect
Innominate actual weight	0.9998	Almost perfect
Innominate length	0.9982	Almost perfect

Table C.3 Cohen's Weighted Kappa for inter-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$)

Trait	Cohen's Kappa	Strength of Agreement
Femur fovea margin lipping	0.634	Moderate
Femur head surface extra bone	0.828	Almost perfect
Femur greater trochanter roughening	0.655	Substantial
Femur trochanteric fossa exostoses	0.762	Substantial
Femur trochanteric medial exostoses	0.750	Substantial
Tibia relative weight	1.000	Almost perfect
Fibula wings	0.688	Substantial
Calcaneus relative weight	1.000	Almost perfect
Clavicle medial epiphysis fusion	0.271	Fair
Clavicle medial epiphysis gravel	0.459	Moderate
Clavicle medial macroporosity	0.286	Fair
Clavicle lateral macroporosity	0.886	Almost perfect
Scapula glenoid fossa lipping	0.259	Fair
Humerus relative weight	-0.071	Poor
Humerus lesser tubercle anterior surface bumps	0.583	Moderate
Humerus lesser tubercle margin shape	0.362	Fair
Humerus greater tubercle pits	0.286	Fair
Humerus medial epicondyle	0.615	Substantial
Humerus lateral epicondyle	0.375	Fair
Radius tuberosity medial crest	0.444	Moderate
Ulna olecranon spur	0.688	Substantial
Trapezium lipping	0.583	Moderate

Table C.3 continued

Trait	Cohen's Kappa	Strength of Agreement
Innominate relative weight	<0.001	Slight
Sacroiliac joint fusion	<0.001	Slight
Iliac crest fusion	0.333	Fair
Iliac crest tuberculum ossification	0.773	Substantial
Ilium anterior inferior iliac spine exostoses	<0.001	Slight
Acetabulum posterior margin lipping	-0.200	Poor
Acetabulum posterior articular surface extra bone	-0.111	Poor
Acetabulum inferior joint lipping	0.444	Moderate
Ischial tuberosity superior margin spur	1.000	Almost perfect
Ischial tuberosity medial spur	NA	NA
Ischial tuberosity bumps	0.333	Fair
Auricular surface inferior surface porosity	<0.001	Slight
Auricular surface superior posterior iliac exostoses	0.091	Slight
Auricular surface inferior posterior iliac exostoses	0.026	Slight
Auricular surface posterior exostoses	0.375	Fair
Pubic symphyseal collar	0.615	Fair
Pubic symphyseal relief	0.297	Fair
Pubic superior protuberance/apex	0.180	Slight
Pubic symphysis ventral margin	0.321	Fair
Pubic symphysis dorsal margin	0.177	Slight
C1 lipping	0.688	Substantial
C1 eburnation	0.615	Substantial
L1 spinous process flattening	0.853	Almost perfect
L1 superior epiphyseal ring fusion	<0.001	Slight
L1 inferior epiphyseal ring fusion	1.000	Almost perfect
L1 superior surface morphology	0.444	Moderate
L1 inferior surface morphology	-0.190	Poor
L5 superior epiphyseal fusion	<0.001	Slight
L5 inferior epiphyseal fusion	<0.001	Slight
L5 superior surface morphology	0.444	Moderate
L5 inferior surface morphology	0.773	Substantial
L5 superior margin shape	1.000	Almost perfect
L5 inferior margin shape	0.419	Moderate
S1 margin shape	0.667	Substantial
S1-2 fusion	0.583	Moderate
Sacral elbow	0.194	Slight
C lipping	0.310	Fair
C candlewax	<0.001	Slight
T lipping	0.583	Moderate
T candlewax	<0.001	Slight
L lipping	1.000	Almost perfect
L candlewax	<0.001	Slight
DISH	<0.001	Slight
R1 fusion	<0.001	Slight
Sternum central dorsal ridge	0.800	Substantial

Table C.3 continued

Trait	Cohen's Kappa	Strength of Agreement
R2 rim profile	0.286	Fair
R2 ossification pattern	0.444	Moderate
R3-10 rim profile	0.103	Slight
R3-10 ossification pattern	0.149	Slight
R3-10 shingle ribs	-0.111	Poor
Parietal depression	<0.001	Slight
Spheno-occipital synchondrosis	0.642	Substantial
Occipital condyle lipping	0.500	Moderate

Table C.4 Lin's Concordance Correlation Coefficient for inter-observer repeatability (left side only for bilateral traits) ($\alpha = 0.05$, $n = 10$) *density calculated by dividing actual weight by length as recommended in the trait scoring guide

Trait	Lin's CCC	Strength of Agreement
Tibia relative density*	0.9936	Almost perfect
Tibia actual weight	0.9963	Almost perfect
Tibia length	0.9911	Almost perfect
Calcaneus relative density*	0.4050	Poor
Calcaneus actual weight	0.5014	Poor
Calcaneus length	0.9817	Substantial
Humerus relative density*	0.9999	Almost perfect
Humerus actual weight	1.0000	Almost perfect
Humerus length	0.9994	Almost perfect
Innominate relative density*	0.9992	Almost perfect
Innominate actual weight	0.9999	Almost perfect
Innominate length	0.9968	Almost perfect

C.2 Correlation of trait to true age

Table C.5 The Spearman's Rank Correlation Coefficient for the correlation between trait and true age in the full (n = 89), male (n = 45), and female (n = 44) samples ($\alpha = 0.05$). (¹According to Chan (2003))

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
Femur fovea margin lipping (L)	0.455	<0.001	Fair	0.696	<0.001	Moderate	0.192	0.224	Poor
Femur fovea margin lipping (R)	0.474	<0.001	Fair	0.608	<0.001	Moderate	0.343	0.026	Fair
Femur head surface extra bone (L)	0.390	<0.001	Fair	0.562	<0.001	Fair	0.236	0.126	Poor
Femur head surface extra bone (R)	0.443	<0.001	Fair	0.634	<0.001	Moderate	0.195	0.206	Poor
Femur greater trochanter roughening (L)	0.421	<0.001	Fair	0.204	0.202	Poor	0.611	<0.001	Moderate
Femur greater trochanter roughening (R)	0.372	0.001	Fair	0.183	0.241	Poor	0.543	0.001	Fair
Femur trochanteric fossa exostoses (L)	0.405	<0.001	Fair	0.382	0.016	Fair	0.433	0.005	Fair
Femur trochanteric fossa exostoses (R)	0.271	0.014	Poor	0.154	0.329	Poor	0.382	0.014	Fair
Femur trochanteric medial exostoses (L)	0.372	0.001	Fair	0.114	0.473	Poor	0.594	<0.001	Fair
Femur trochanteric medial exostoses (R)	0.362	0.001	Fair	0.195	0.211	Poor	0.519	0.001	Fair
Tibia relative weight (L)	0.328	0.004	Fair	0.206	0.193	Poor	0.467	0.004	Fair
Tibia relative weight (R)	0.260	0.024	Poor	0.259	0.110	Poor	0.355	0.033	Fair
Tibia actual weight (L)	-0.108	0.343	Poor	-0.011	0.944	Poor	-0.484	0.004	Fair
Tibia actual weight (R)	-0.075	0.524	Poor	0.053	0.743	Poor	-0.460	0.007	Fair
Tibia length (L)	0.068	0.543	Poor	0.265	0.094	Poor	-0.225	0.166	Poor
Tibia length (R)	0.045	0.696	Poor	0.276	0.089	Poor	-0.274	0.106	Poor
Fibula wings (L)	0.308	0.006	Fair	0.406	0.010	Fair	0.184	0.257	Poor
Fibula wings (R)	0.193	0.090	Poor	0.321	0.048	Fair	0.086	0.594	Poor
Calcaneus relative weight (L)	0.205	0.064	Poor	0.167	0.291	Poor	0.248	0.117	Poor
Calcaneus relative weight (R)	0.169	0.129	Poor	0.146	0.343	Poor	0.214	0.186	Poor
Calcaneus actual weight (L)	-0.179	0.112	Poor	-0.069	0.664	Poor	-0.378	0.020	Fair
Calcaneus actual weight (R)	-0.177	0.116	Poor	-0.095	0.540	Poor	-0.342	0.040	Fair
Calcaneus length (L)	0.198	0.075	Poor	0.367	0.020	Fair	0.145	0.358	Poor

Table C.5 continued

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
Calcaneus length (R)	0.280	0.013	Poor	0.396	0.012	Fair	0.242	0.137	Poor
Clavicle medial epiphysis fusion (L)	0.644	<0.001	Moderate	0.625	<0.001	Moderate	0.669	<0.001	Moderate
Clavicle medial epiphysis fusion (R)	0.640	<0.001	Moderate	0.666	<0.001	Moderate	0.604	<0.001	Moderate
Clavicle medial epiphysis gravel (L)	0.240	0.037	Poor	0.308	0.057	Fair	0.217	0.187	Poor
Clavicle medial epiphysis gravel (R)	0.324	0.004	Fair	0.511	0.001	Fair	0.069	0.668	Poor
Clavicle medial macroporosity (L)	0.027	0.820	Poor	0.003	0.984	Poor	NA	NA	NA
Clavicle medial macroporosity (R)	0.220	0.053	Poor	0.319	0.049	Fair	0.020	0.901	Poor
Clavicle lateral macroporosity (L)	0.330	0.004	Fair	0.387	0.016	Fair	0.251	0.126	Poor
Clavicle lateral macroporosity (R)	0.385	0.001	Fair	0.379	0.017	Fair	0.415	0.010	Fair
Scapula glenoid fossa lipping (L)	0.611	<0.001	Moderate	0.655	<0.001	Moderate	0.565	<0.001	Fair
Scapula glenoid fossa lipping (R)	0.676	<0.001	Moderate	0.662	<0.001	Moderate	0.696	<0.001	Moderate
Humerus relative weight (L)	0.190	0.084	Poor	-0.004	0.977	Poor	0.378	0.017	Fair
Humerus relative weight (R)	0.284	0.009	Poor	0.114	0.456	Poor	0.478	0.002	Fair
Humerus actual weight (L)	-0.049	0.662	Poor	0.071	0.644	Poor	-0.576	<0.001	Fair
Humerus actual weight (R)	-0.161	0.142	Poor	0.018	0.908	Poor	-0.698	<0.001	Moderate
Humerus length (L)	0.088	0.425	Poor	0.307	0.046	Fair	-0.209	0.186	Poor
Humerus length (R)	0.052	0.634	Poor	0.298	0.050	Poor	-0.210	0.178	Poor
Humerus lesser tubercle anterior surface bumps (L)	0.329	0.002	Fair	0.232	0.128	Poor	0.418	0.007	Fair
Humerus lesser tubercle anterior surface bumps (R)	0.399	<0.001	Fair	0.299	0.047	Poor	0.485	0.002	Fair
Humerus lesser tubercle margin shape (L)	0.351	0.001	Fair	0.277	0.069	Poor	0.432	0.006	Fair
Humerus lesser tubercle margin shape (R)	0.367	0.001	Fair	0.273	0.070	Poor	0.471	0.003	Fair
Humerus greater tubercle pits (L)	0.429	<0.001	Fair	0.476	0.002	Fair	0.376	0.015	Fair
Humerus greater tubercle pits (R)	0.439	<0.001	Fair	0.453	0.003	Fair	0.427	0.007	Fair
Humerus medial epicondyle (L)	0.381	<0.001	Fair	0.298	0.054	Poor	0.485	0.002	Fair
Humerus medial epicondyle (R)	0.373	<0.001	Fair	0.321	0.033	Fair	0.444	0.004	Fair
Humerus lateral epicondyle (L)	0.397	<0.001	Fair	0.366	0.018	Fair	0.441	0.004	Fair
Humerus lateral epicondyle (R)	0.442	<0.001	Fair	0.437	0.004	Fair	0.456	0.003	Fair
Radius tuberosity medial crest (L)	0.224	0.040	Poor	0.328	0.036	Fair	0.085	0.583	Poor

Table C.5 continued

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
Radius tuberosity medial crest (R)	0.272	0.012	Poor	0.342	0.027	Fair	0.185	0.226	Poor
Ulna olecranon spur (L)	0.304	0.005	Fair	0.415	0.007	Fair	0.173	0.262	Poor
Ulna olecranon spur (R)	0.379	<0.001	Fair	0.372	0.015	Fair	0.399	0.010	Fair
Trapezium lipping (L)	0.131	0.234	Poor	0.070	0.650	Poor	0.198	0.210	Poor
Trapezium lipping (R)	0.203	0.063	Poor	0.310	0.044	Fair	0.042	0.788	Poor
Innominate relative weight (L)	0.302	0.007	Fair	0.143	0.373	Poor	0.467	0.003	Fair
Innominate relative weight (R)	0.353	0.001	Fair	0.250	0.109	Poor	0.477	0.003	Fair
Innominate actual weight (L)	-0.175	0.121	Poor	0.157	0.326	Poor	-0.588	<0.001	Fair
Innominate actual weight (R)	-0.195	0.083	Poor	0.117	0.454	Poor	-0.555	0.001	Fair
Innominate length (L)	0.215	0.055	Poor	0.529	0.001	Fair	0.057	0.718	Poor
Innominate length (R)	0.218	0.050	Poor	0.536	0.001	Fair	0.030	0.850	Poor
Sacroiliac joint fusion (L)	0.277	0.010	Poor	0.305	0.045	Fair	0.246	0.106	Poor
Sacroiliac joint fusion (R)	0.126	0.239	Poor	0.198	0.193	Poor	0.042	0.783	Poor
Iliac crest fusion (L)	0.590	<0.001	Fair	0.578	<0.001	Fair	0.609	<0.001	Moderate
Iliac crest fusion (R)	0.580	<0.001	Fair	0.542	<0.001	Fair	0.612	<0.001	Moderate
Iliac crest tuberculum ossification (L)	0.418	<0.001	Fair	0.444	0.004	Fair	0.387	0.011	Fair
Iliac crest tuberculum ossification (R)	0.320	0.003	Fair	0.370	0.015	Fair	0.300	0.049	Fair
Ilium anterior inferior Iliac spine exostoses (L)	0.253	0.019	Poor	0.126	0.408	Poor	0.353	0.022	Fair
Ilium anterior inferior iliac spine exostoses (R)	0.116	0.283	Poor	0.137	0.375	Poor	0.117	0.442	Poor
Acetabulum posterior margin lipping (L)	0.576	<0.001	Fair	0.653	<0.001	Moderate	0.505	0.001	Fair
Acetabulum posterior margin lipping (R)	0.539	<0.001	Fair	0.659	<0.001	Moderate	0.439	0.004	Fair
Acetabulum posterior articular surface extra bone (L)	0.290	0.007	Poor	0.210	0.169	Poor	0.362	0.018	Fair
Acetabulum posterior articular surface extra bone (R)	0.312	0.004	Fair	0.300	0.049	Fair	0.324	0.034	Fair
Acetabulum inferior joint lipping (L)	0.425	<0.001	Fair	0.393	0.013	Fair	0.442	0.004	Fair
Acetabulum inferior joint lipping (R)	0.523	<0.001	Fair	0.527	0.001	Fair	0.517	0.001	Fair
Ischial tuberosity superior margin spur (L)	0.270	0.012	Poor	0.316	0.038	Fair	0.224	0.147	Poor
Ischial tuberosity superior margin spur (R)	0.298	0.006	Poor	0.362	0.017	Fair	0.222	0.145	Poor
Ischial tuberosity medial spur (L)	0.128	0.234	Poor	0.163	0.284	Poor	NA	1.000	NA

Table C.5 continued

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
Ischial tuberosity medial spur (R)	0.212	0.048	Poor	0.298	0.050	Poor	NA	1.000	NA
Ischial tuberosity bumps (L)	0.604	<0.001	Moderate	0.515	0.001	Fair	0.712	<0.001	Moderate
Ischial tuberosity bumps (R)	0.564	<0.001	Fair	0.479	0.002	Fair	0.652	<0.001	Moderate
Auricular surface inferior surface porosity (L)	0.235	0.034	Poor	0.239	0.130	Poor	0.248	0.112	Poor
Auricular surface inferior surface porosity (R)	0.097	0.376	Poor	<0.001	1.000	None	0.187	0.231	Poor
Auricular surface superior posterior iliac exostoses (L)	0.389	<0.001	Fair	0.421	0.008	Fair	0.329	0.033	Fair
Auricular surface superior posterior iliac exostoses (R)	0.427	<0.001	Fair	0.471	0.002	Fair	0.352	0.023	Fair
Auricular surface inferior posterior iliac exostoses (L)	0.255	0.026	Poor	0.268	0.107	Poor	0.240	0.135	Poor
Auricular surface inferior posterior iliac exostoses (R)	0.226	0.041	Poor	0.185	0.236	Poor	0.260	0.100	Poor
Auricular surface posterior exostoses (L)	0.350	0.002	Fair	0.447	0.005	Fair	0.259	0.102	Poor
Auricular surface posterior exostoses (R)	0.510	<0.001	Fair	0.547	<0.001	Fair	0.475	0.002	Fair
Pubic symphyseal collar (L)	0.073	0.517	Poor	0.264	0.090	Poor	-0.150	0.362	Poor
Pubic symphyseal collar (R)	0.071	0.534	Poor	0.111	0.488	Poor	NA	NA	NA
Pubic symphyseal relief (L)	0.583	<0.001	Fair	0.597	<0.001	Fair	0.603	<0.001	Moderate
Pubic symphyseal relief (R)	0.669	<0.001	Moderate	0.675	<0.001	Moderate	0.696	<0.001	Moderate
Pubic superior protuberance/apex (L)	0.523	<0.001	Fair	0.463	0.005	Fair	0.612	<0.001	Moderate
Pubic superior protuberance/apex (R)	0.564	<0.001	Fair	0.516	0.002	Fair	0.641	<0.001	Moderate
Pubic symphysis ventral margin (L)	0.570	<0.001	Fair	0.526	0.001	Fair	0.686	<0.001	Moderate
Pubic symphysis ventral margin (R)	0.644	<0.001	Moderate	0.677	<0.001	Moderate	0.649	<0.001	Moderate
Pubic symphysis dorsal margin (L)	0.469	<0.001	Fair	0.553	0.001	Fair	0.407	0.013	Fair
Pubic symphysis dorsal margin (R)	0.594	<0.001	Fair	0.686	<0.001	Moderate	0.507	0.002	Fair
C1 lipping	0.539	<0.001	Fair	0.590	<0.001	Fair	0.464	0.006	Fair
C1 eburnation	0.332	0.003	Fair	0.343	0.023	Fair	0.329	0.055	Fair
L1 spinous process flattening	0.657	<0.001	Moderate	0.604	<0.001	Moderate	0.716	<0.001	Moderate
L1 superior epiphyseal ring fusion	0.404	<0.001	Fair	0.499	0.001	Fair	0.264	0.091	Poor
L1 inferior epiphyseal ring fusion	0.384	0.001	Fair	0.497	0.001	Fair	0.240	0.129	Poor
L1 superior surface morphology	0.394	<0.001	Fair	0.493	0.002	Fair	0.255	0.103	Poor
L1 inferior surface morphology	0.390	<0.001	Fair	0.527	0.001	Fair	0.231	0.144	Poor

Table C.5 continued

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
L5 superior epiphyseal fusion	0.487	<0.001	Fair	0.58	<0.001	Fair	0.355	0.025	Fair
L5 inferior epiphyseal fusion	0.482	<0.001	Fair	0.529	0.001	Fair	0.421	0.008	Fair
L5 superior surface morphology	0.400	<0.001	Fair	0.437	0.005	Fair	0.353	0.025	Fair
L5 inferior surface morphology	0.449	<0.001	Fair	0.528	0.001	Fair	0.356	0.026	Fair
L5 superior margin shape	0.696	<0.001	Moderate	0.772	<0.001	Moderate	0.643	<0.001	Moderate
L5 inferior margin shape	0.474	<0.001	Fair	0.529	0.001	Fair	0.376	0.024	Fair
S1 margin shape	0.559	<0.001	Fair	0.550	0.001	Fair	0.592	<0.001	Fair
S1-2 fusion	0.739	<0.001	Moderate	0.732	<0.001	Moderate	0.756	<0.001	Moderate
Sacral elbow	0.289	0.008	Poor	0.386	0.012	Fair	0.147	0.354	Poor
C lipping	0.541	<0.001	Fair	0.546	<0.001	Fair	0.551	<0.001	Fair
C candlewax	0.540	<0.001	Fair	0.557	<0.001	Fair	0.533	0.001	Fair
T lipping	0.603	<0.001	Moderate	0.537	0.001	Fair	0.665	<0.001	Moderate
T candlewax	0.504	<0.001	Fair	0.520	0.001	Fair	0.482	0.002	Fair
L lipping	0.625	<0.001	Moderate	0.672	<0.001	Moderate	0.608	<0.001	Moderate
L candlewax	0.625	<0.001	Moderate	0.702	<0.001	Moderate	0.584	<0.001	Fair
DISH	0.407	<0.001	Fair	0.458	0.003	Fair	0.339	0.028	Fair
R1 fusion (L)	0.244	0.031	Poor	0.214	0.176	Poor	0.262	0.110	Poor
R1 fusion (R)	0.203	0.073	Poor	NA	NA	NA	0.305	0.060	Fair
Sternum central dorsal ridge	-0.232	0.070	Poor	NA	NA	NA	-0.317	0.093	Fair
R2 rim profile (L)	0.320	0.032	Fair	0.260	0.203	Poor	0.36	0.108	Fair
R2 rim profile (R)	0.317	0.048	Fair	0.373	0.104	Fair	0.200	0.383	Poor
R2 ossification pattern (L)	0.261	0.080	Poor	0.318	0.120	Fair	0.222	0.321	Poor
R2 ossification pattern (R)	0.300	0.061	Poor	0.521	0.023	Fair	0.058	0.801	Poor
R3-10 rim profile (L)	0.517	<0.001	Fair	0.651	<0.001	Moderate	0.372	0.030	Fair
R3-10 rim profile (R)	0.487	<0.001	Fair	0.623	<0.001	Moderate	0.328	0.056	Fair
R3-10 ossification pattern (L)	0.378	0.001	Fair	0.594	<0.001	Fair	0.054	0.754	Poor
R3-10 ossification pattern (R)	0.314	0.007	Fair	0.482	0.003	Fair	0.053	0.758	Poor
R2-10 shingle ribs (L)	-0.054	0.644	Poor	-0.062	0.696	Poor	NA	NA	NA

Table C.5 continued

Trait	Full sample			Male sub-sample			Female sub-sample		
	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹	Rho	p-value	Strength ¹
R2-10 shingle ribs (R)	NA	NA	NA	NA	NA	NA	NA	NA	NA
L Parietal depression	0.133	0.227	Poor	0.170	0.266	Poor	0.095	0.553	Poor
R Parietal depression	0.140	0.200	Poor	0.155	0.309	Poor	0.121	0.446	Poor
Spheno-occipital synchondrosis	0.241	0.027	Poor	0.340	0.026	Fair	0.154	0.329	Poor
L Occipital condyle lipping	0.033	0.766	Poor	0.058	0.708	Poor	0.009	0.956	Poor
R Occipital condyle lipping	0.034	0.755	Poor	0.318	0.037	Fair	-0.201	0.209	Poor

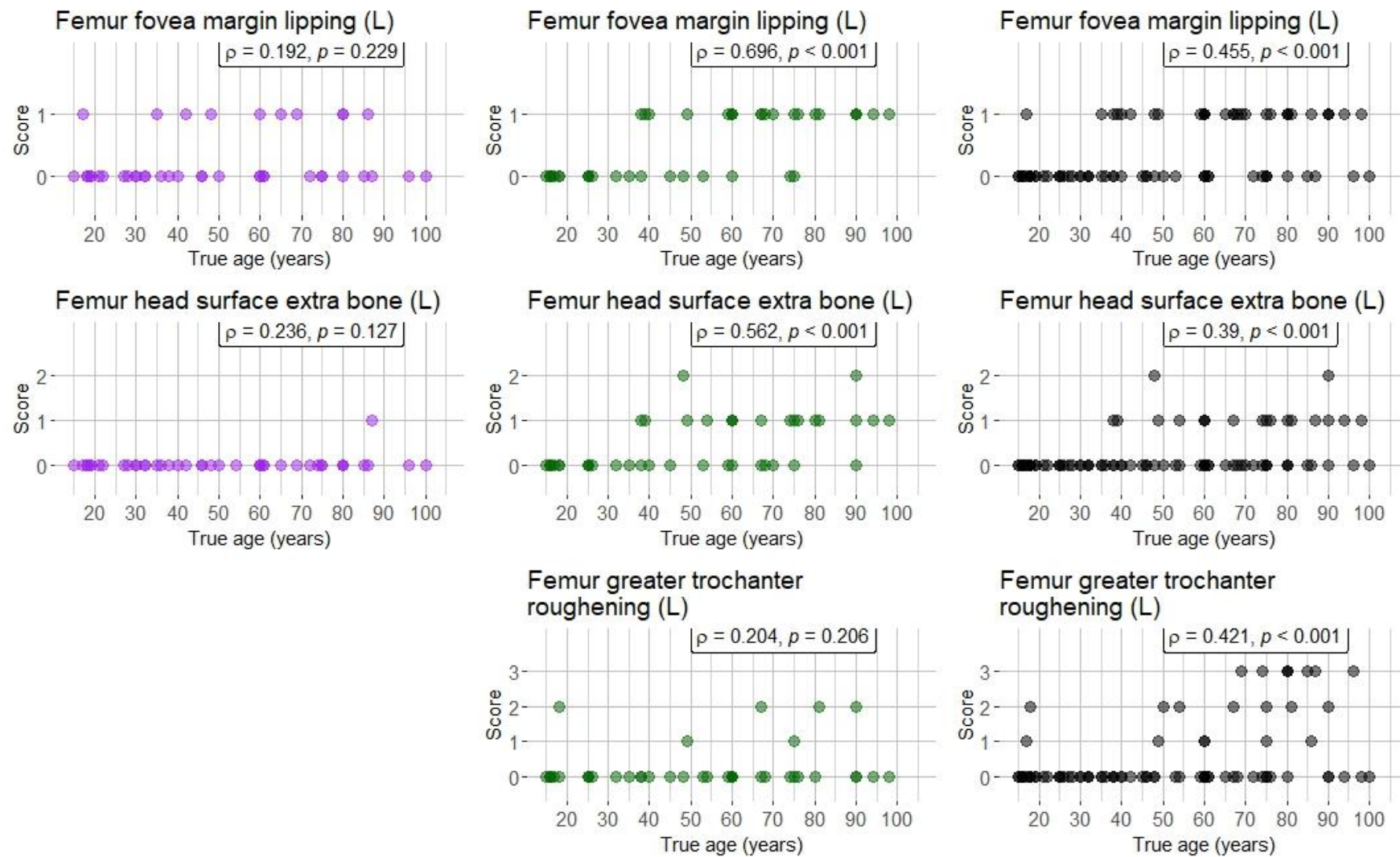


Fig. C.1 Scatterplots of trait score vs age for the left femur fovea head margin lipping, head extra bone, and greater trochanter roughening for the female (n = 44), male (n = 55), and pooled-sex (n = 89) samples from left to right

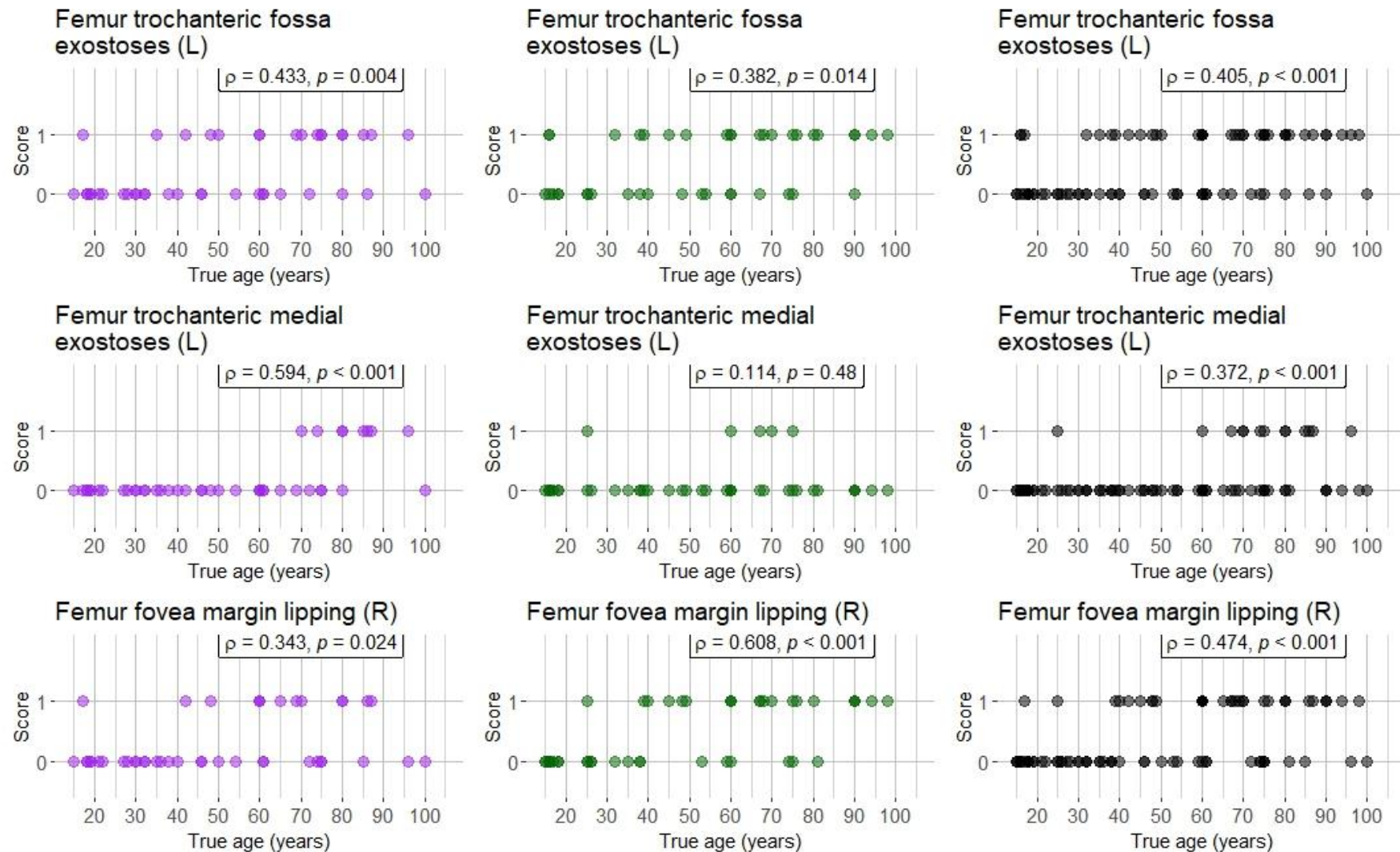


Fig. C.2 Scatterplots of trait score vs age for left femur trochanteric fossa exostoses, trochanteric medial exostoses, and right femur fovea margin for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

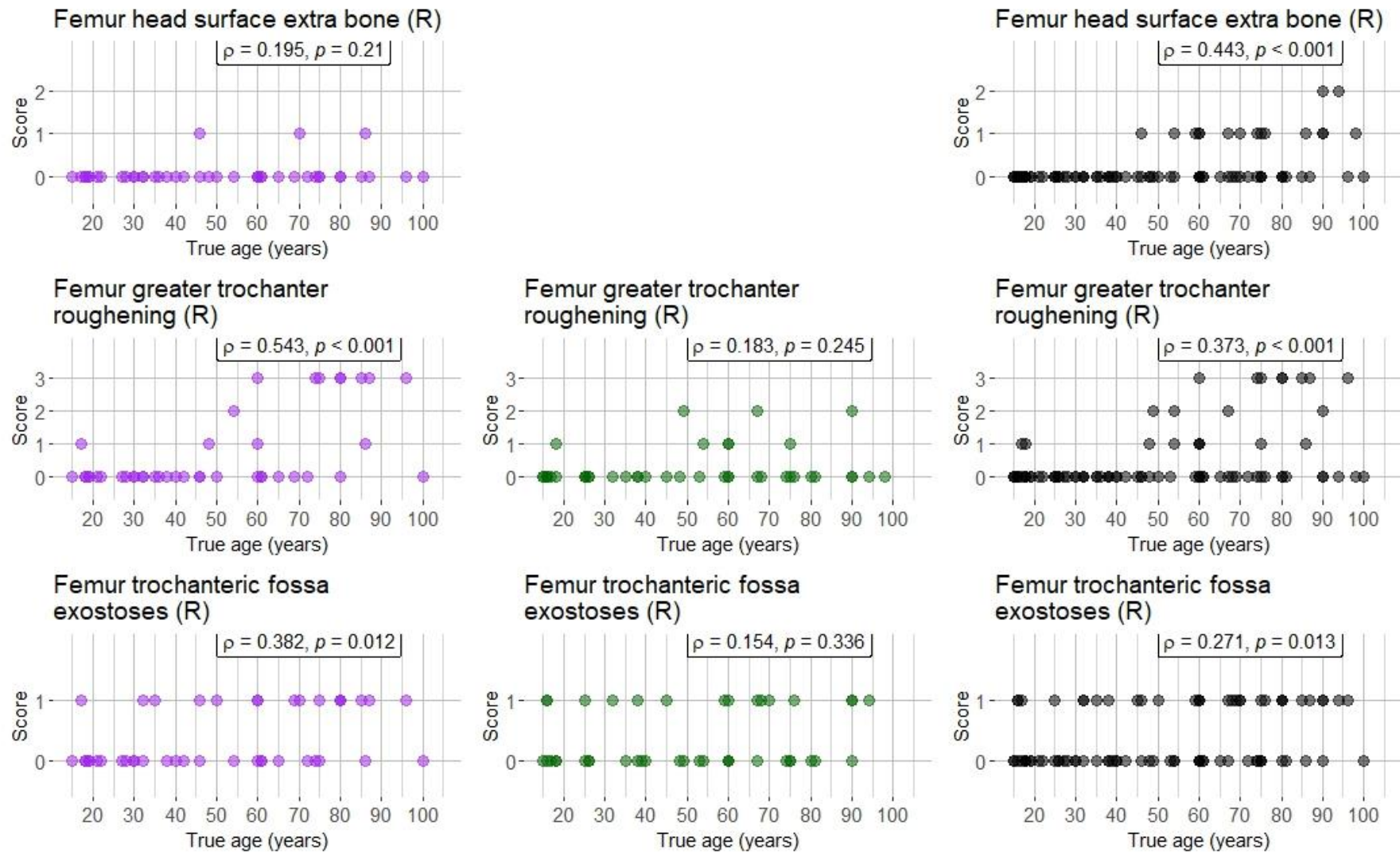


Fig. C.3 Scatterplots of trait score vs age for right femur head extra bone, greater trochanter roughening, and trochanteric fossa exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

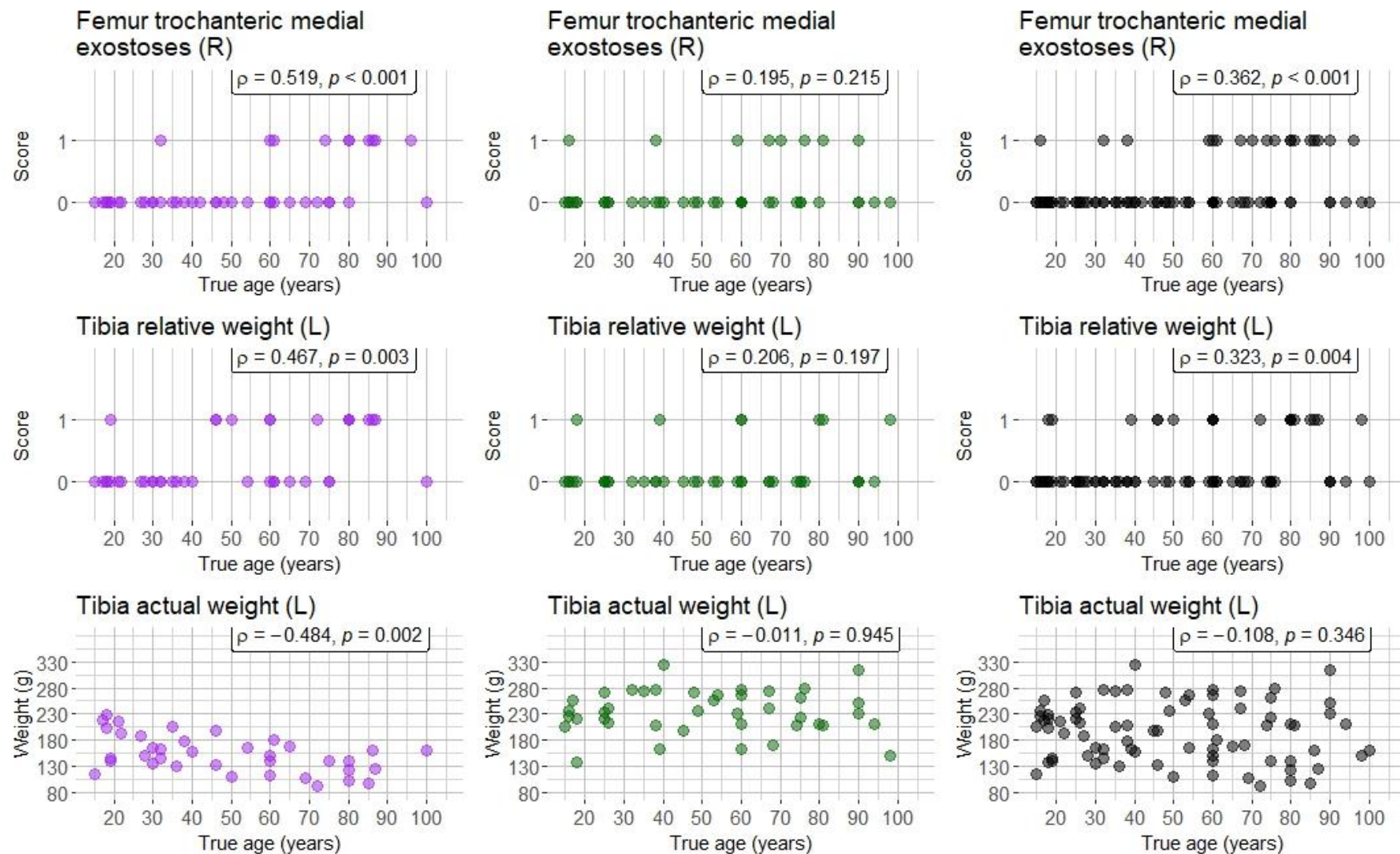


Fig. C.4 Scatterplots of trait score vs age for right femur trochanteric medial exostoses, left tibia relative weight, and actual weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

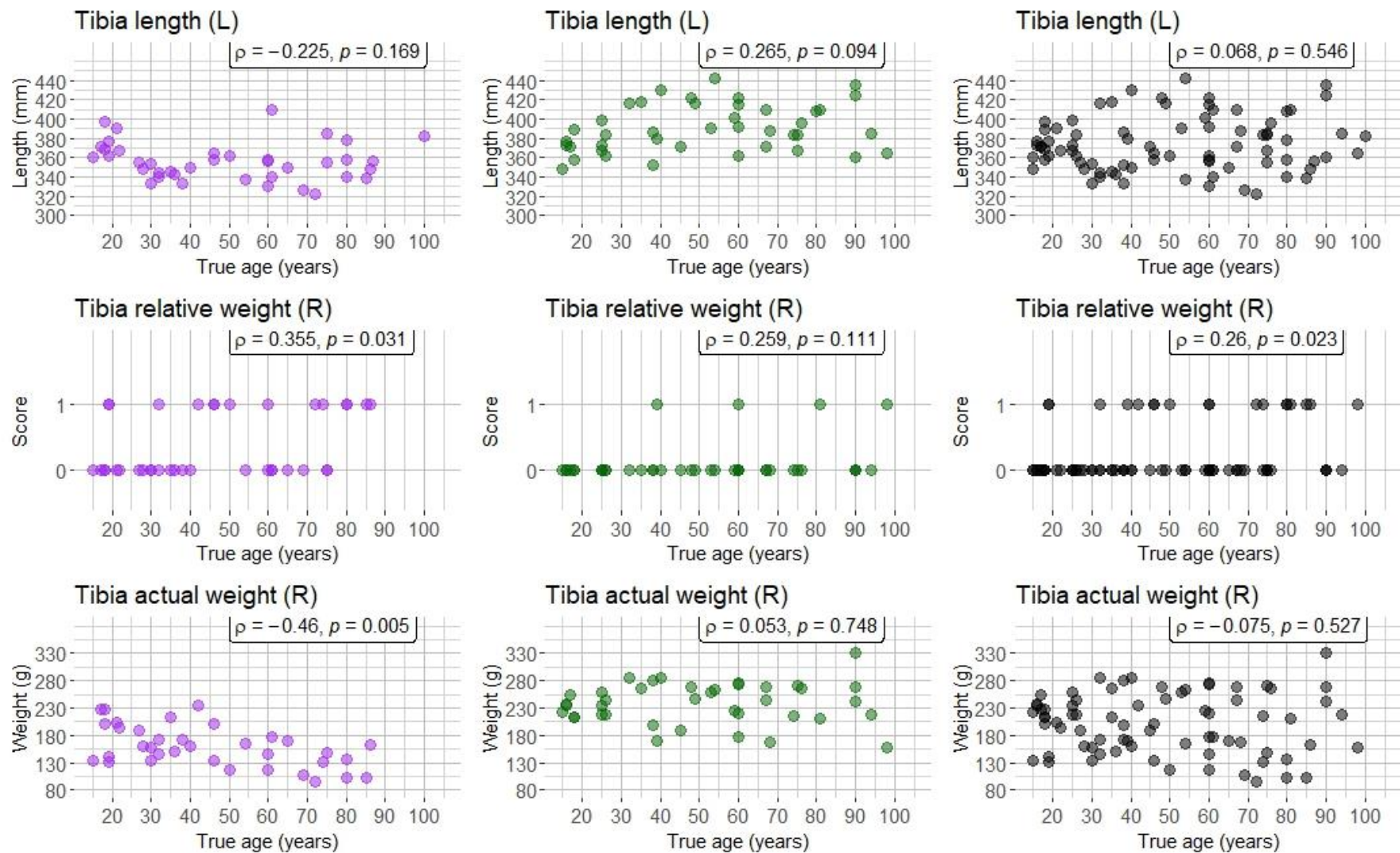


Fig. C.5 Scatterplots of trait score vs age for left tibia length, right tibia relative weight, and actual weight for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

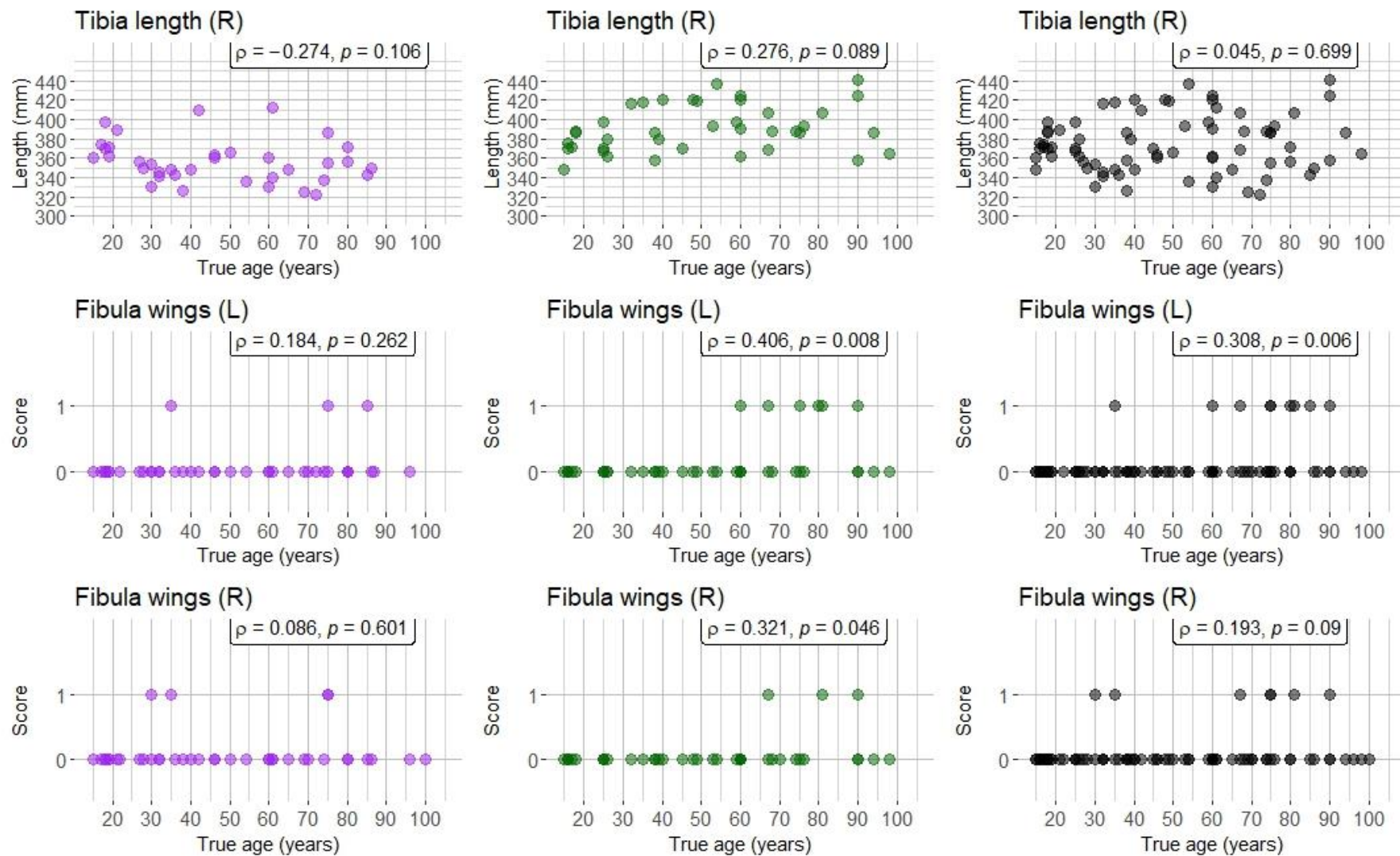


Fig. C.6 Scatterplots of trait score vs age for right tibia length, left fibula wings, and right fibula wings for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

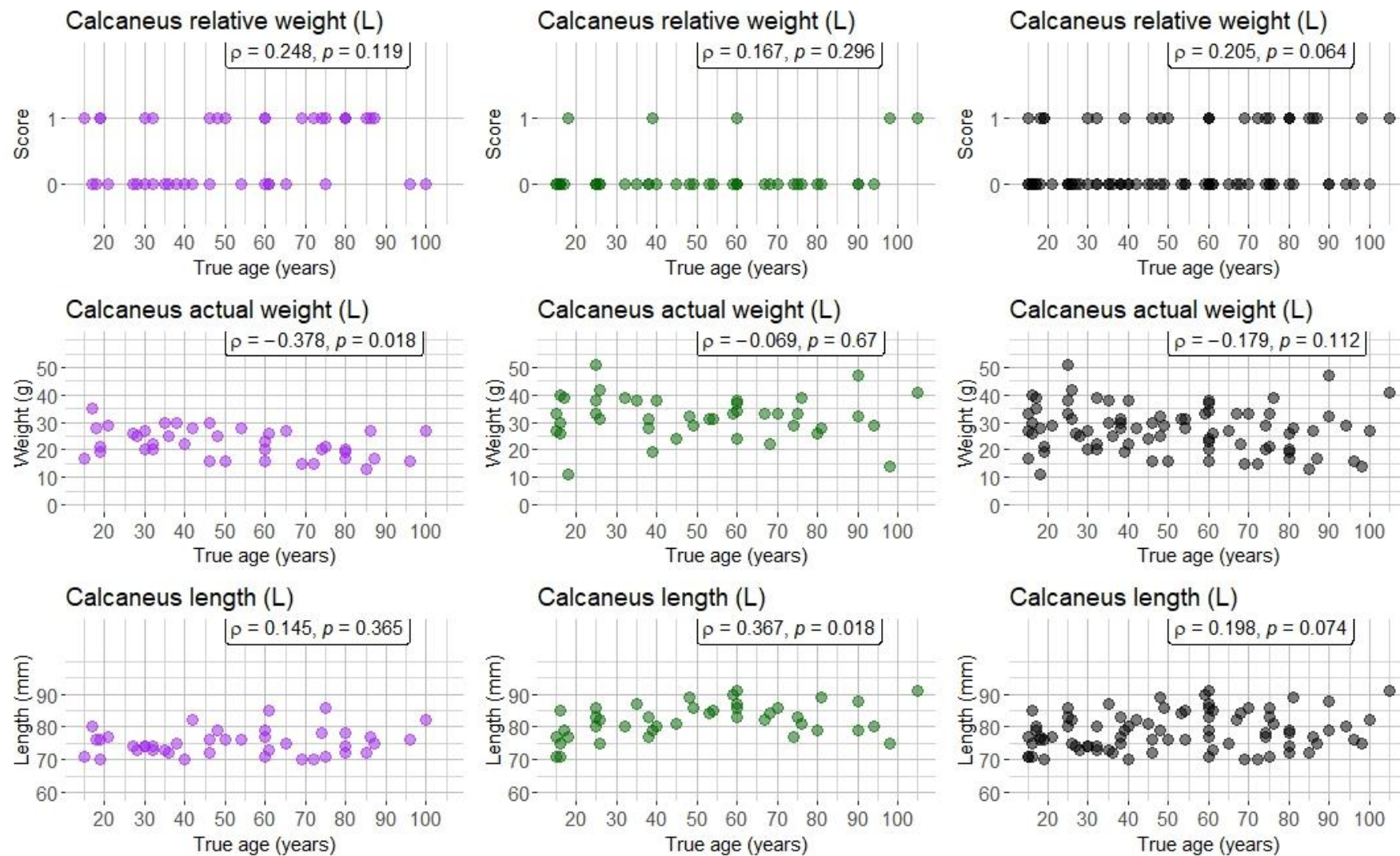


Fig. C.7 Scatterplots for trait score vs age for left calcaneus relative weight, actual weight, and length for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

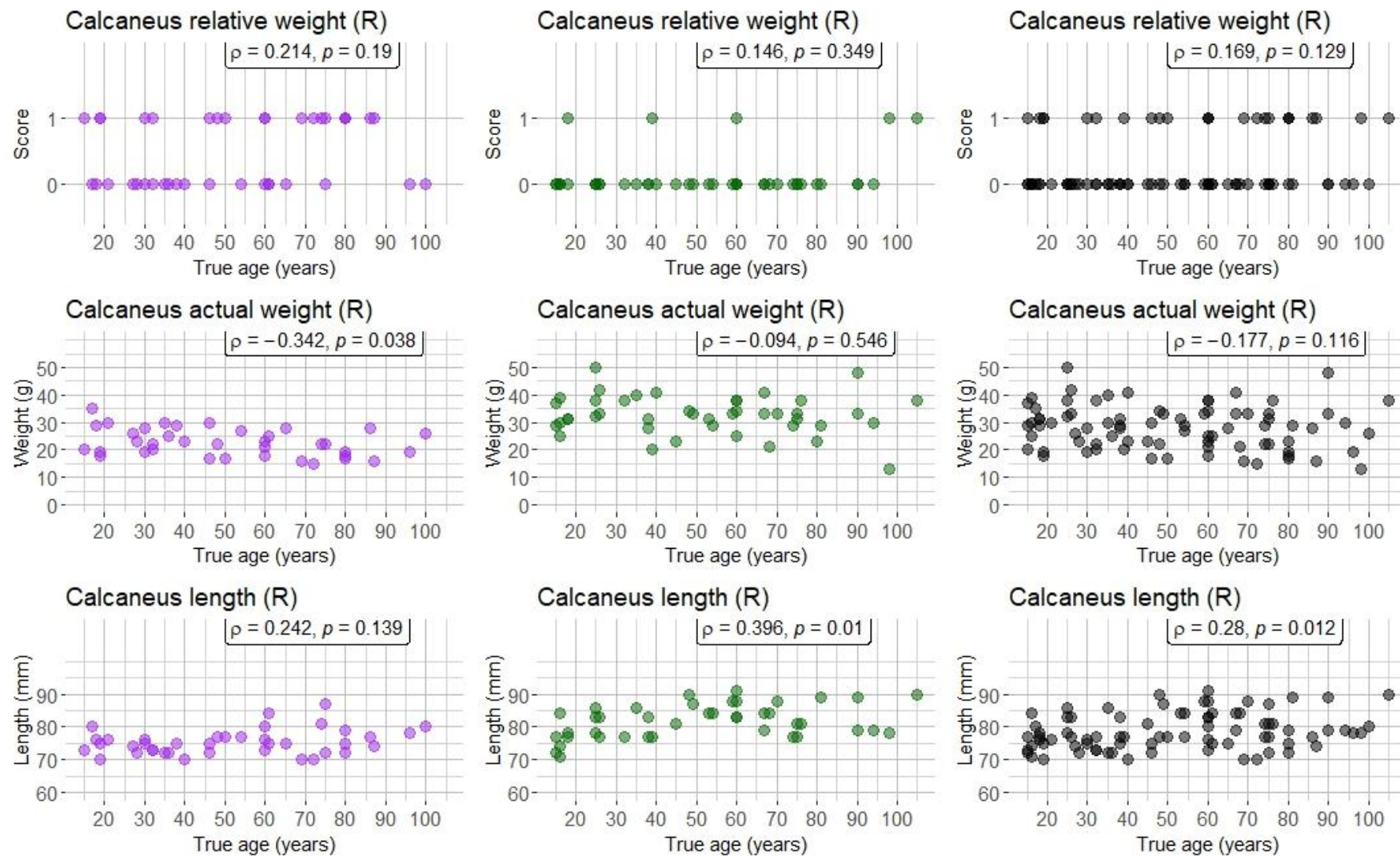


Fig. C.8 Scatterplots for trait score vs age for right calcaneus relative weight, actual weight, and length for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

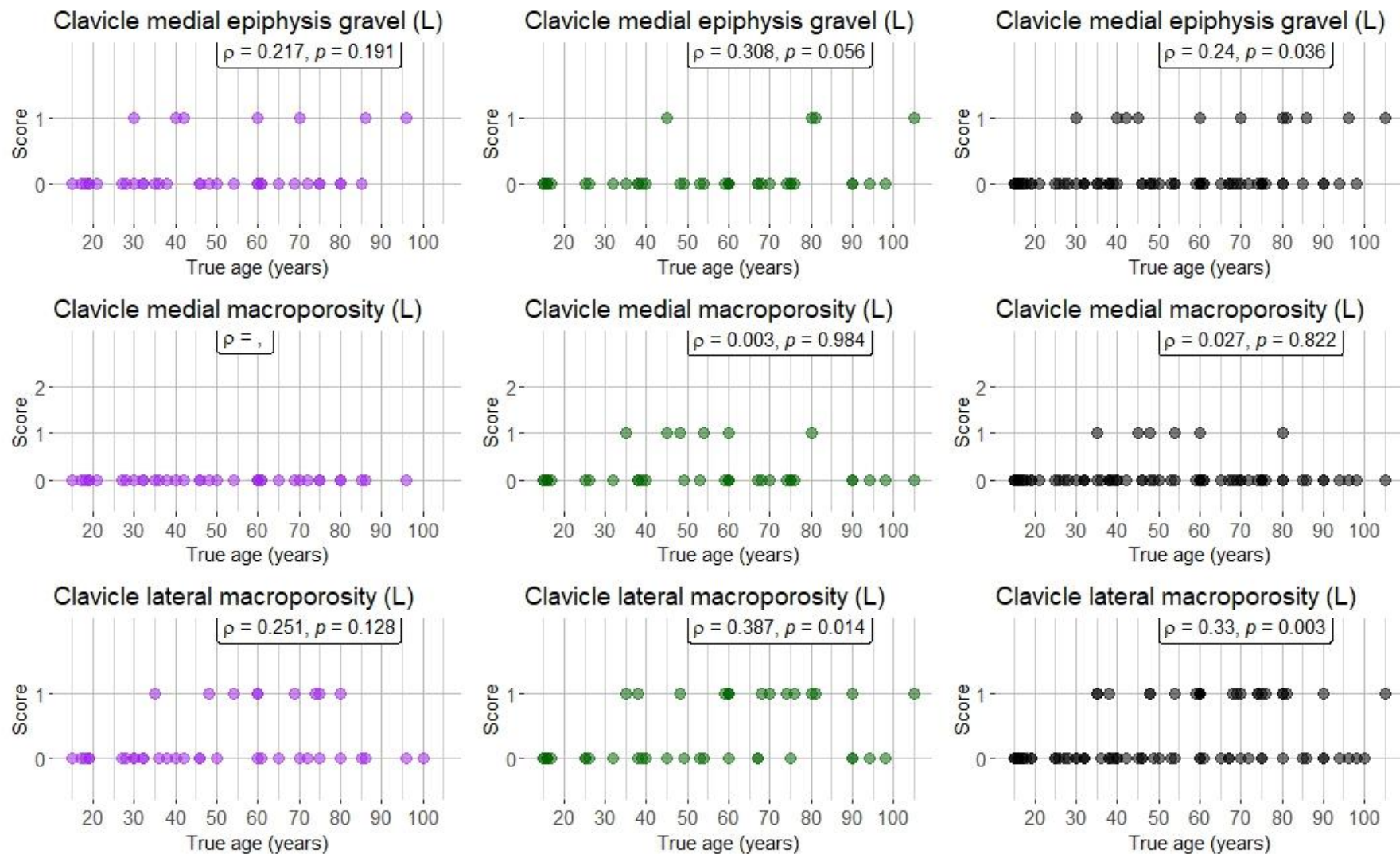


Fig. C.9 Scatterplots of trait score vs age for left clavicle medial epiphysis gravel, medial macroporosity, and lateral macroporosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

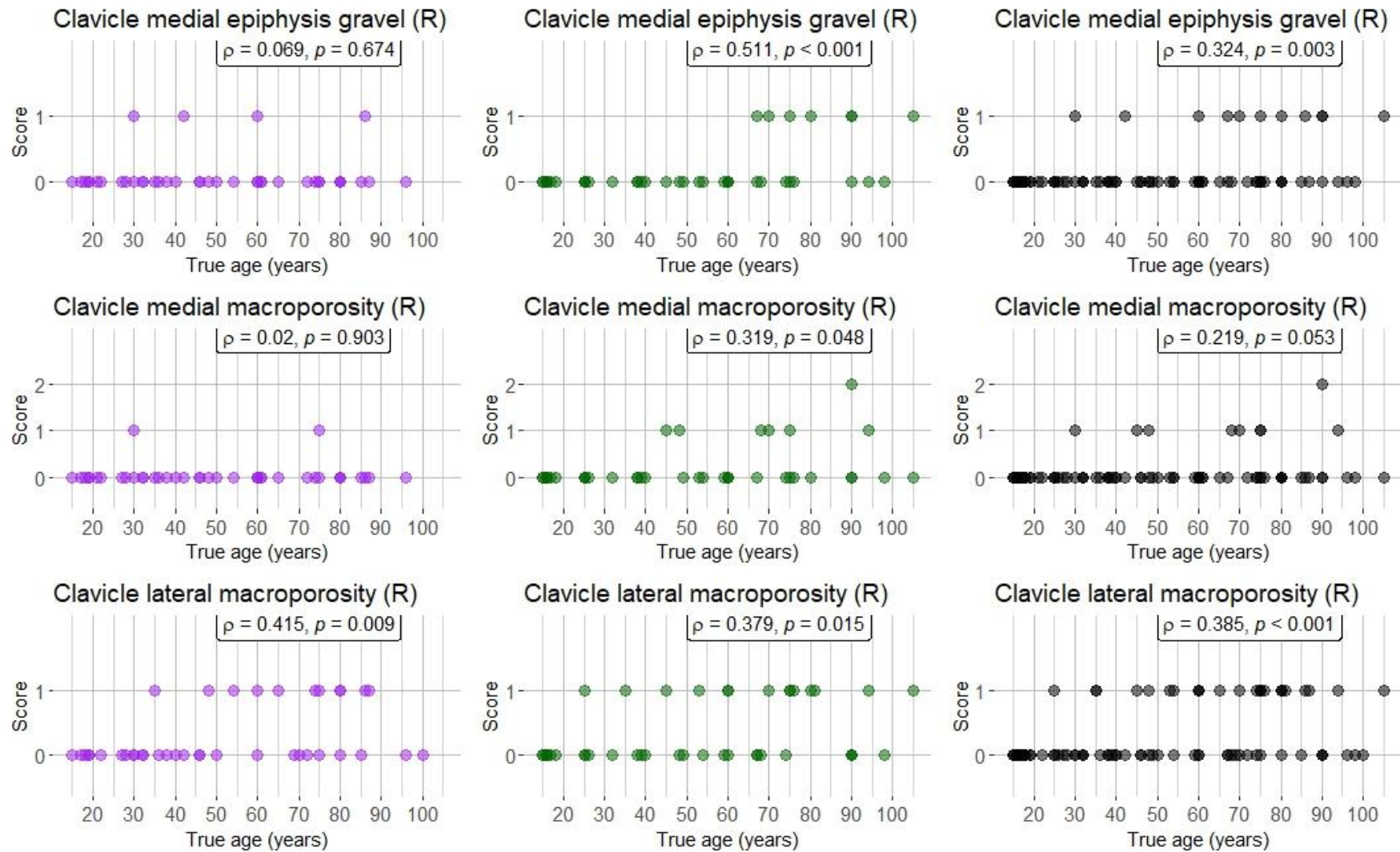


Fig. C.10 Scatterplots of trait score vs age for right clavicle medial epiphysis gravel, medial macroporosity, and lateral macroporosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

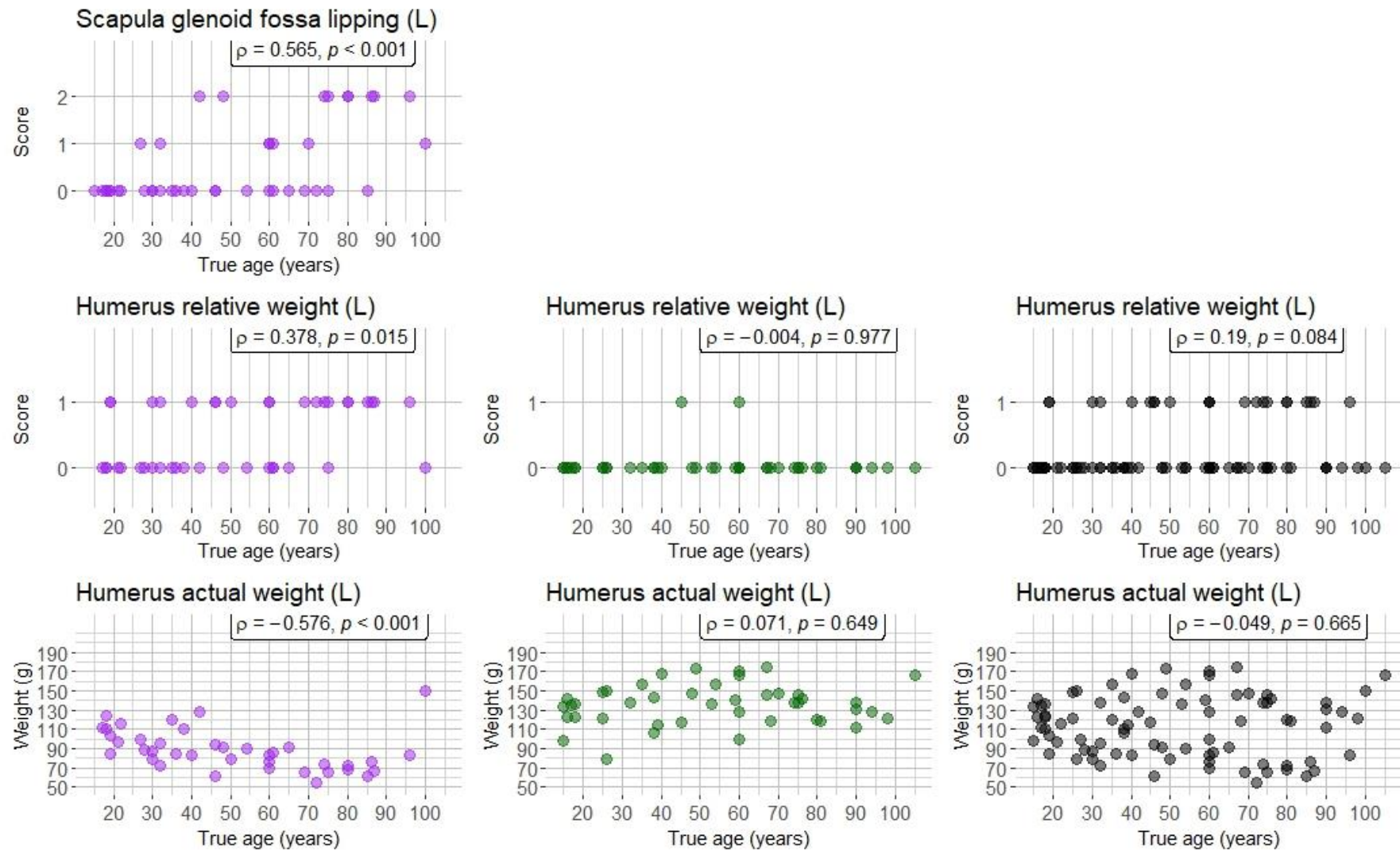


Fig. C.11 Scatterplots of trait score vs age for left scapula glenoid fossa lipping, humerus relative weight, and actual weight for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

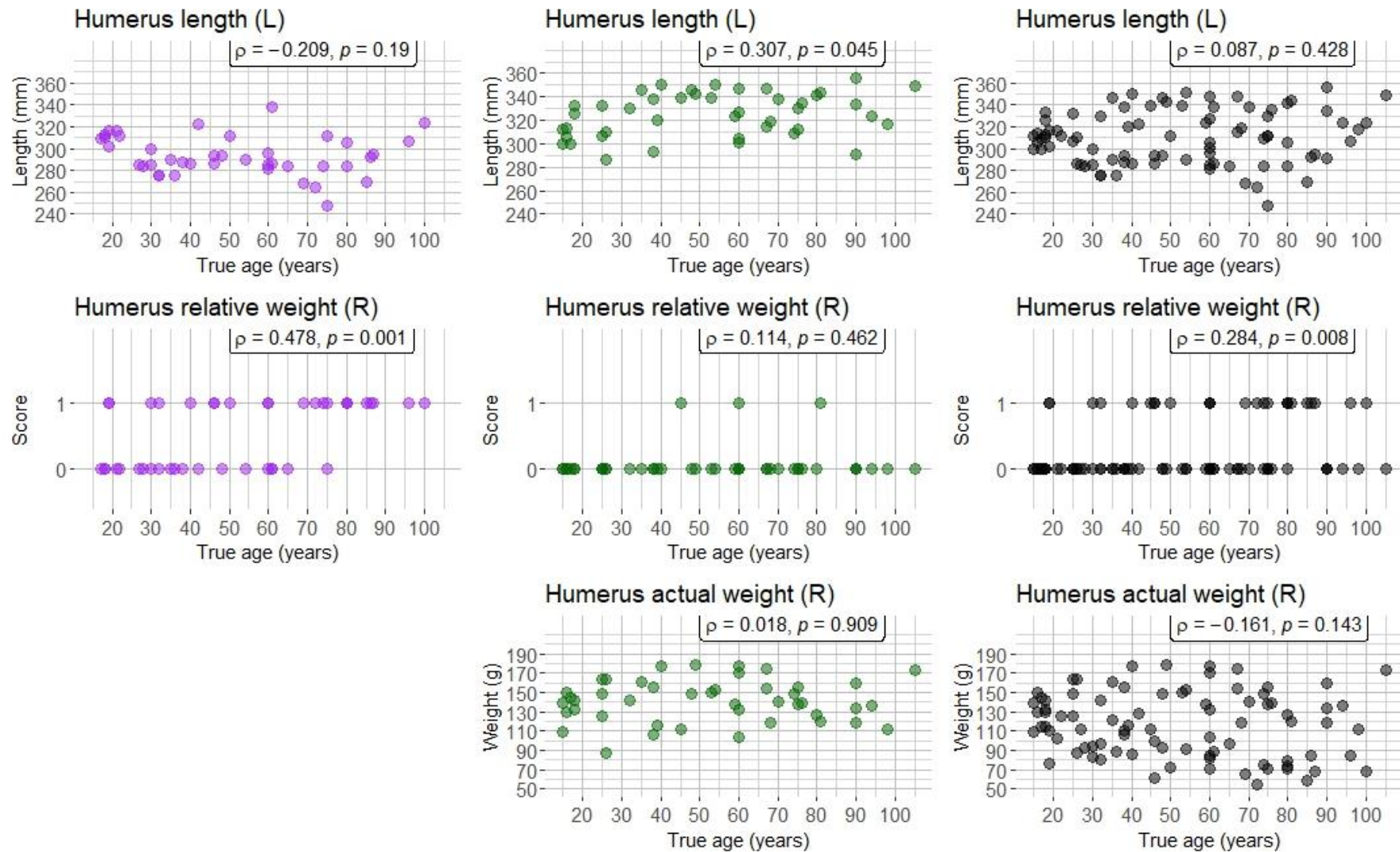


Fig. C.12 Scatterplots of trait score vs age for left humerus length, right humerus relative weight, and actual weight for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

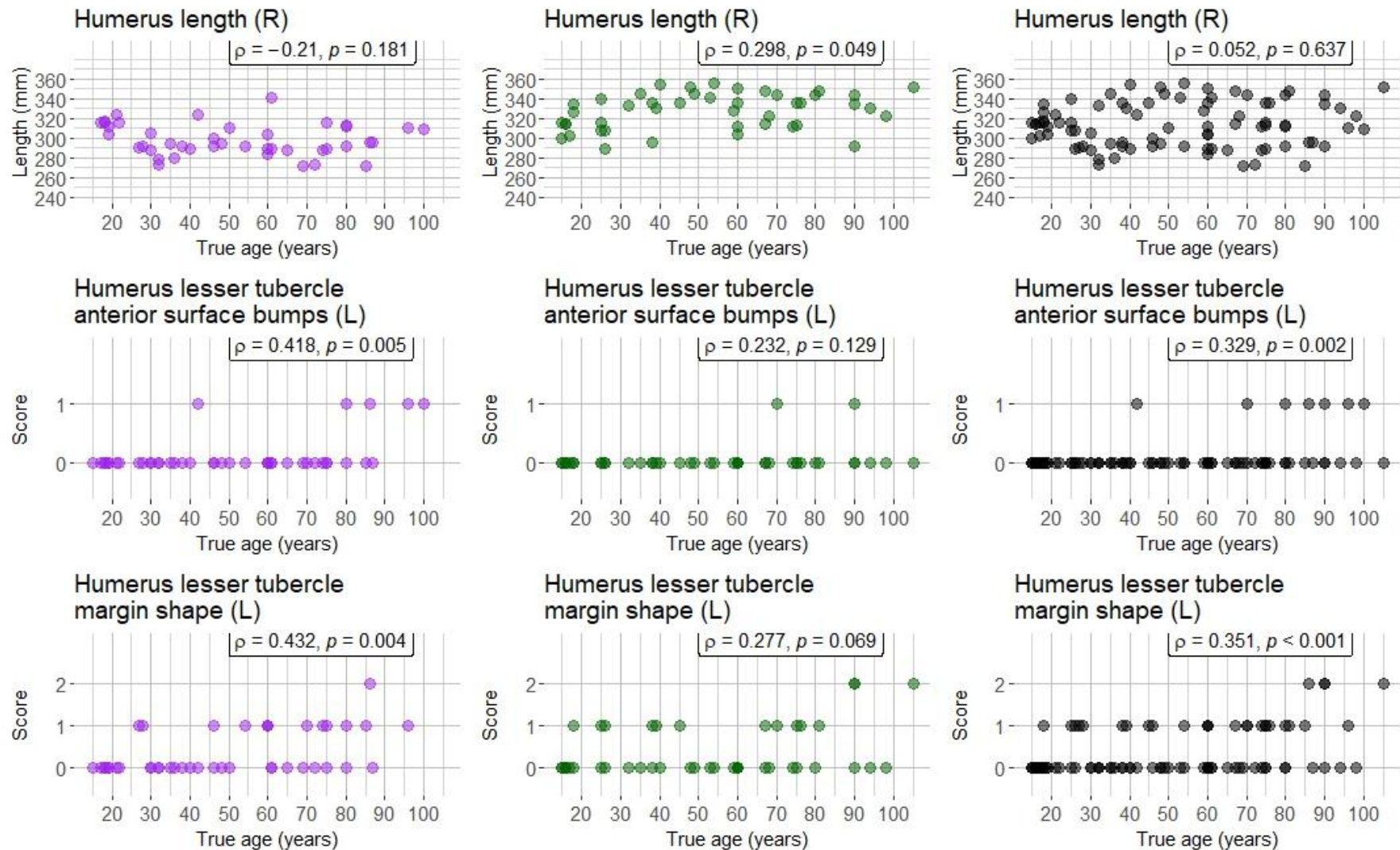


Fig. C.13 Scatterplots of trait score vs age for right humerus length, left humerus lesser tubercle anterior surface bumps, and lesser tubercle margin shape for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

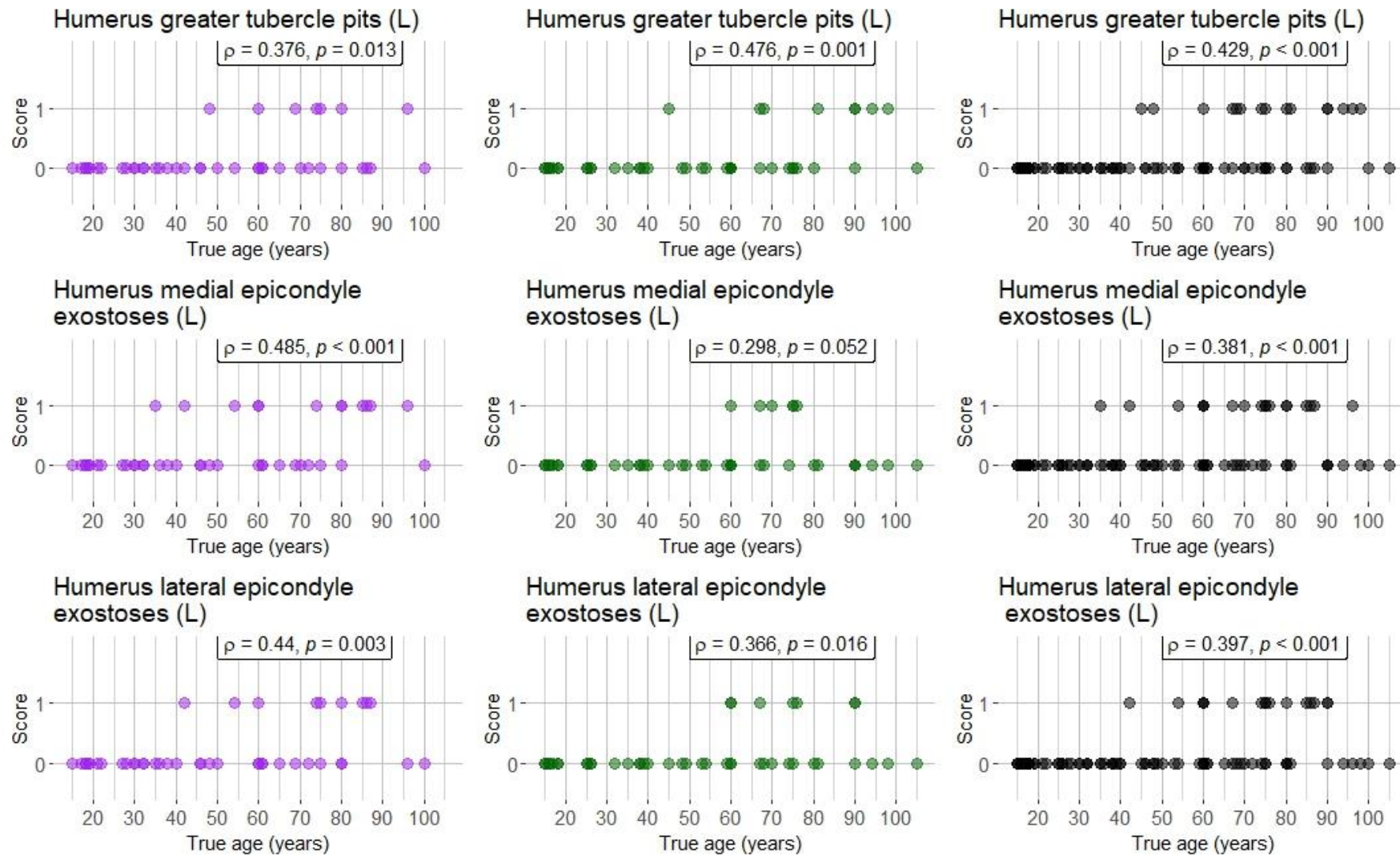


Fig. C.14 Scatterplots of trait score vs age for left humerus greater tubercle pits, medial epicondyle, and lateral epicondyle for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

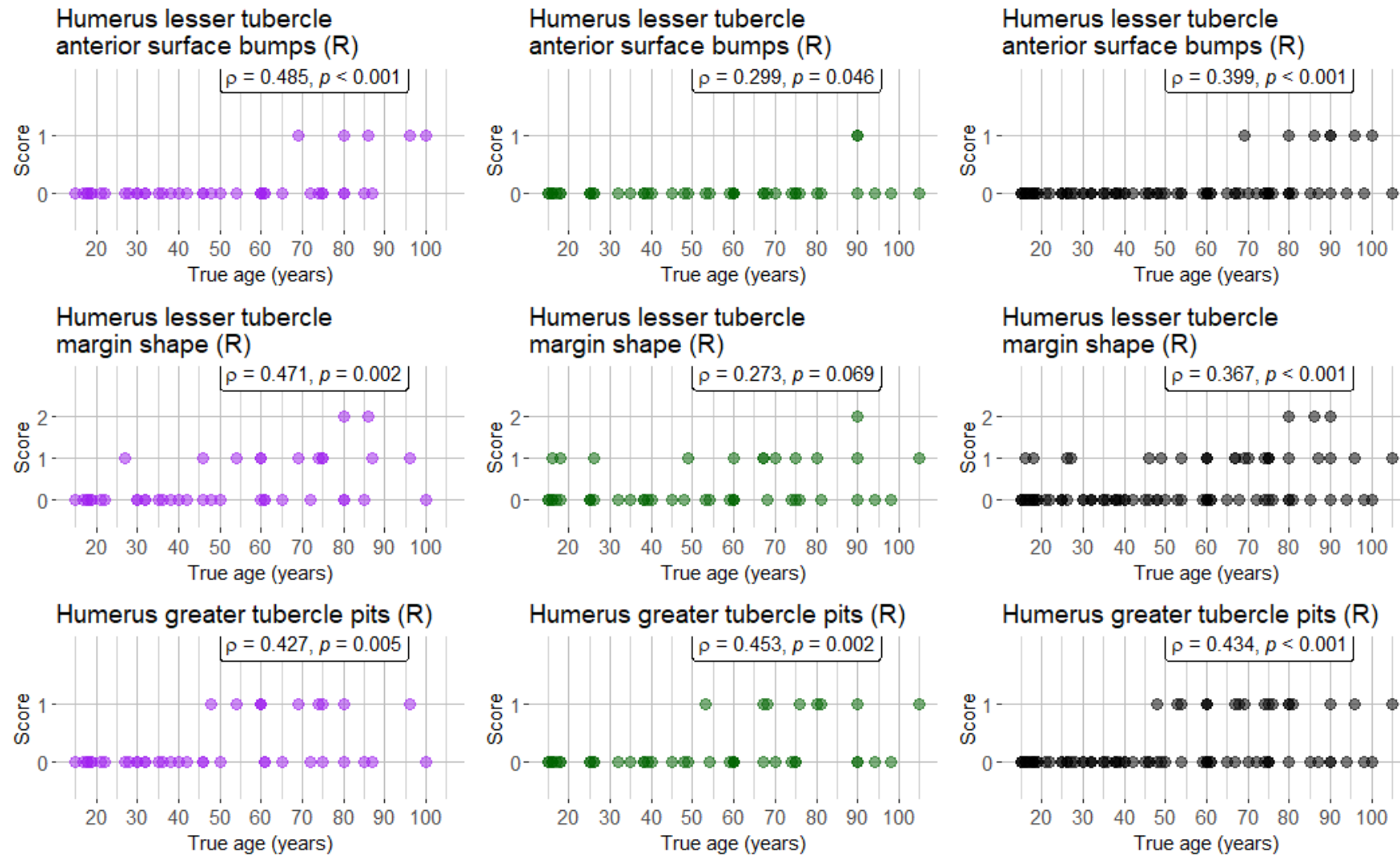


Fig. C.15 Scatterplots of trait score vs age for right humerus lesser tubercle anterior surface bumps, lesser tubercle margin shape, and greater tubercle pits for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

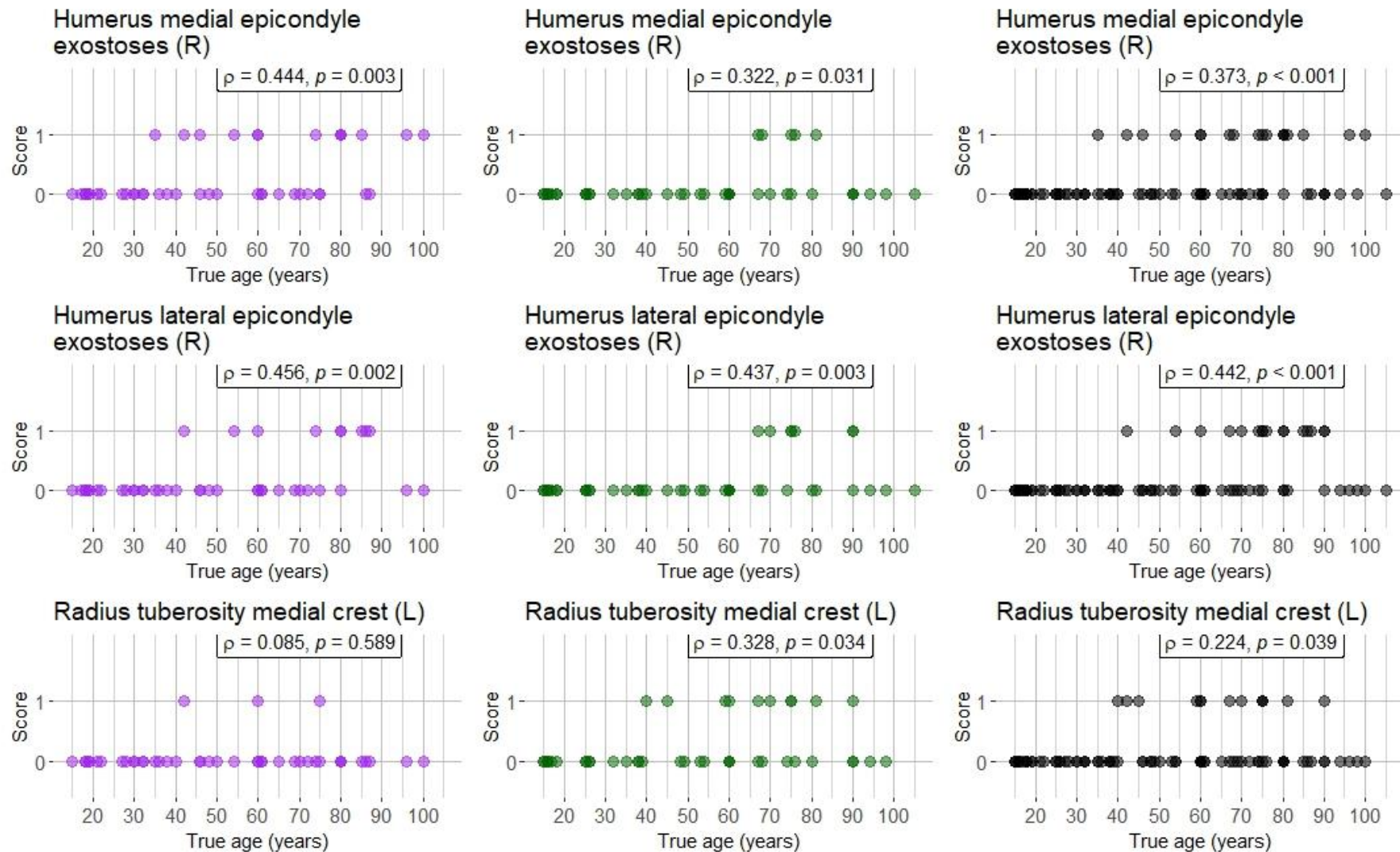


Fig. C.16 Scatterplots for trait score vs age for right humerus epicondyle, lateral epicondyle, and left radius tuberosity medial crest for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right:

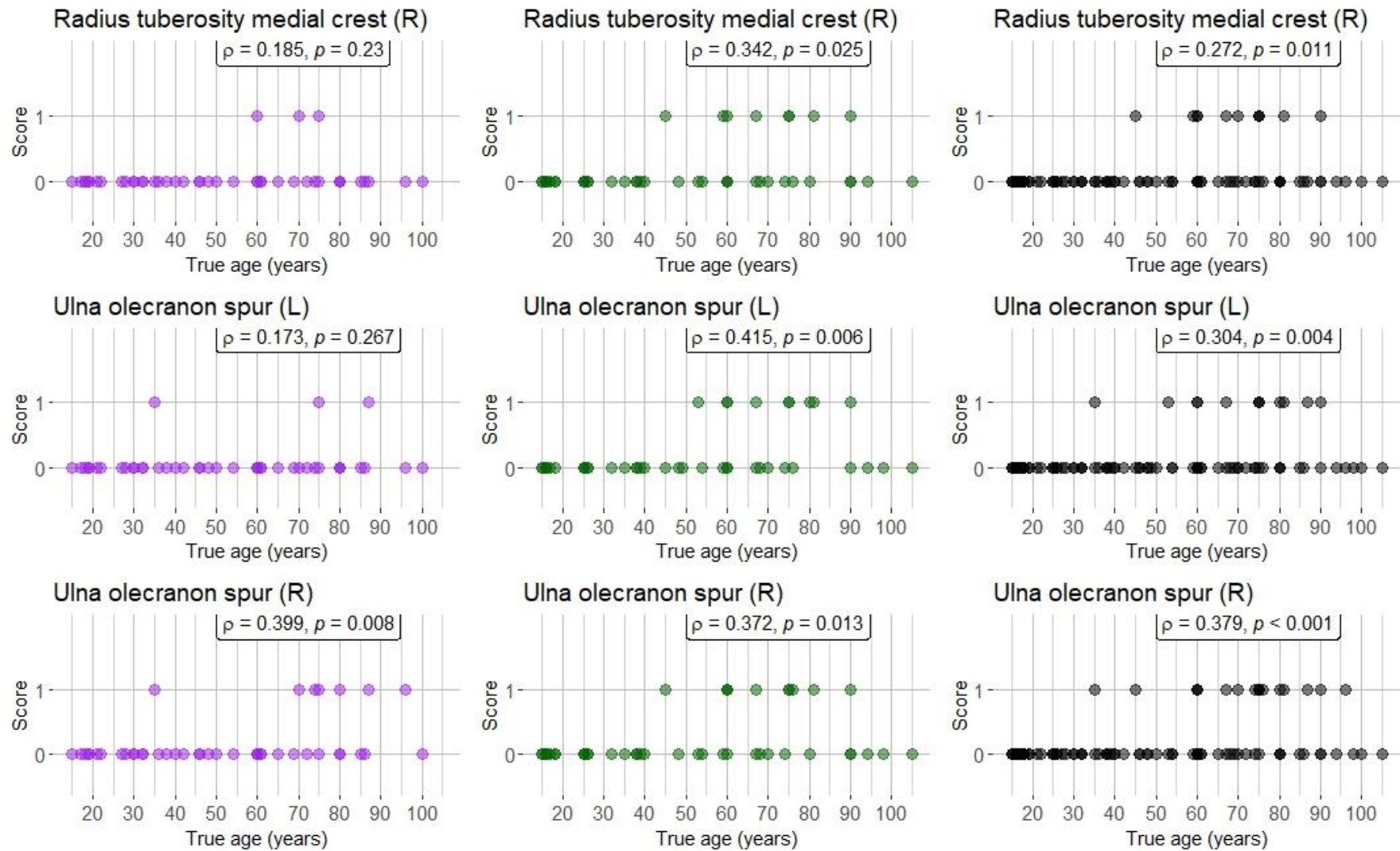


Fig. C.17 Scatterplots for trait score vs age for right radius tuberosity medial crest, left ulna olecranon spur, and right ulna olecranon spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

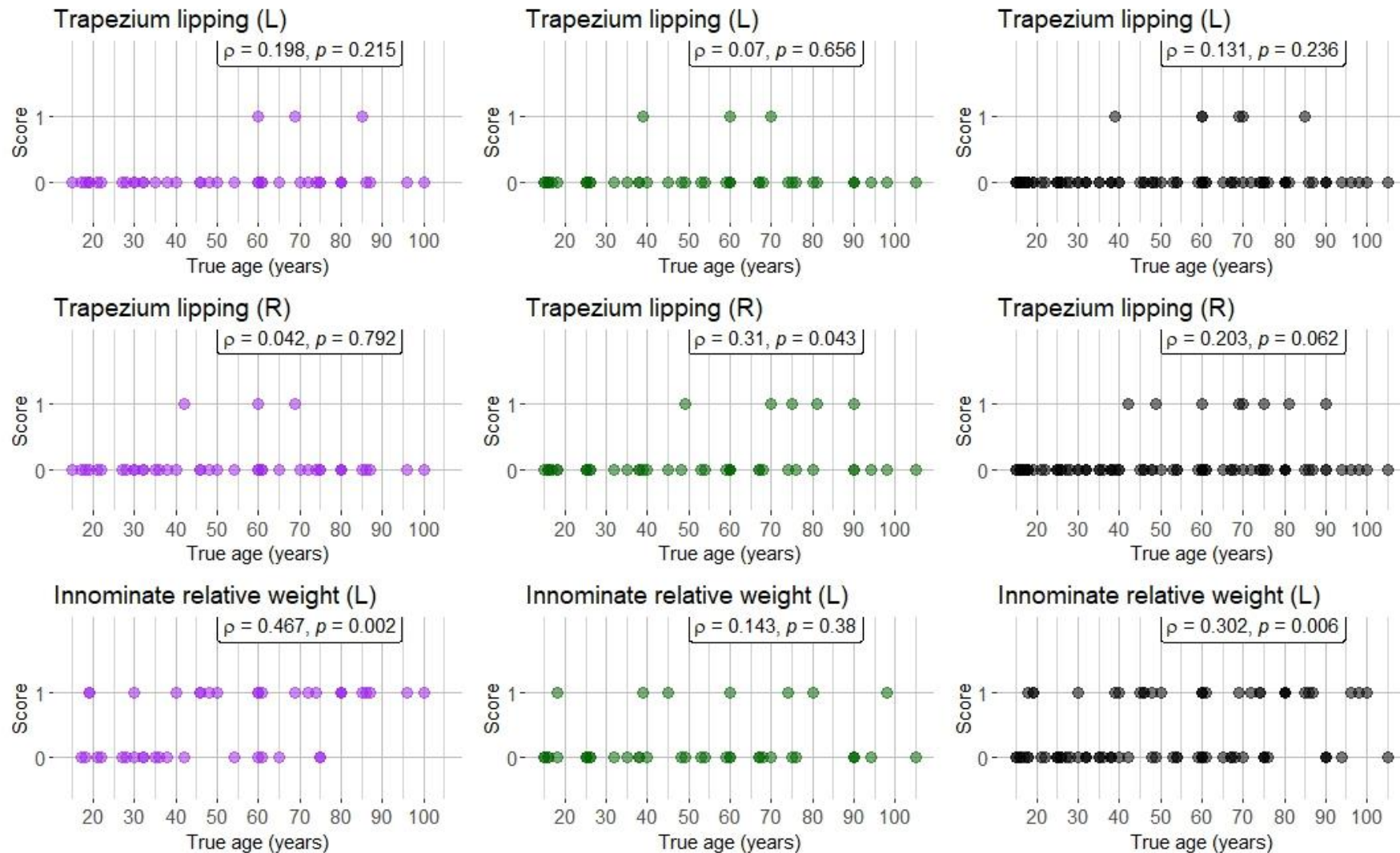


Fig. C.18 Scatterplots for trait score vs age for left trapezium lipping, right trapezium lipping, and left innominate relative weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

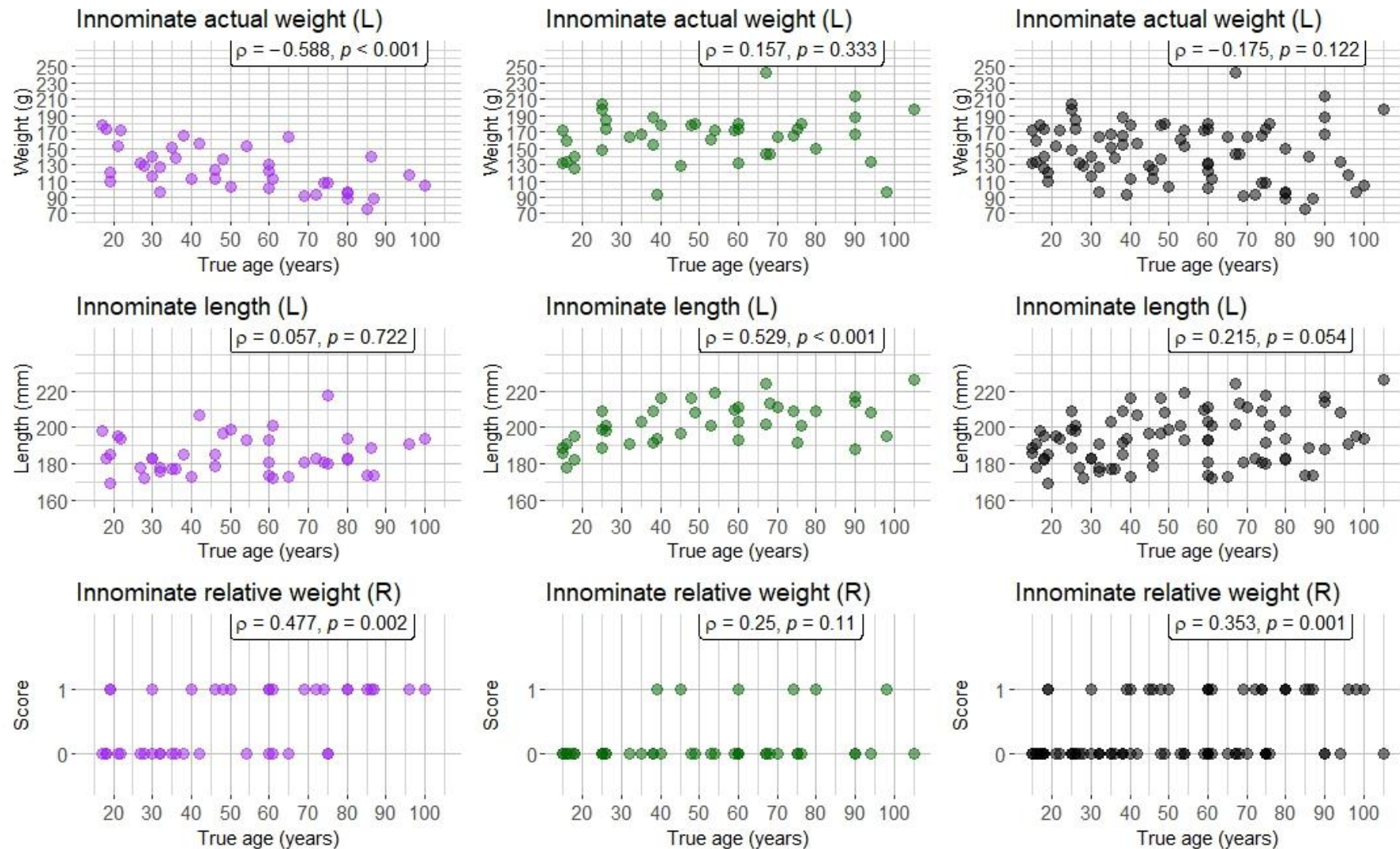


Fig. C.19 Scatterplots for trait score vs age for left innominate actual weight, length, and right innominate relative weight for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

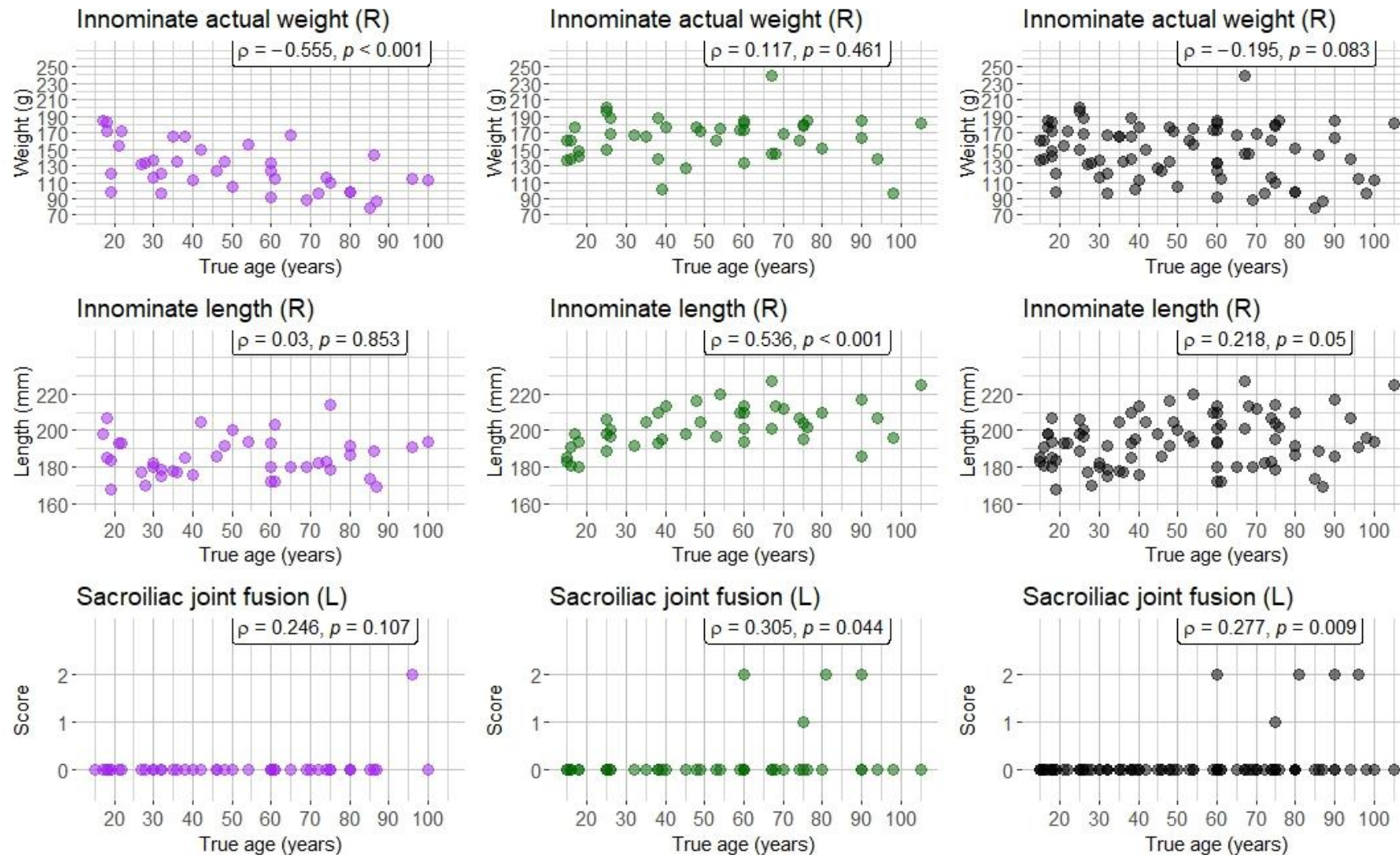


Fig. C.20 Scatterplots for trait score vs age for right innominate actual weight, length, and left sacroiliac joint fusion for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

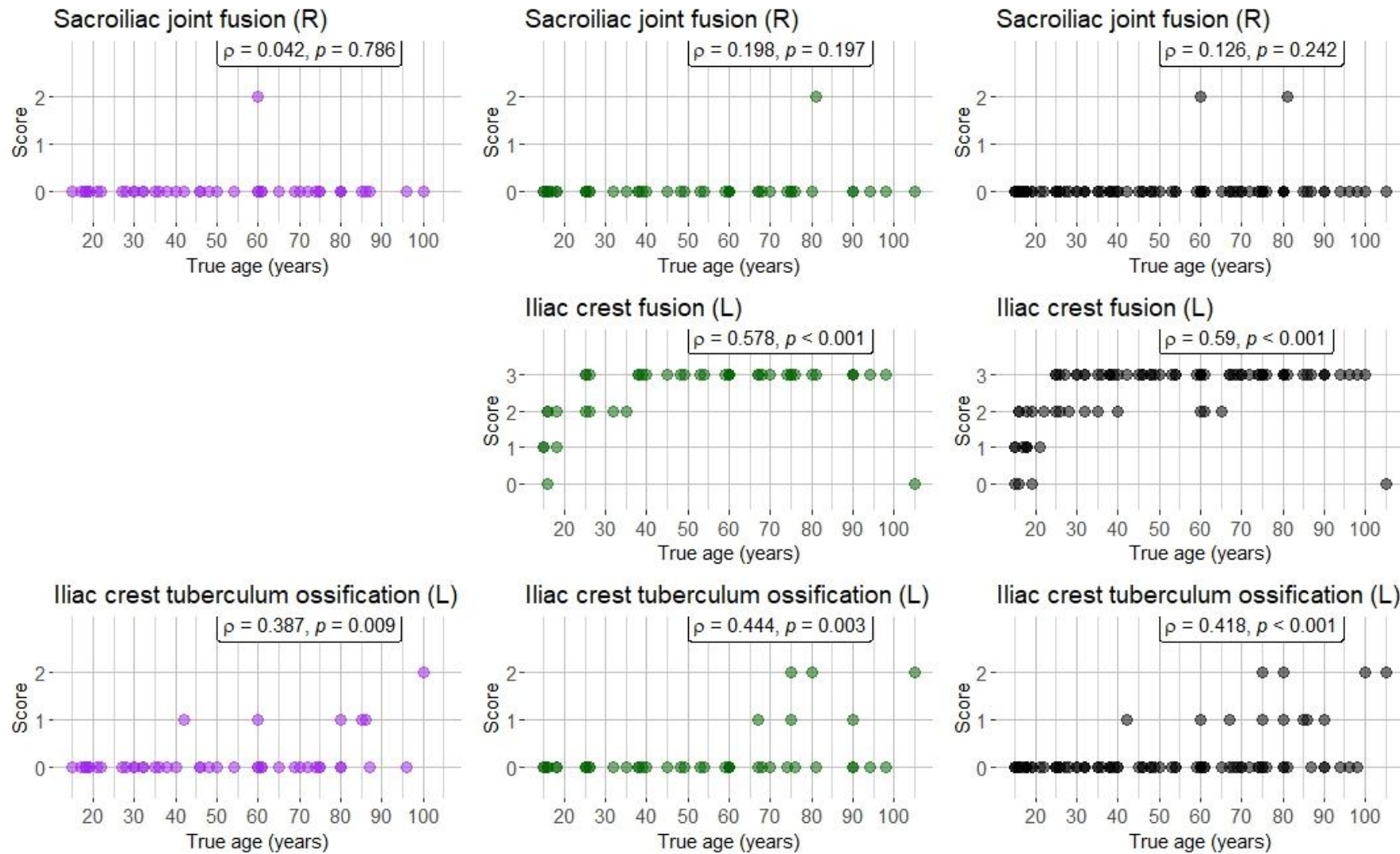


Fig. C.21 Scatterplots for trait score vs age for right sacroiliac joint fusion, left iliac crest fusion, and tuberculum ossification for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

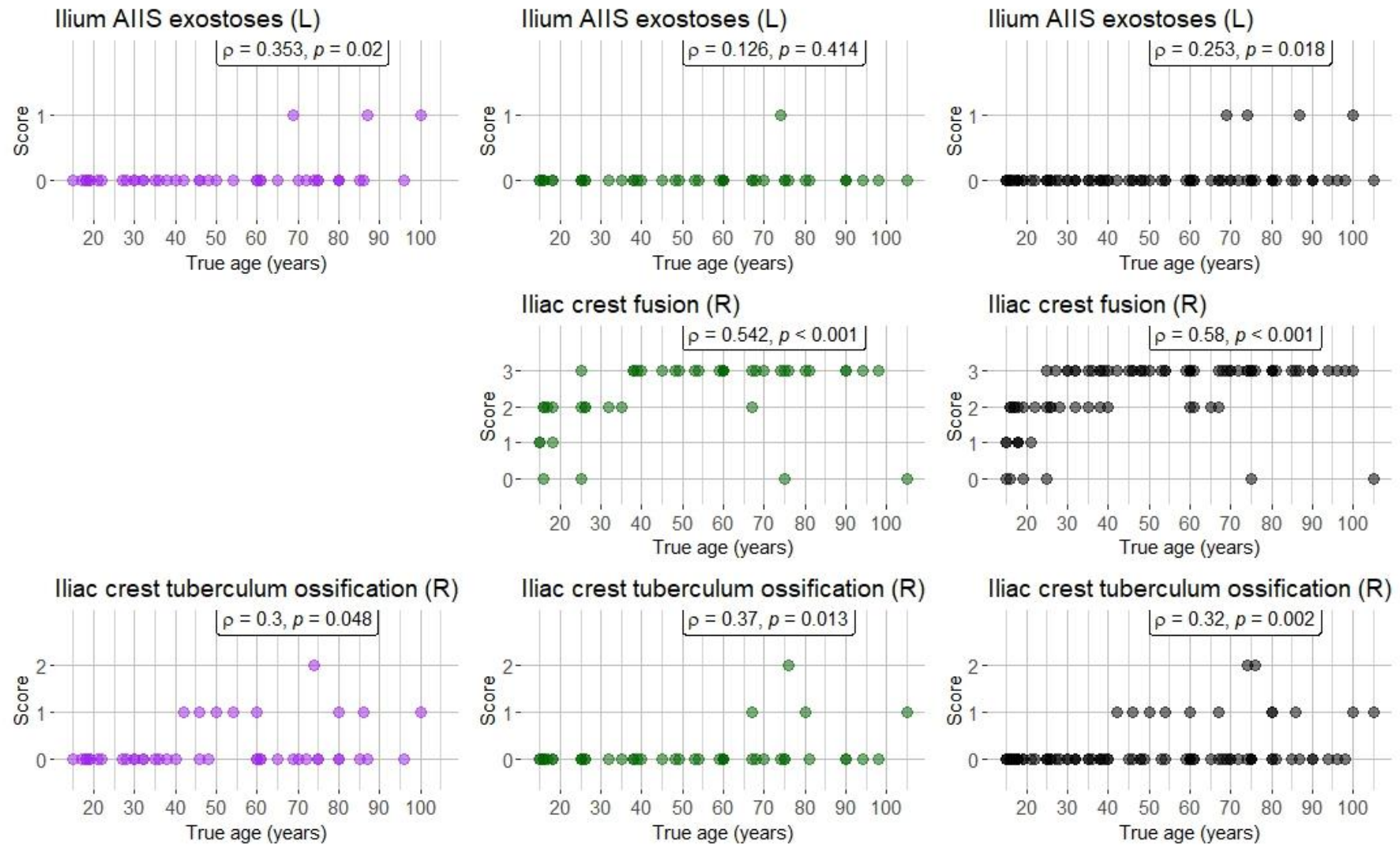


Fig. C.22 Scatterplots for trait score vs age for left ilium anterior inferior iliac exostoses, right iliac crest fusion, and tuberculum ossification for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

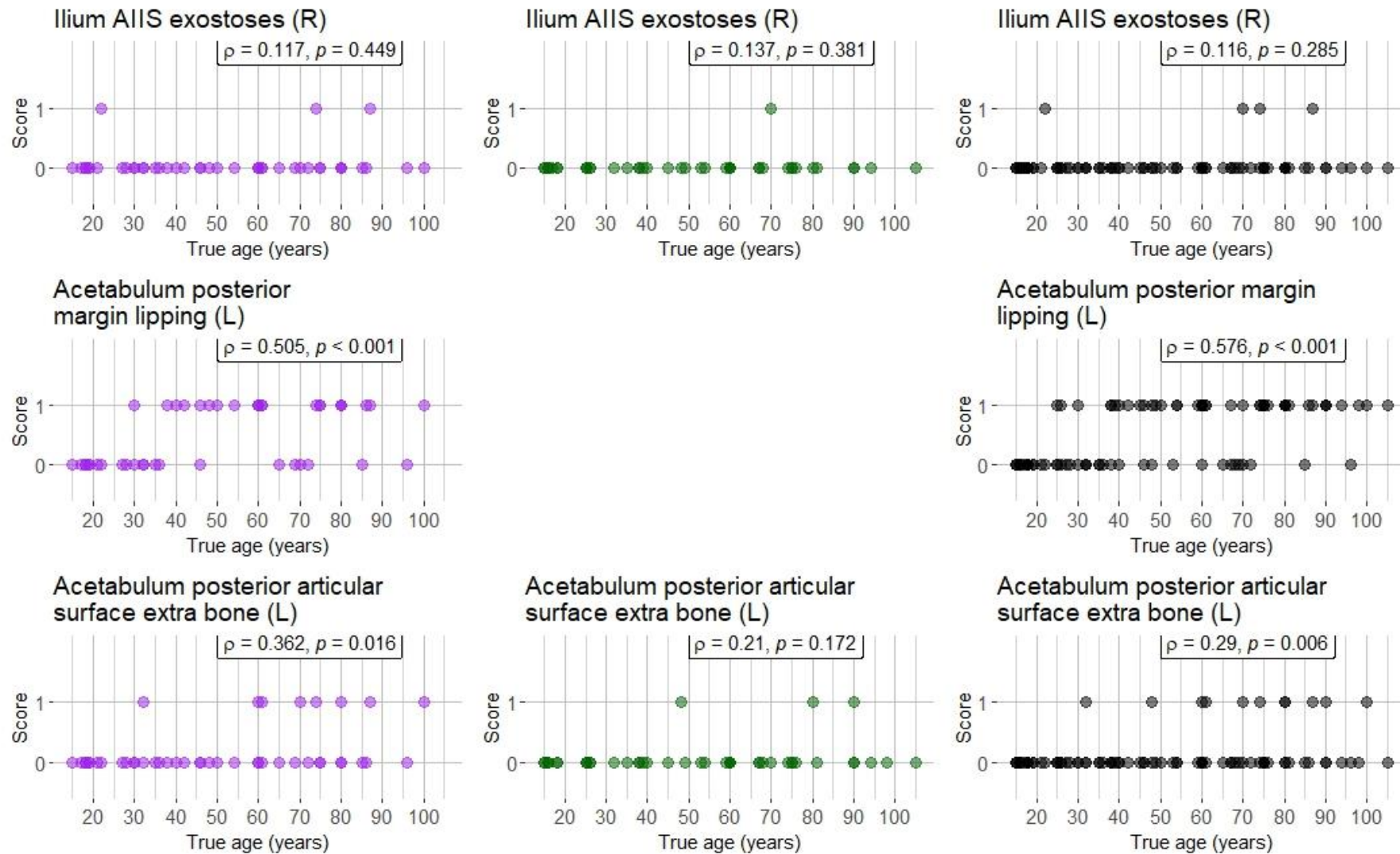


Fig. C.23 Scatterplots for trait score vs age for right ilium anterior inferior iliac exostoses, left acetabulum posterior margin lipping, and posterior articular surface extra bone for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

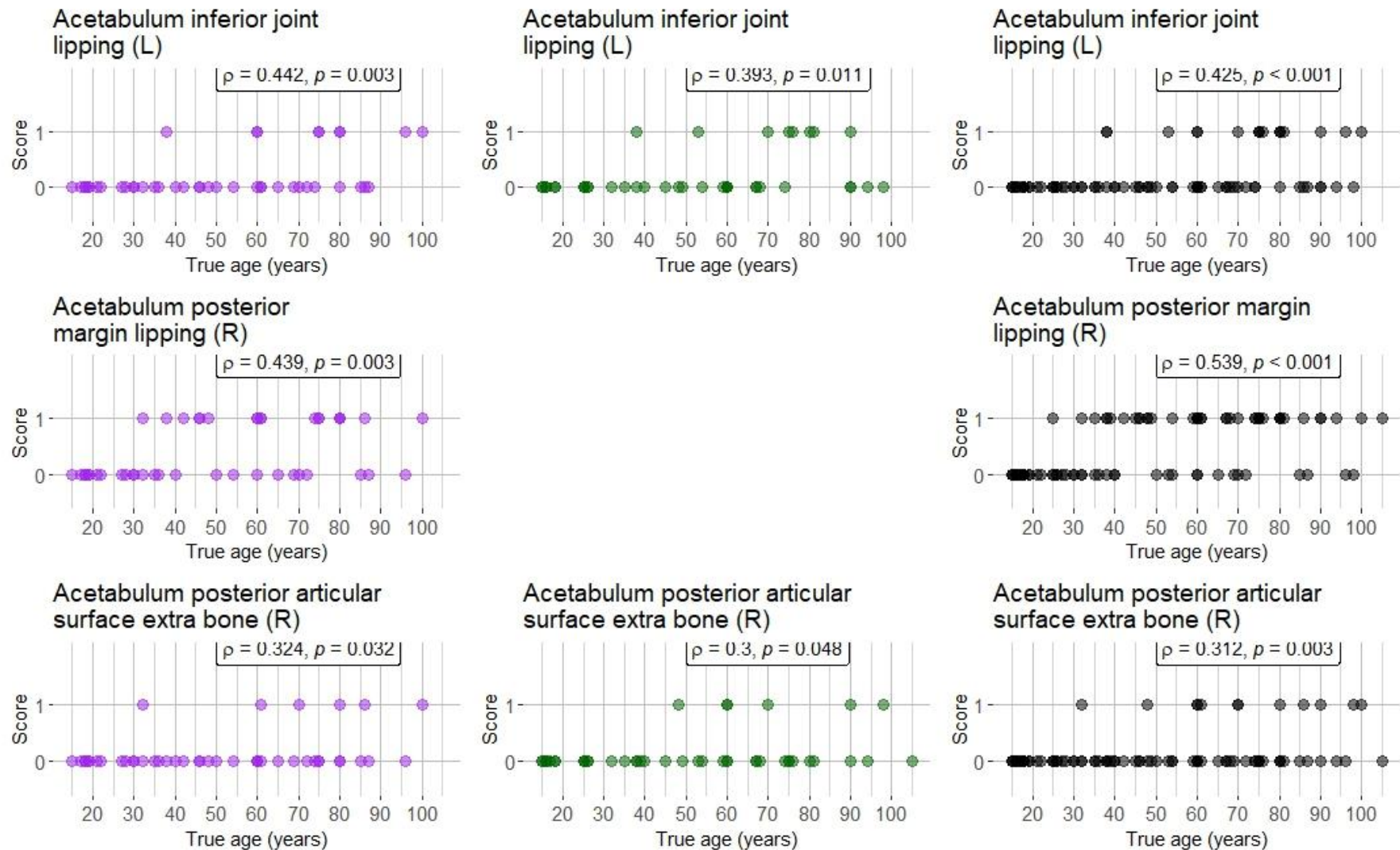


Fig. C.24 Scatterplots for trait score vs age for left acetabulum inferior joint lipping, right acetabulum posterior margin lipping, and posterior articular surface extra bone for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

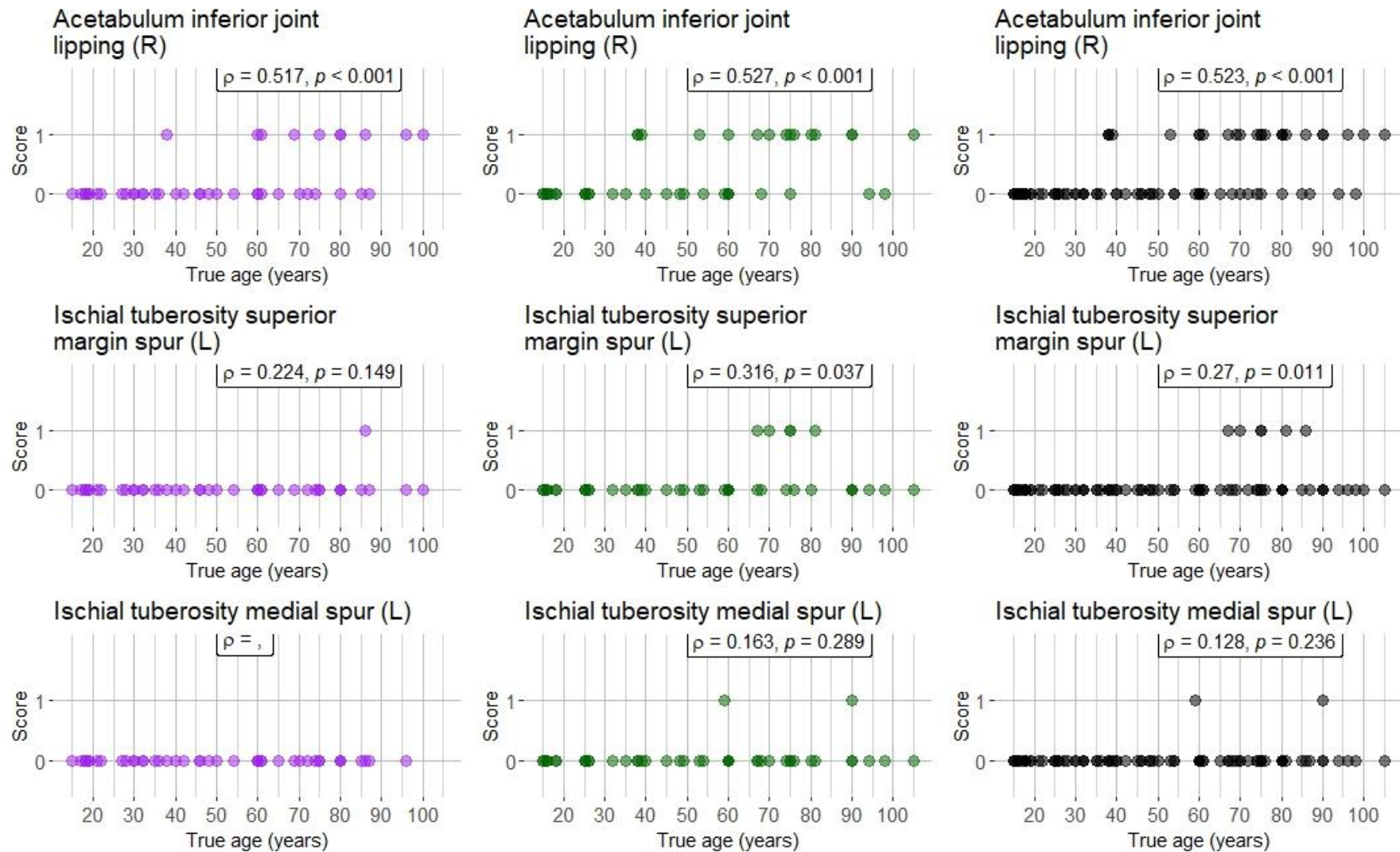


Fig. C.25 Scatterplots for trait score vs age for right acetabulum inferior joint lipping, left ischial tuberosity superior margin spur, and medial spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

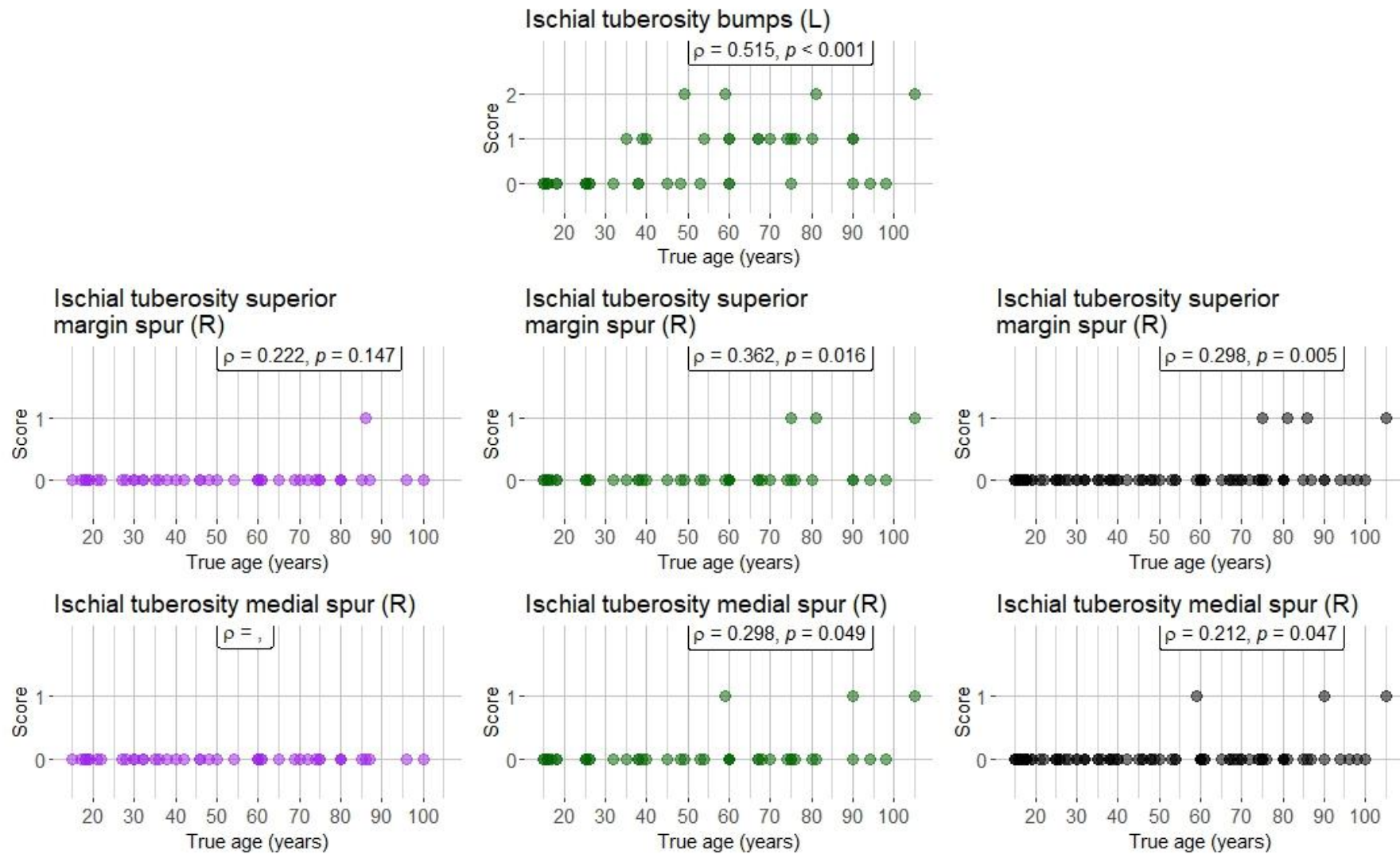


Fig. C.26 Scatterplots for trait score vs age for left ischial tuberosity bumps, right ischial tuberosity superior margin spur, and medial spur for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

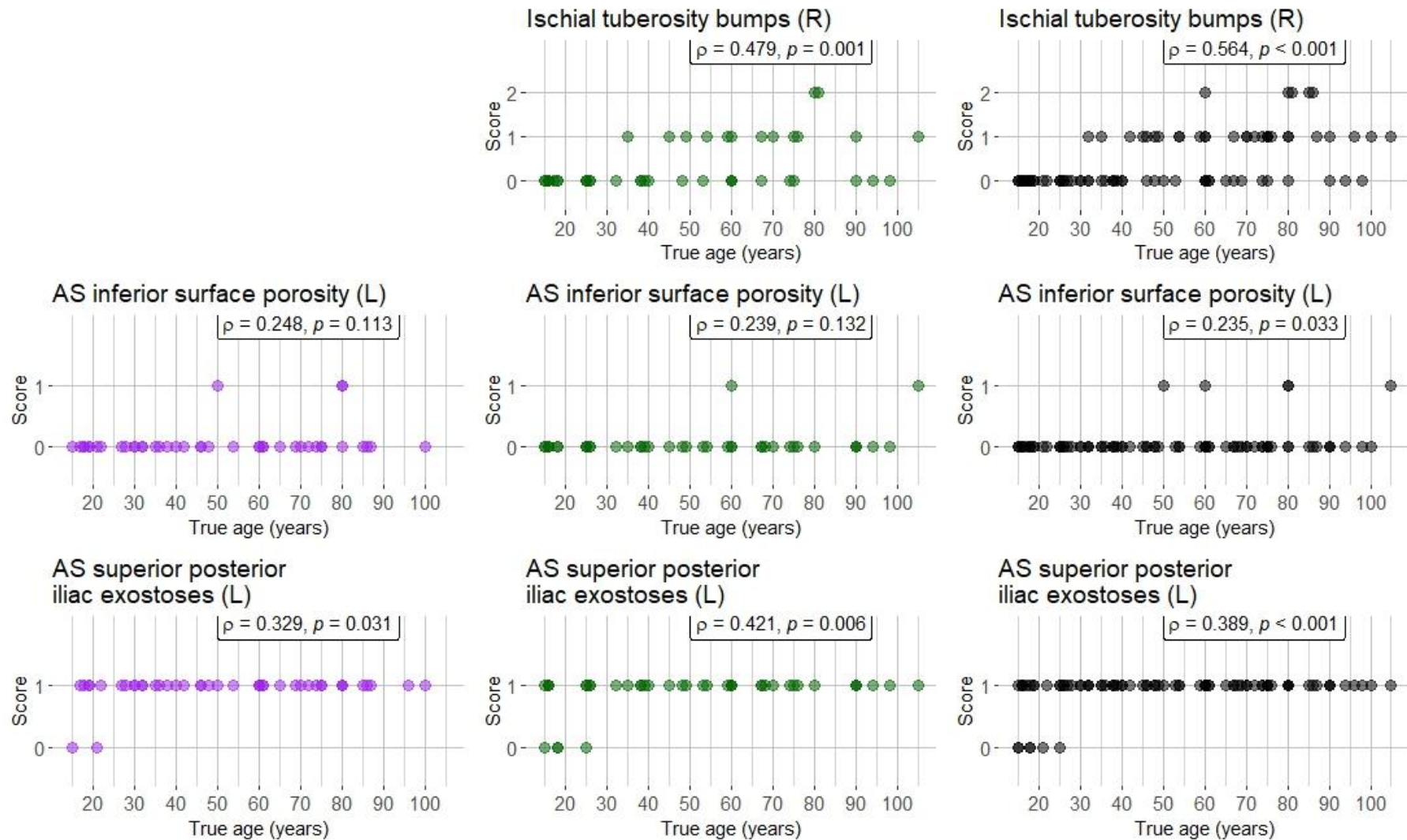


Fig. C.27 Scatterplots for trait score vs age for right ischial tuberosity bumps, left auricular surface inferior surface porosity, and superior posterior exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

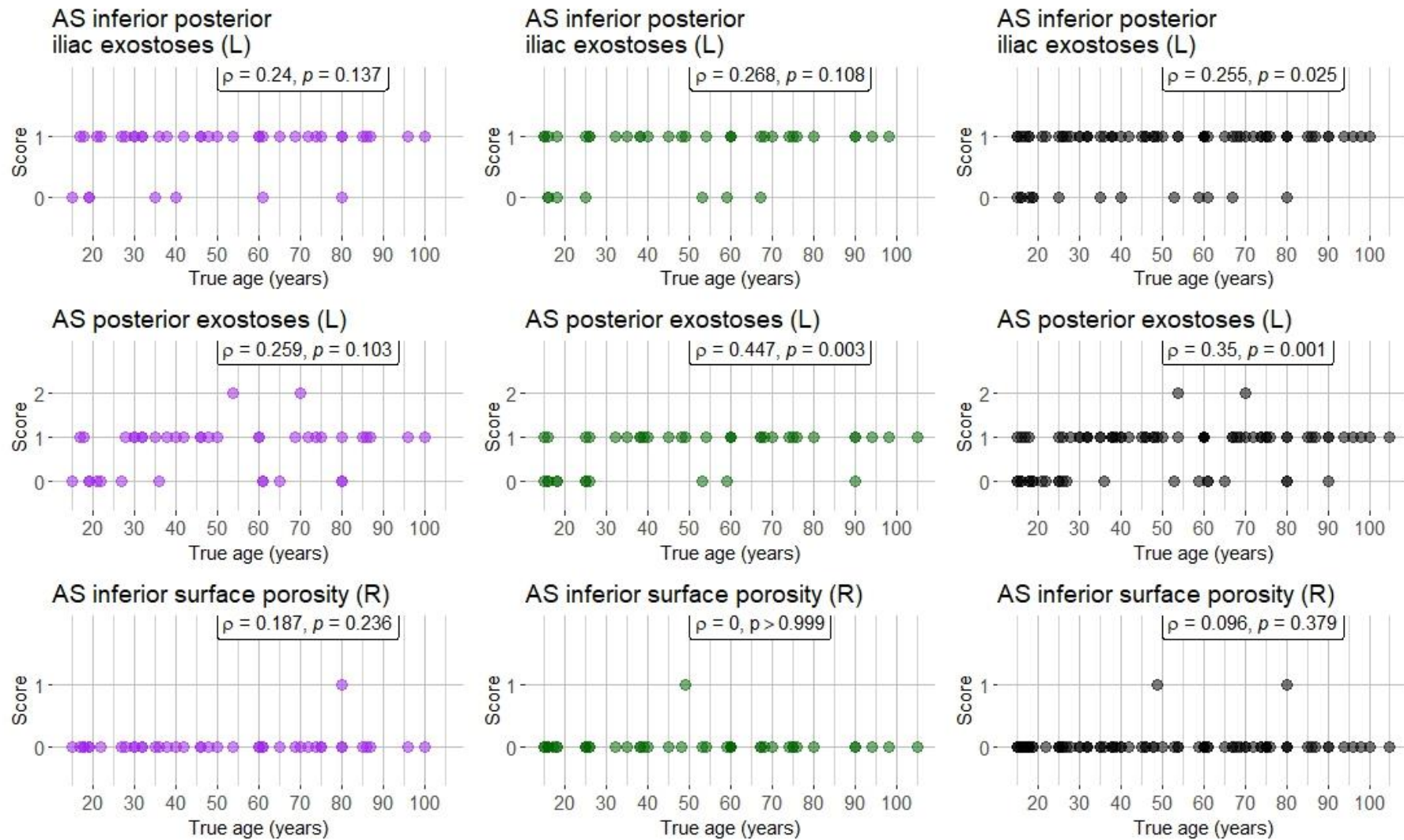


Fig. C.28 Scatterplots for trait score vs age for left auricular surface inferior posterior exostoses, posterior exostoses, and right auricular surface inferior surface porosity for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

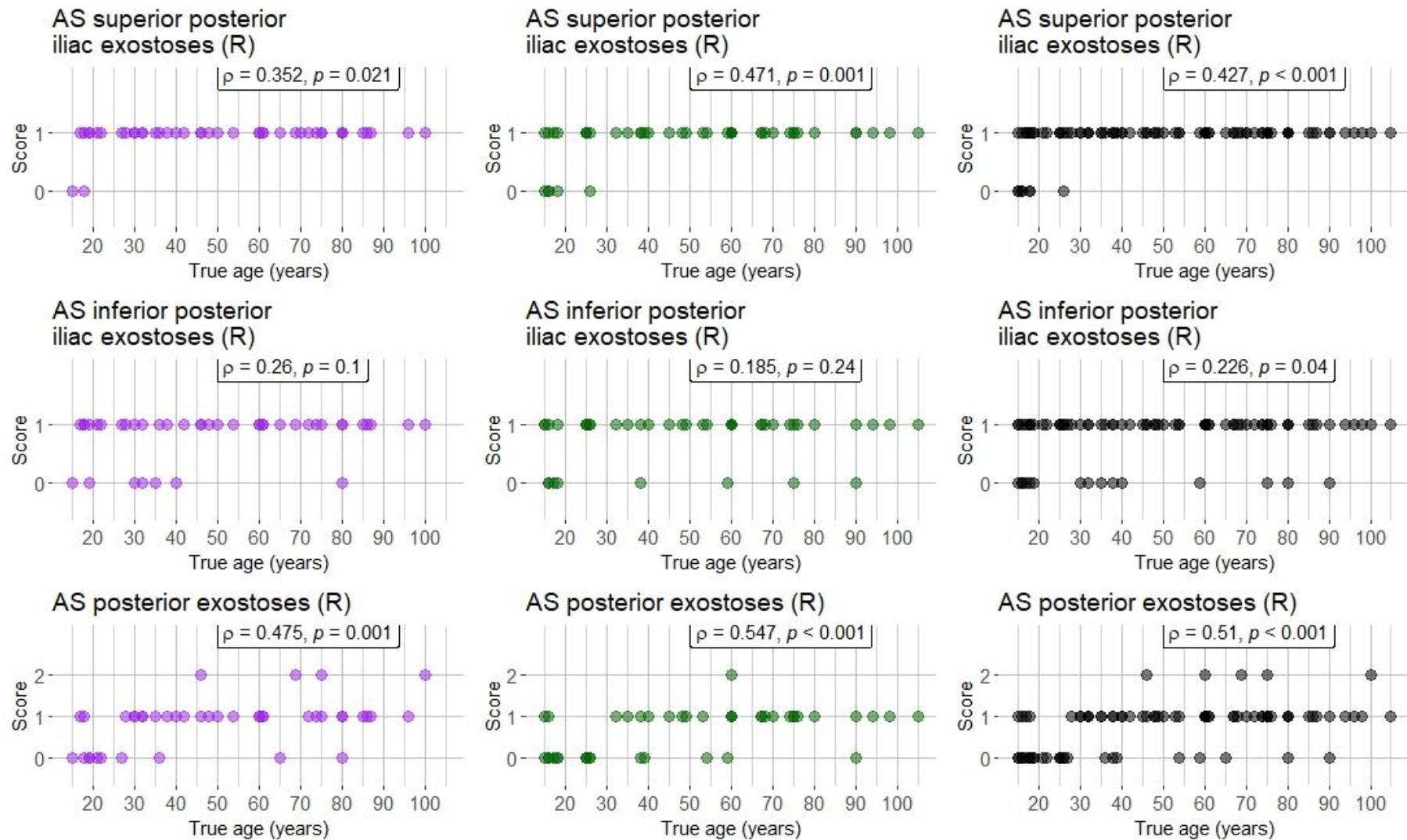


Fig. C.29 Scatterplots for trait score vs age for right auricular surface superior posterior exostoses, inferior posterior exostoses, and posterior exostoses for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

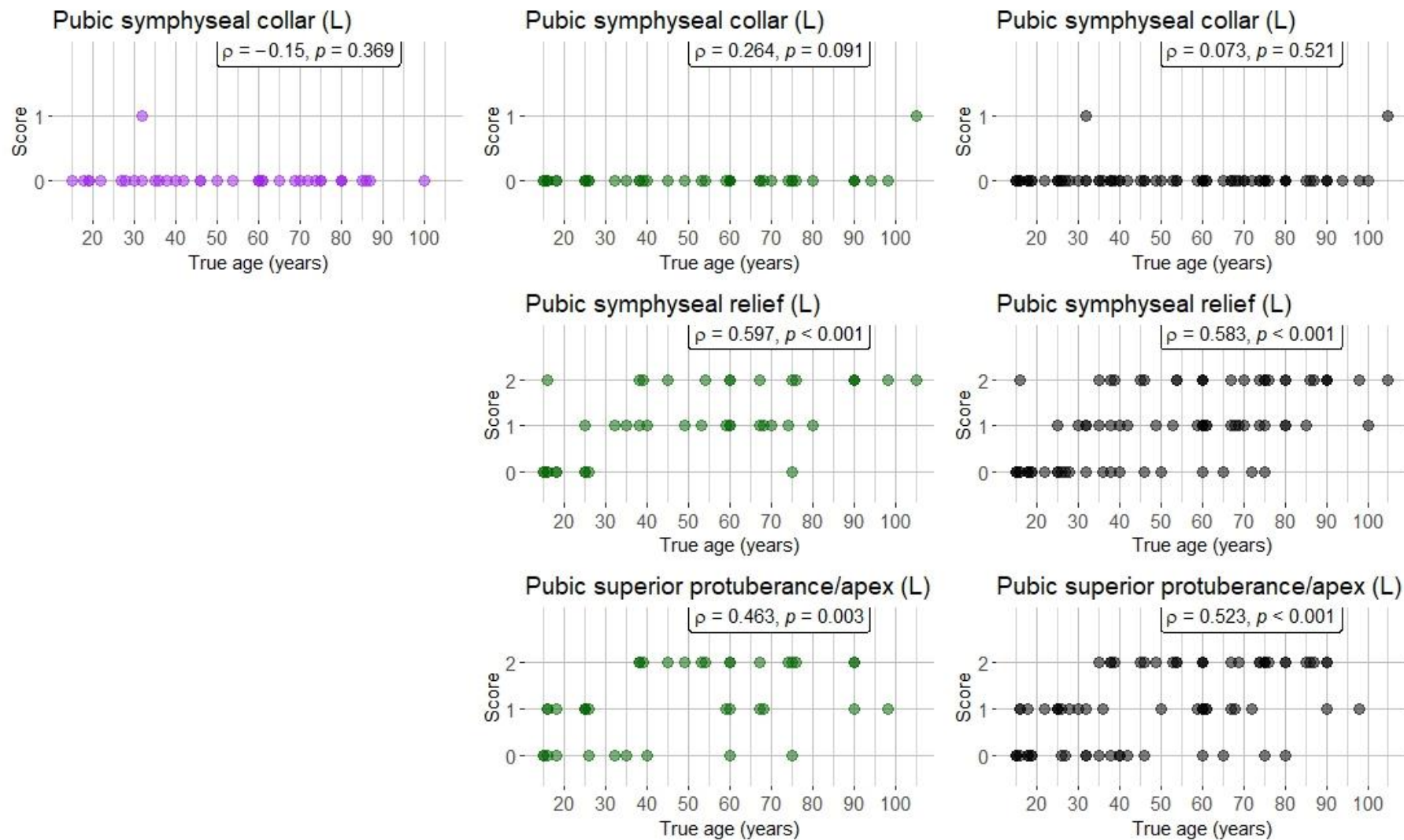


Fig. C.30 Scatterplots for trait score vs age for left pubic symphyseal collar, relief, and superior protuberance for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

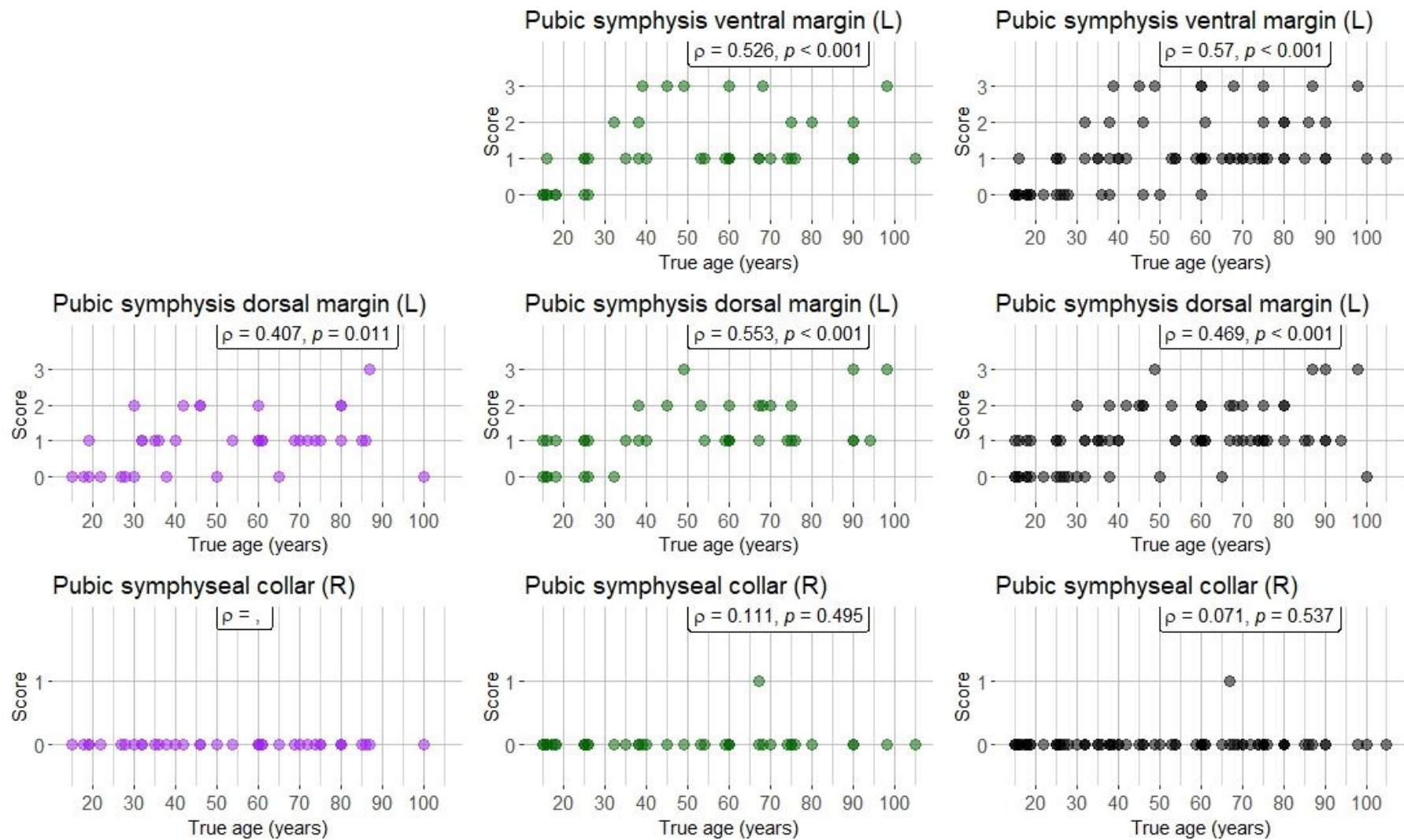


Fig. C.31 Scatterplots for trait score vs age, for left pubic symphysis ventral margin, dorsal margin, and right pubic symphysis collar for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

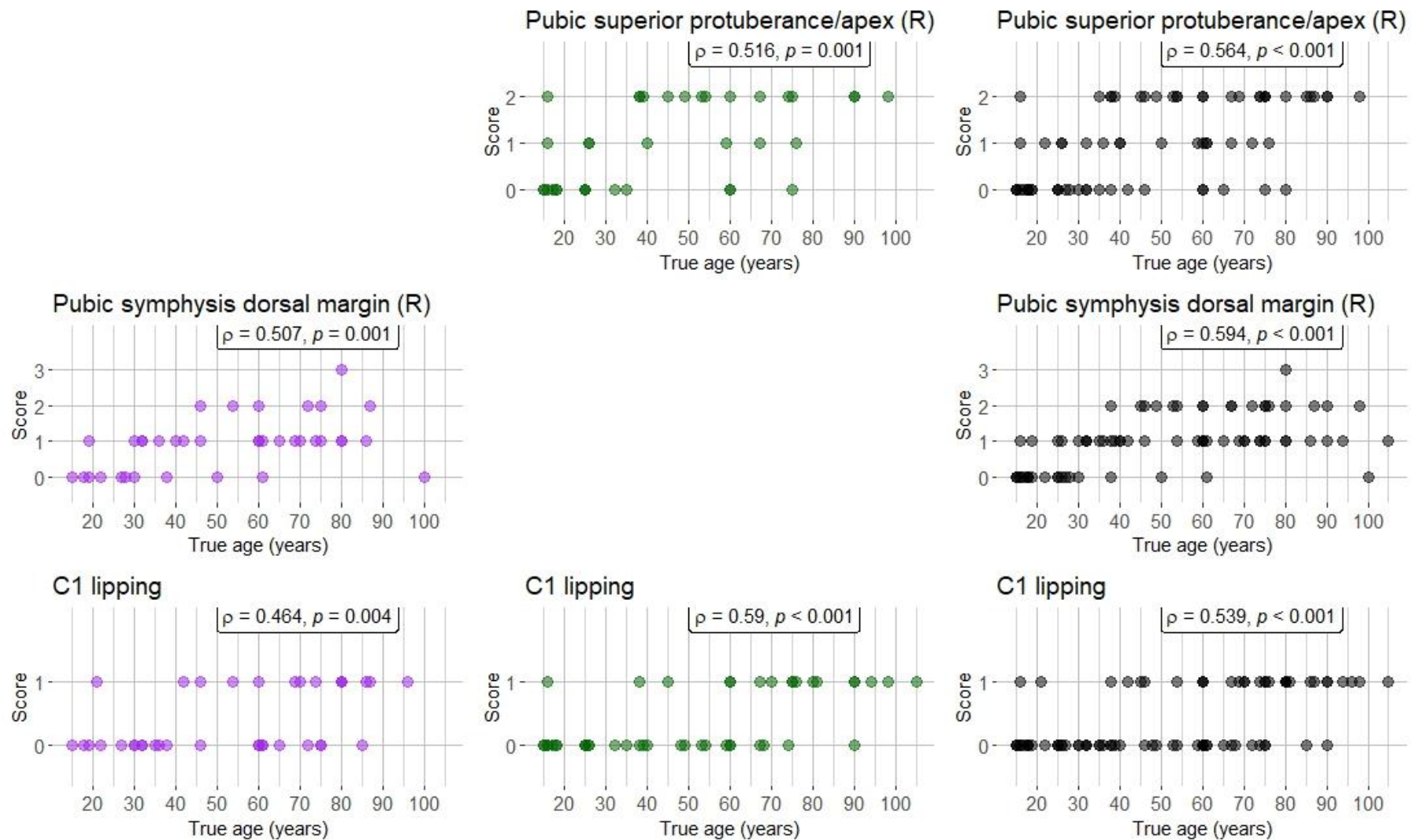


Fig. C.32 Scatterplots for trait score vs age for right pubic superior protuberance, dorsal margin, and C1 lipping for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

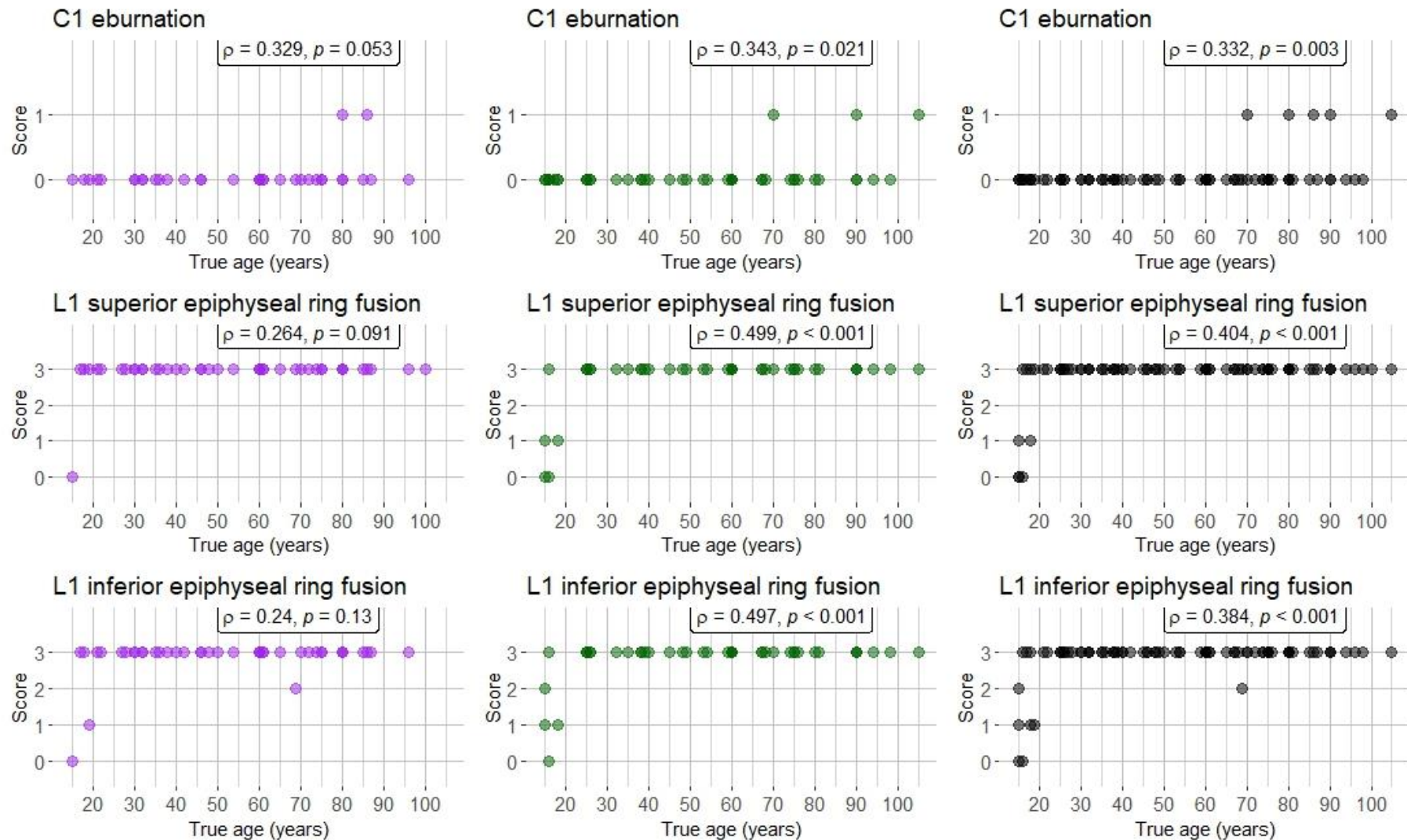


Fig. C.33 Scatterplots for trait score vs age for C1 eburation, L1 superior epiphyseal ring fusion, and inferior epiphyseal ring fusion for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

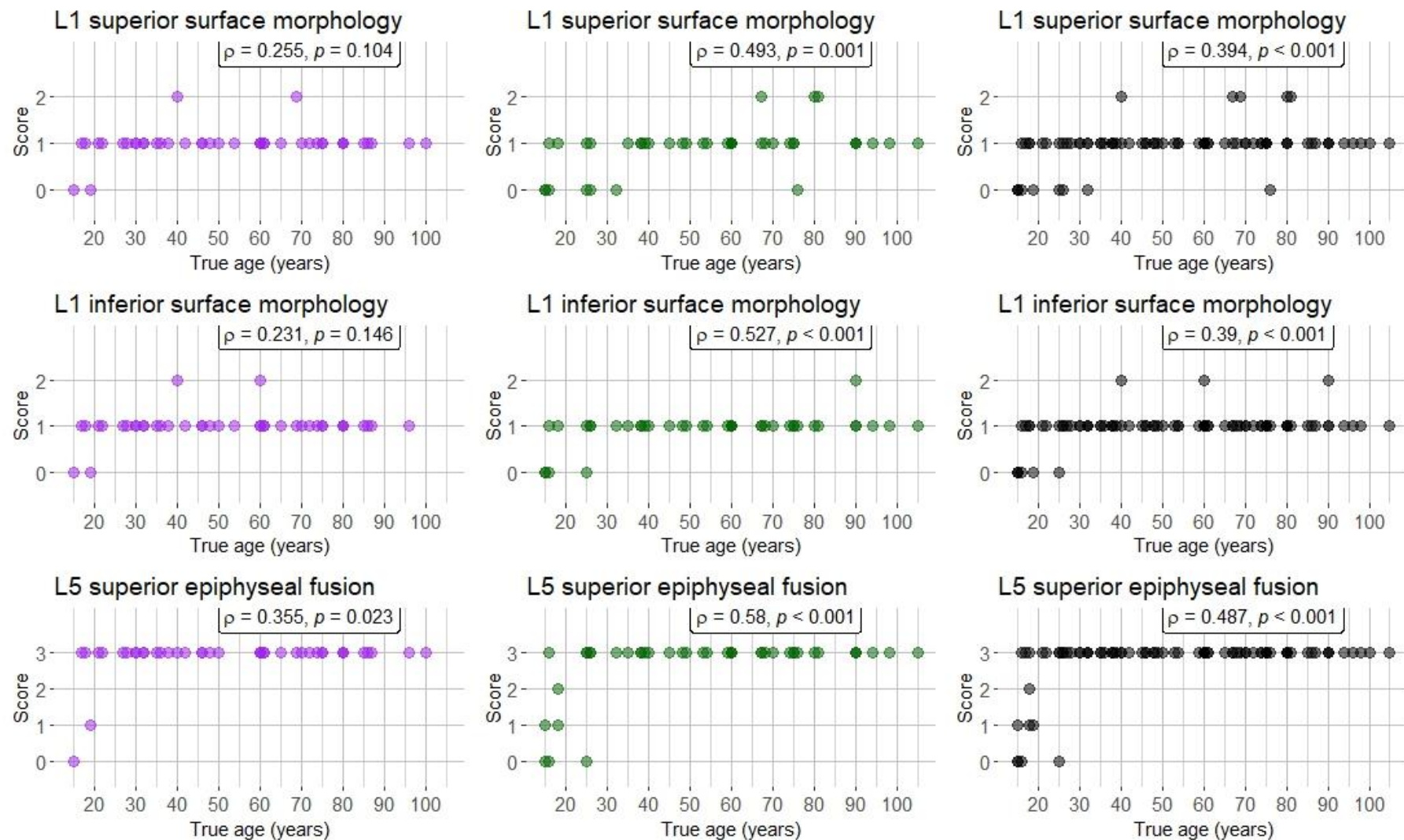


Fig. C.34 Scatterplots for trait score vs age for L1 superior surface morphology, inferior surface morphology, and L5 superior epiphyseal ring fusion for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

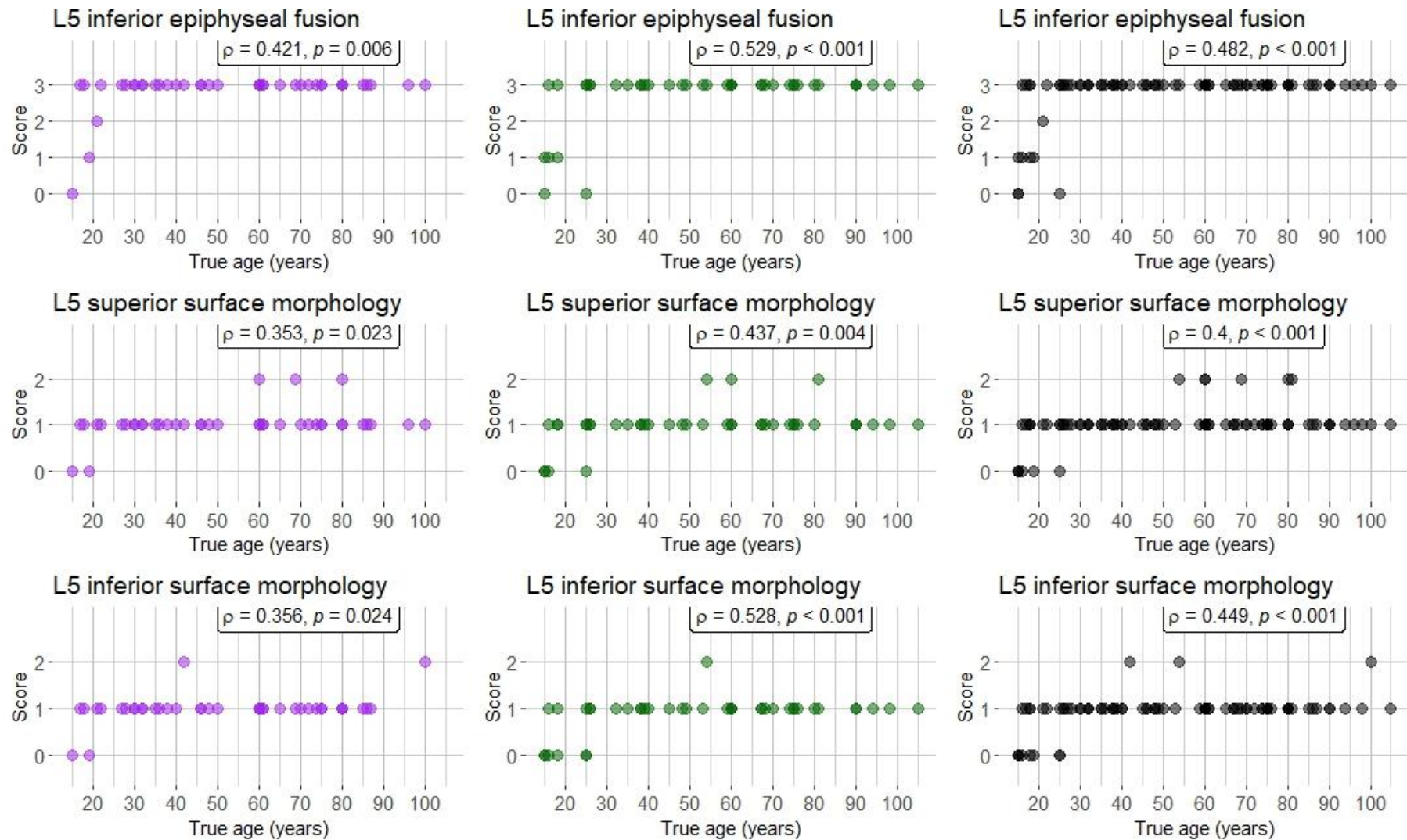


Fig. C.35 Scatterplots for trait score vs age for L5 inferior epiphyseal ring fusion, superior surface morphology, and inferior surface morphology for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

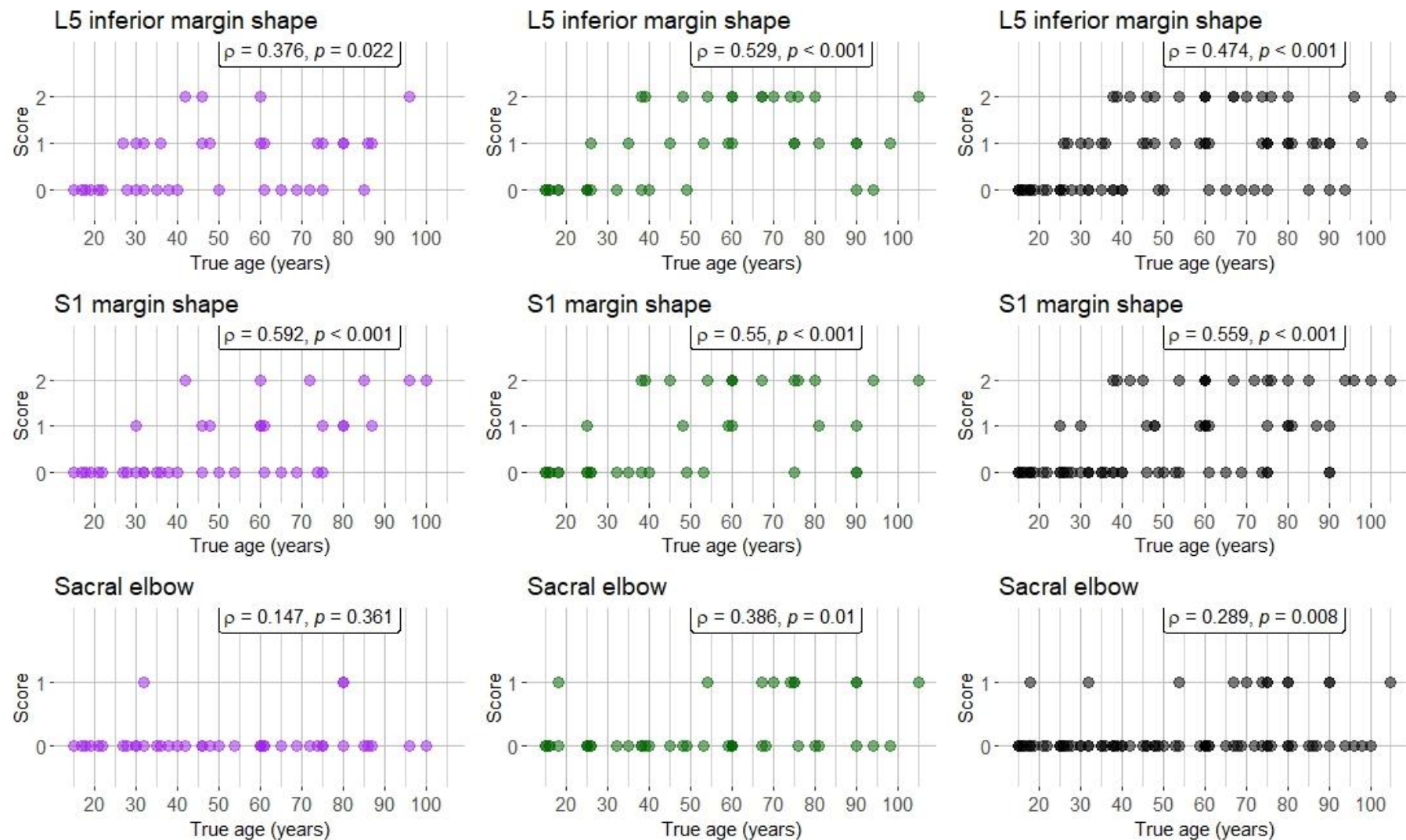


Fig. C.36 Scatterplots for trait score vs age for L5 superior margin shape, inferior margin shape, and S1 margin shape for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

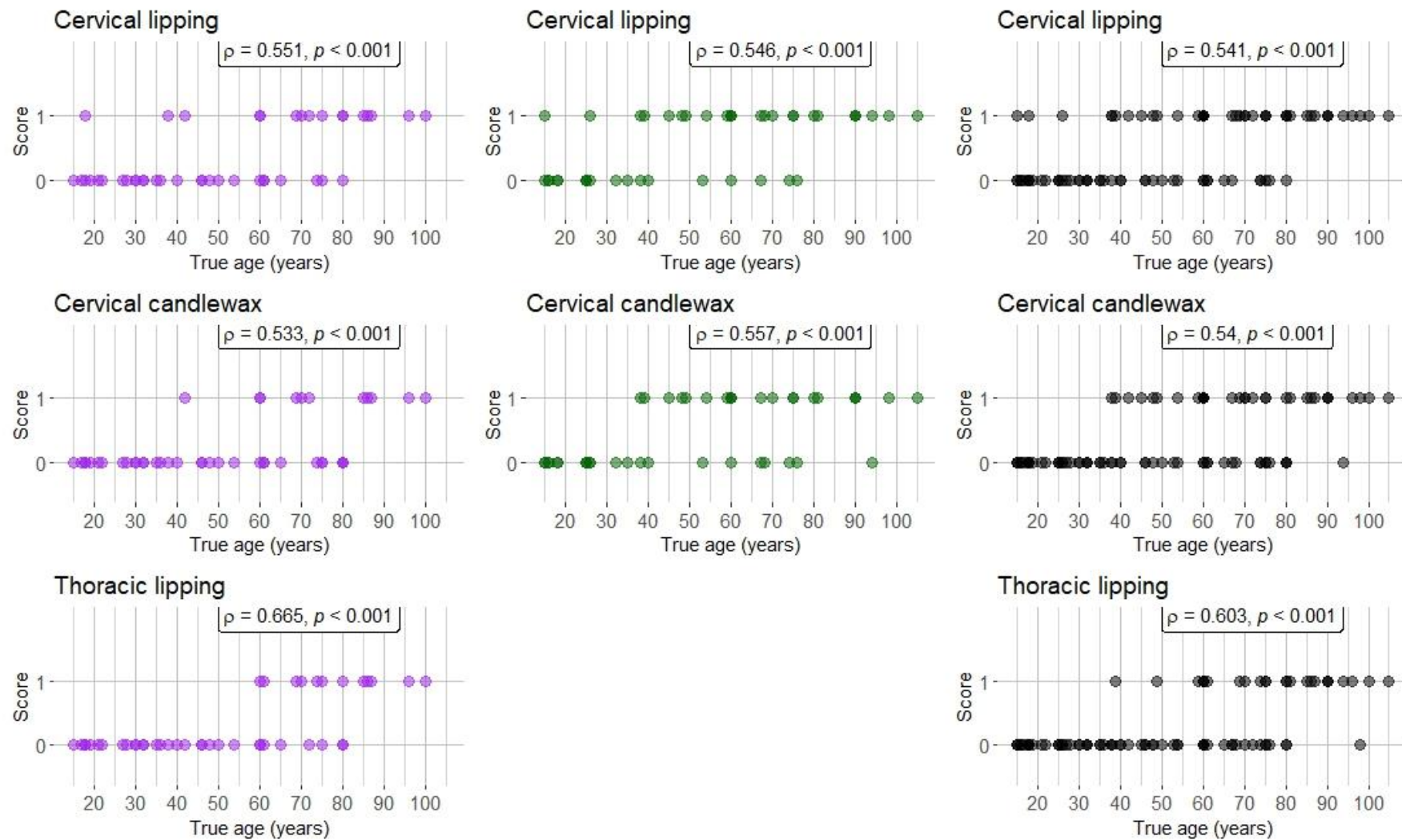


Fig. C.37 Scatterplots of trait score vs age cervical lipping, candlewax, and thoracic lipping for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

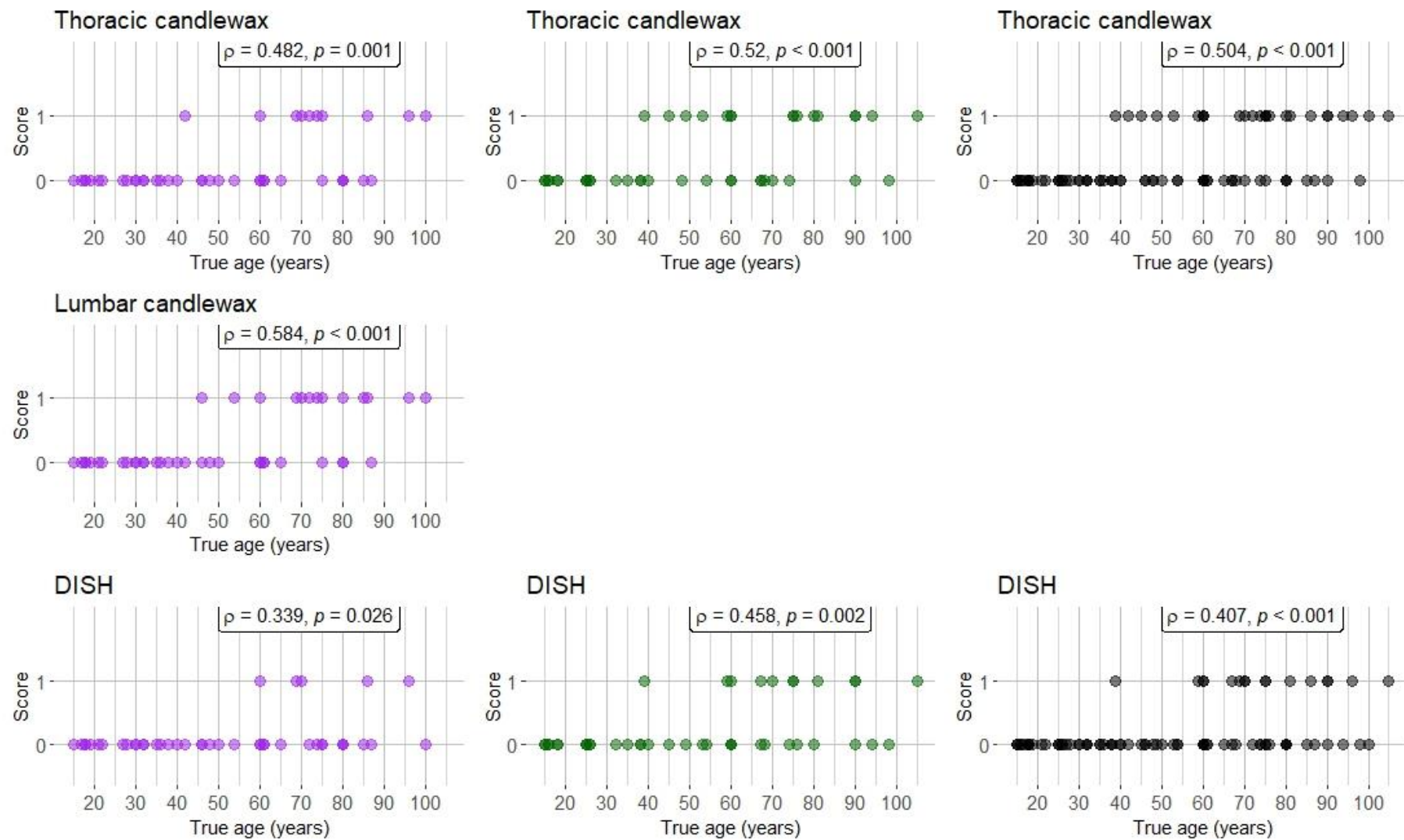


Fig. C.38 Scatterplots for trait score vs age for thoracic candlewax, lumbar candlewax, and DISH for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

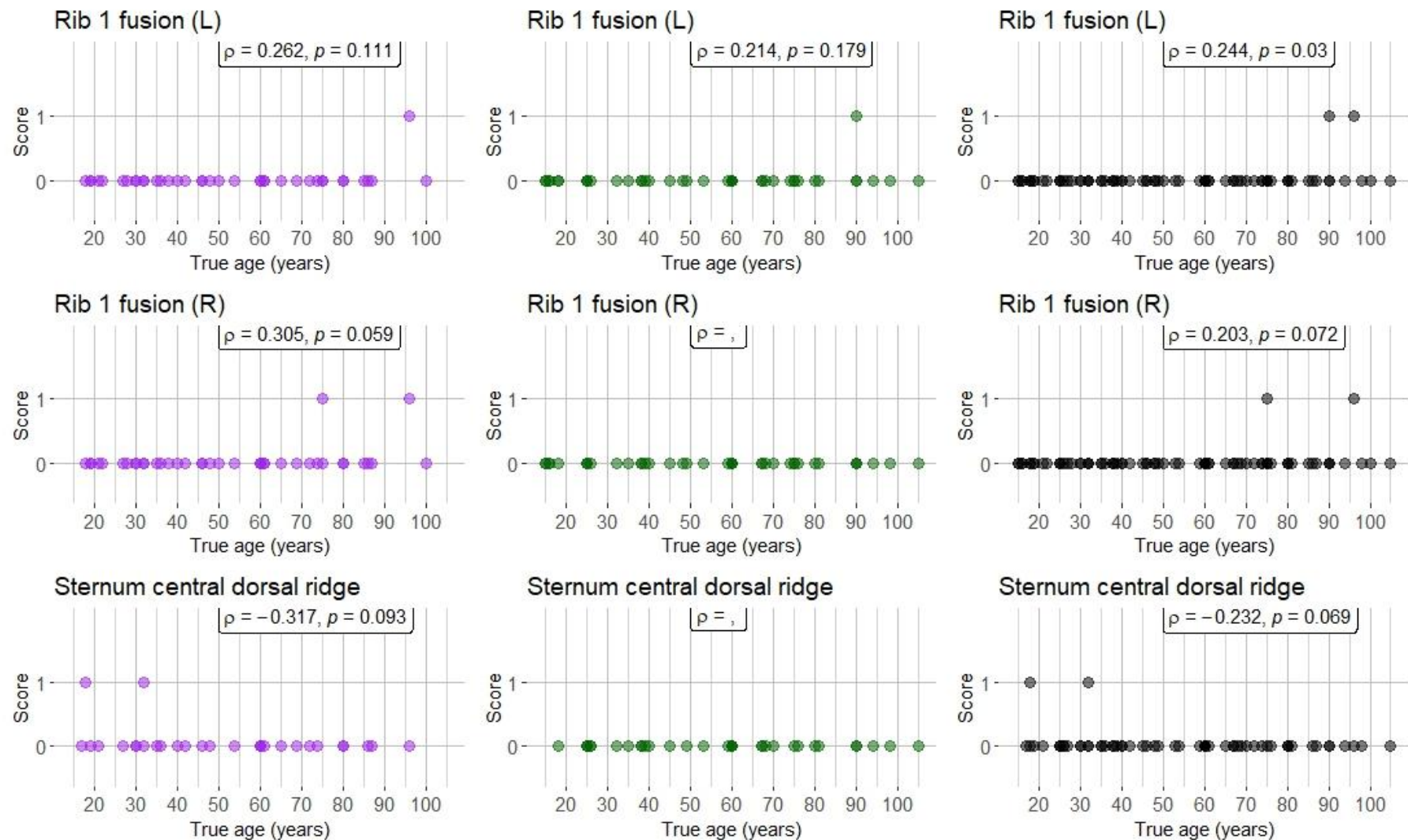


Fig. C.39 Scatterplots for trait score vs age for left rib 1 fusion, right rib 1 fusion, and sternum central dorsal ridge full sample for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

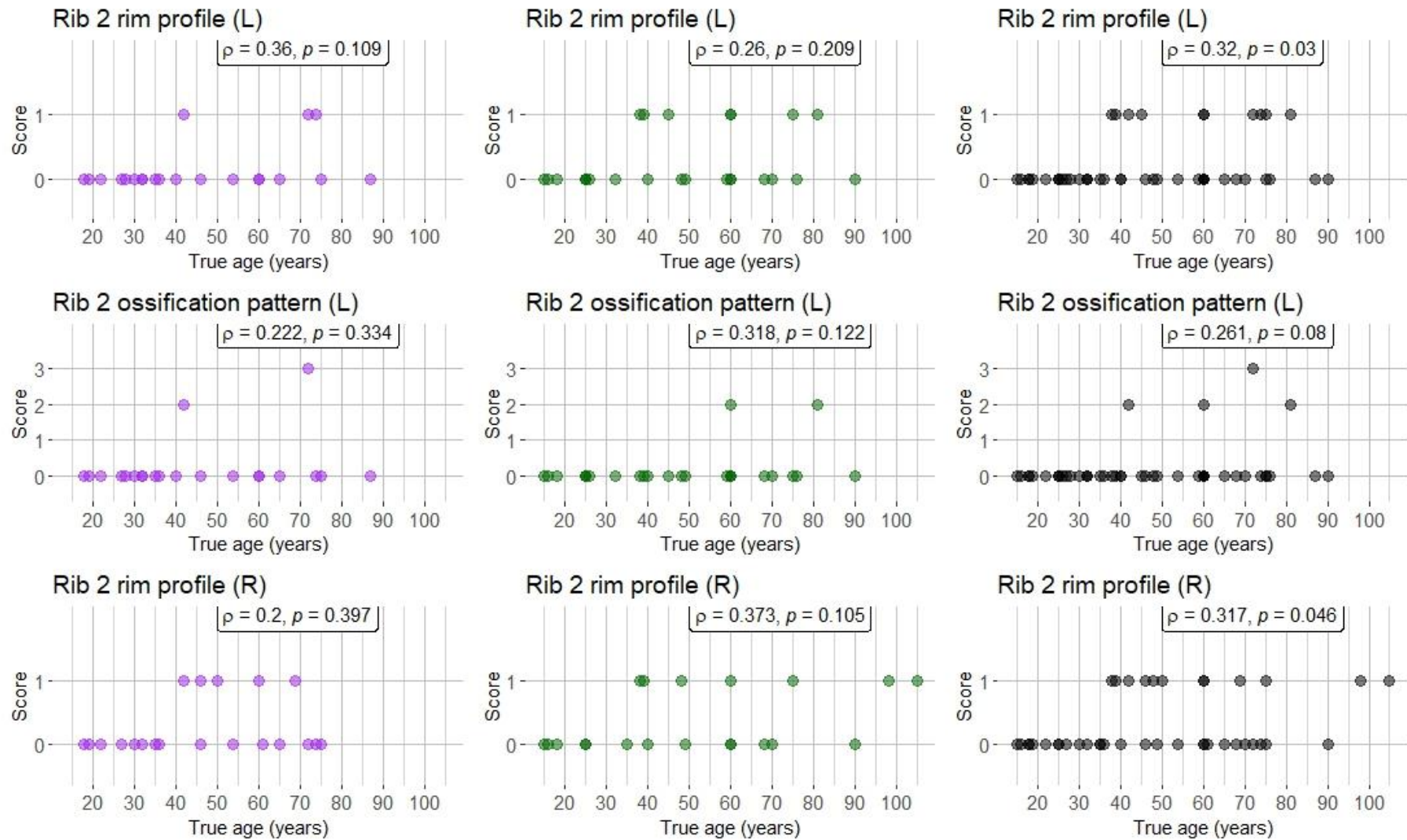


Fig. C.40 Scatterplots for trait score vs age for left rib 2 rim profile, ossification pattern, and right rib 2 rim profile for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

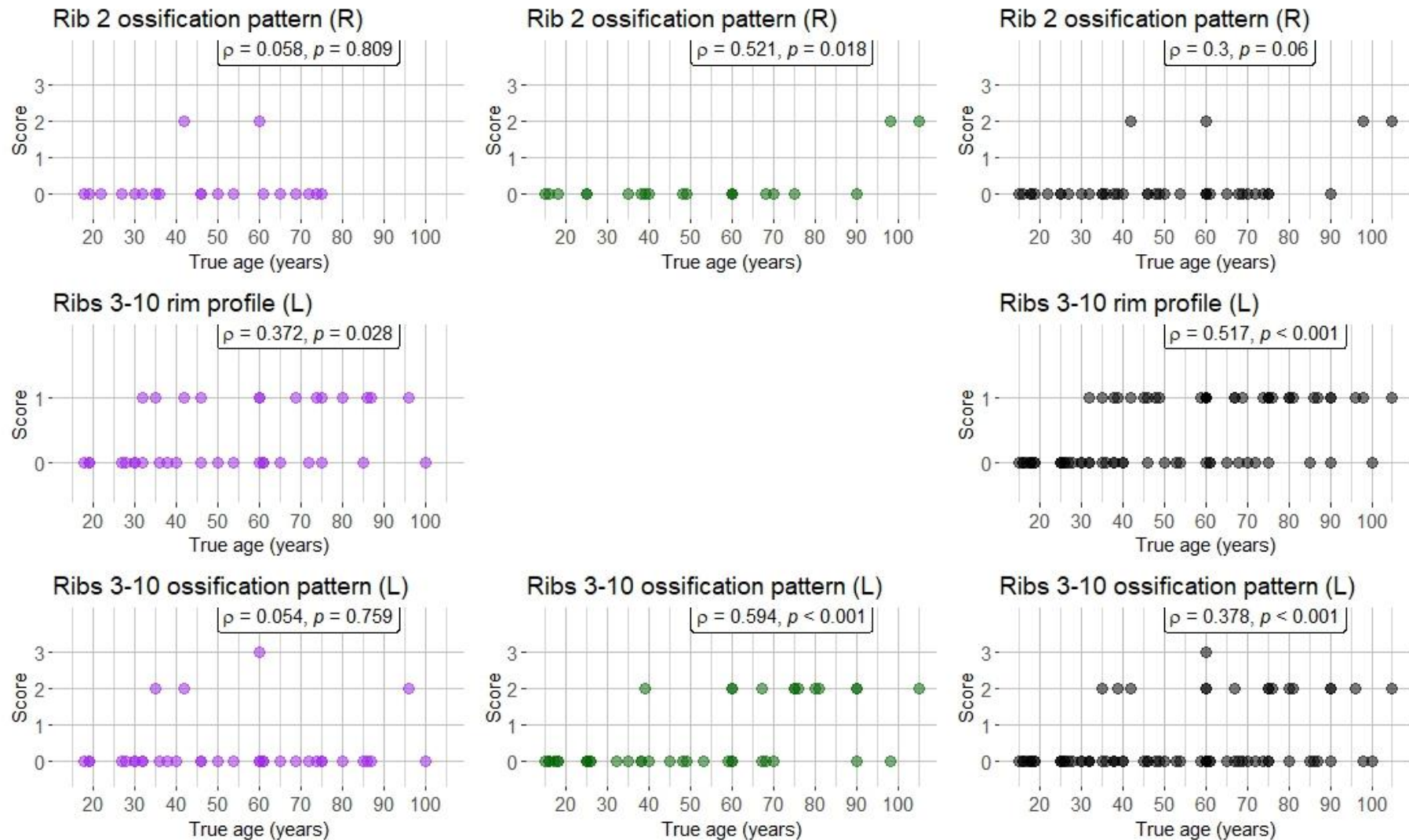


Fig. C.41 Scatterplots for trait score vs age for right rib 2 ossification pattern, left ribs 3-10 rim profile, and ossification pattern for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

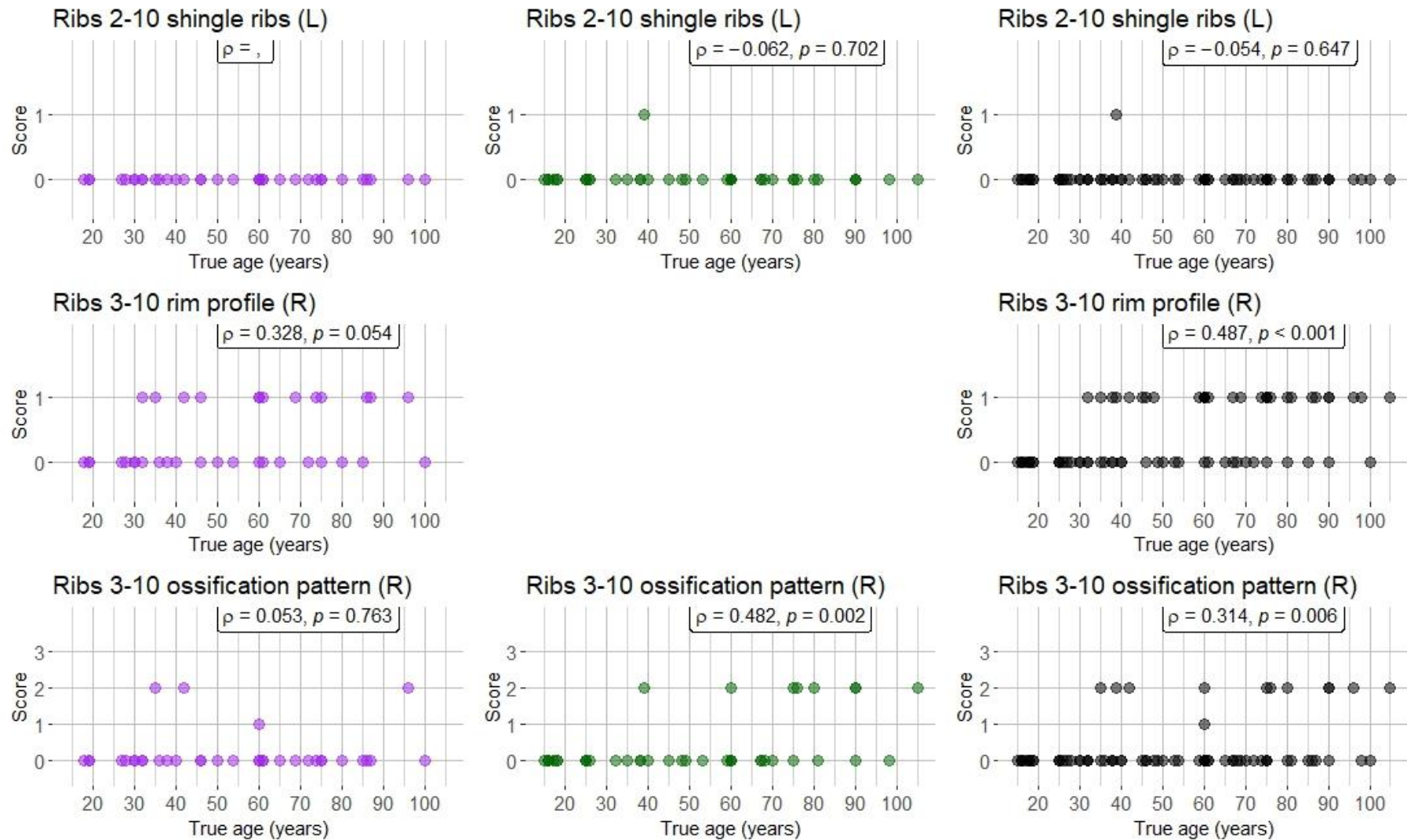


Fig. C.42 Scatterplots for trait score vs age for left ribs 2-10 shingle ribs, right ribs 3-10 rim profile, and ossification pattern for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

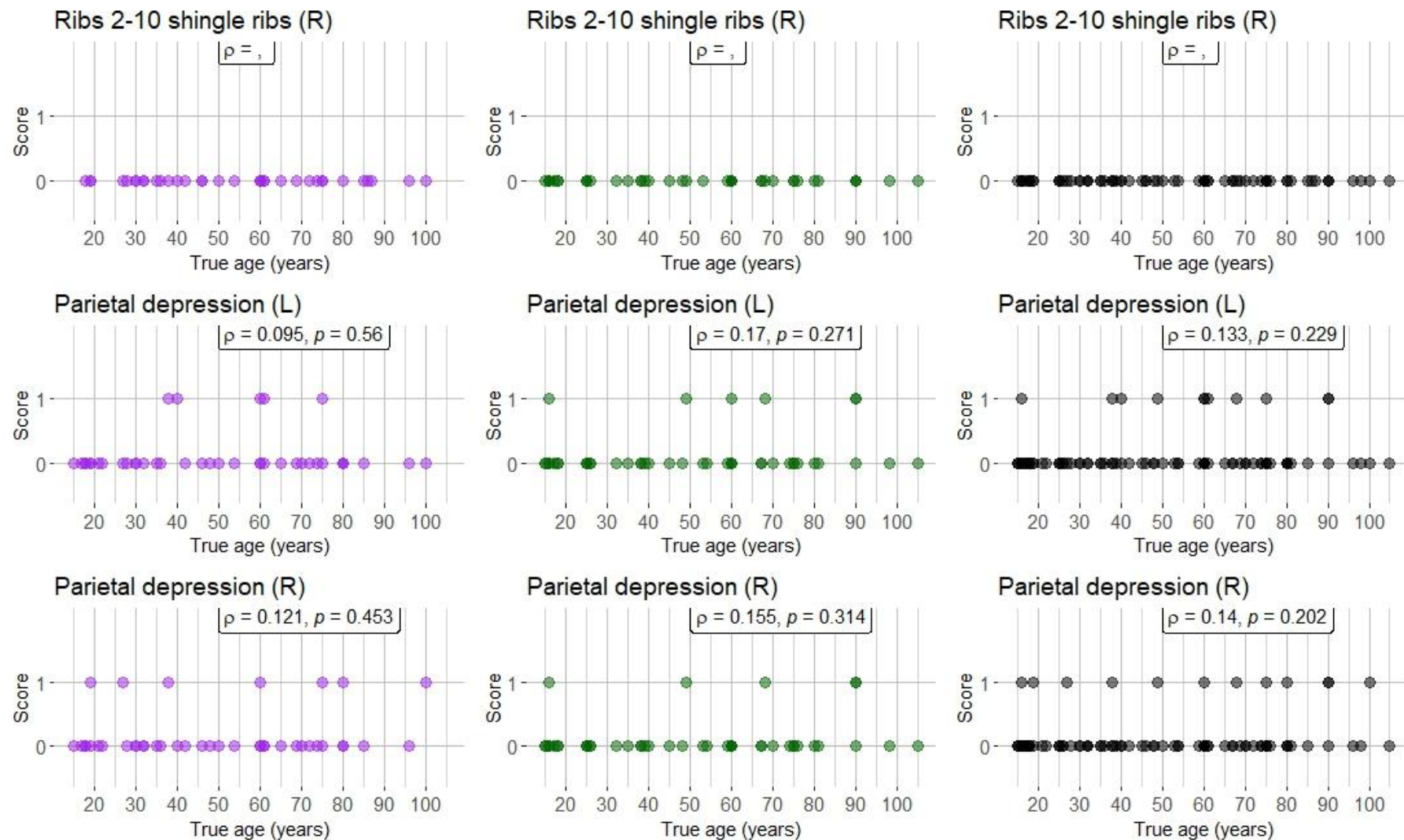


Fig. C.43 Scatterplots for trait score vs age for right ribs 2-10 shingle ribs, left parietal depression, and right parietal depression for the female (n = 44), male (n = 45), and pooled-sex (n = 89) samples from left to right

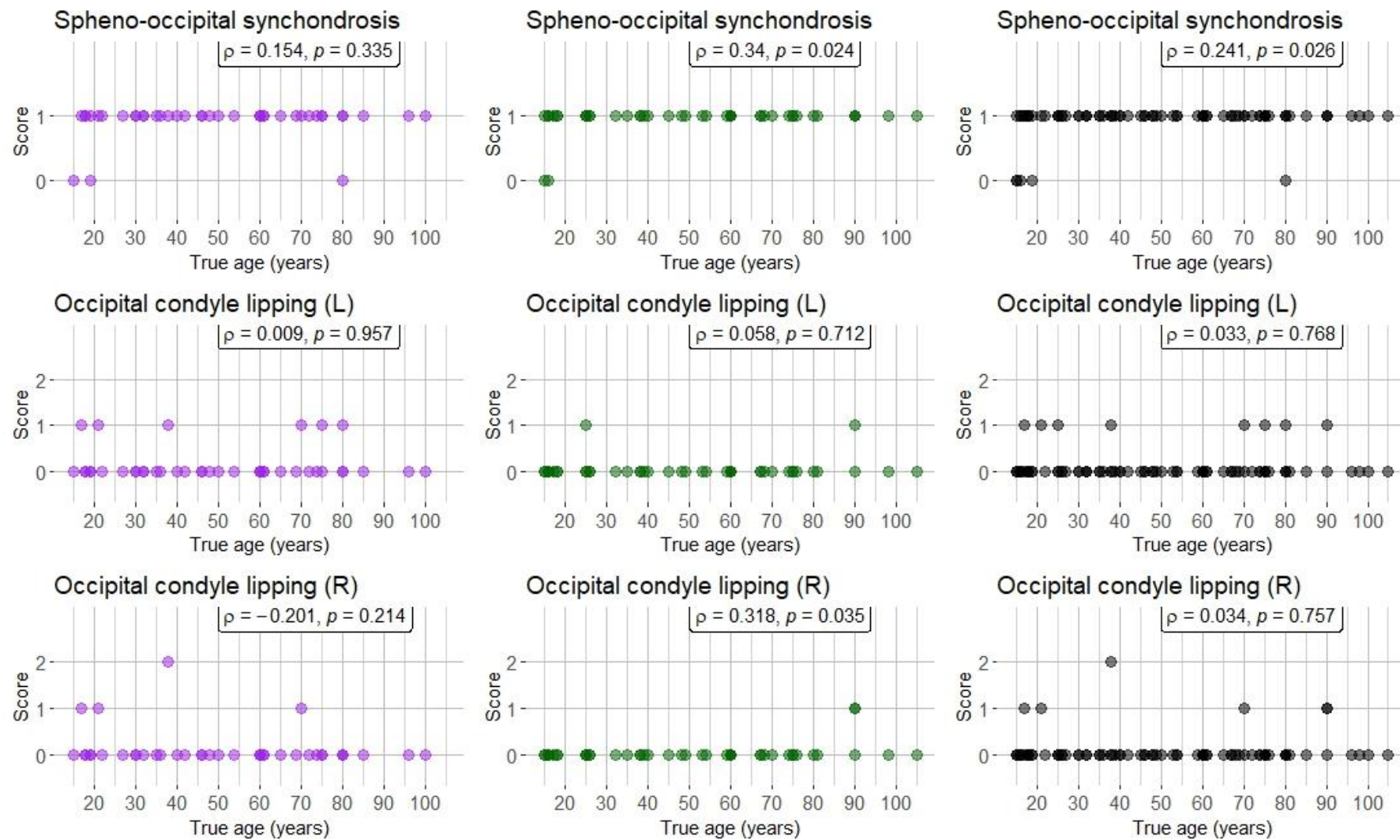


Fig. C.44 Scatterplots for trait score vs age for spheno-occipital synchondrosis, left occipital condyle lipping, and right occipital condyle lipping for the female ($n = 44$), male ($n = 45$), and pooled-sex ($n = 89$) samples from left to right

C.3 Differences between males and females

Table C.6 Fisher's Z transformation ($\alpha = 0.05$) comparing male ($n = 45$) vs female ($n = 44$) age-trait correlations

Trait	Fisher's Z	p-value	H ₀
Femur fovea margin lipping (L)	3.026	0.003	Rejected
Femur fovea margin lipping (R)	1.587	0.113	Retained
Femur head surface extra bone (L)	1.796	0.073	Retained
Femur head surface extra bone (R)	2.509	0.012	Rejected
Femur greater trochanter roughening (L)	-2.292	0.022	Rejected
Femur greater trochanter roughening (R)	-1.929	0.054	Retained
Femur trochanteric fossa exostoses (L)	-0.282	0.778	Retained
Femur trochanteric fossa exostoses (R)	-1.127	0.260	Retained
Femur trochanteric medial exostoses (L)	-2.598	0.009	Rejected
Femur trochanteric medial exostoses (R)	-1.720	0.086	Retained
Tibia relative weight (L)	0.064	0.949	Retained
Tibia relative weight (R)	-0.482	0.630	Retained
Tibia actual weight (L)	2.358	0.018	Rejected
Tibia actual weight (R)	2.507	0.012	Rejected
Tibia length (L)	2.277	0.023	Rejected
Tibia length (R)	2.571	0.010	Rejected
Fibula wings (L)	1.114	0.265	Retained
Fibula wings (R)	1.212	0.262	Retained
Calcaneus relative weight (L)	-0.383	0.702	Retained
Calcaneus relative weight (R)	-0.321	0.748	Retained
Calcaneus actual weight (L)	1.497	0.134	Retained
Calcaneus actual weight (R)	1.191	0.234	Retained
Calcaneus length (L)	1.086	0.278	Retained
Calcaneus length (R)	0.785	0.433	Retained
Clavicle medial epiphysis fusion (L)	-0.346	0.729	Retained
Clavicle medial epiphysis fusion (R)	0.474	0.636	Retained
Clavicle medial epiphysis gravel (L)	0.448	0.655	Retained
Clavicle medial epiphysis gravel (R)	2.254	0.024	Rejected
Clavicle medial macroporosity (L)	NA	NA	NA
Clavicle medial macroporosity (R)	1.416	0.159	Retained
Clavicle lateral macroporosity (L)	0.692	0.489	Retained
Clavicle lateral macroporosity (R)	-0.199	0.842	Retained
Scapula glenoid fossa lipping (L)	0.653	0.514	Retained
Scapula glenoid fossa lipping (R)	-0.283	0.777	Retained
Humerus relative weight (L)	-1.829	0.067	Retained
Humerus relative weight (R)	-1.851	0.064	Retained
Humerus actual weight (L)	3.314	0.001	Rejected
Humerus actual weight (R)	4.012	<0.001	Rejected
Humerus length (L)	2.413	0.046	Rejected
Humerus length (R)	2.375	0.018	Rejected
Humerus lesser tubercle anterior surface bumps (L)	-0.952	0.341	Retained
Humerus lesser tubercle anterior surface bumps (R)	-1.009	0.313	Retained

Table C.6 continued

Trait	Fisher's Z	p-value	H₀
Humerus lesser tubercle margin shape (L)	-0.812	0.417	Retained
Humerus lesser tubercle margin shape (R)	-1.052	0.293	Retained
Humerus greater tubercle pits (L)	0.559	0.576	Retained
Humerus greater tubercle pits (R)	0.144	0.886	Retained
Humerus medial epicondyle (L)	-1.013	0.311	Retained
Humerus medial epicondyle (R)	-0.652	0.514	Retained
Humerus lateral epicondyle (L)	-0.405	0.686	Retained
Humerus lateral epicondyle (R)	-0.109	0.913	Retained
Radius tuberosity medial crest (L)	1.164	0.245	Retained
Radius tuberosity medial crest (R)	0.774	0.439	Retained
Ulna olecranon spur (L)	1.216	0.224	Retained
Ulna olecranon spur (R)	-0.143	0.886	Retained
Trapezium lipping (L)	-0.595	0.552	Retained
Trapezium lipping (R)	1.270	0.204	Retained
Innominate relative weight (L)	-1.654	0.098	Retained
Innominate relative weight (R)	-1.199	0.231	Retained
Innominate actual weight (L)	3.798	<0.001	Rejected
Innominate actual weight (R)	3.383	0.001	Rejected
Innominate length (L)	2.419	0.016	Rejected
Innominate length (R)	2.587	0.010	Rejected
Sacroiliac joint fusion (L)	0.291	0.771	Retained
Sacroiliac joint fusion (R)	0.724	0.469	Retained
Iliac crest fusion (L)	-0.219	0.827	Retained
Iliac crest fusion (R)	-0.478	0.633	Retained
Iliac crest tuberculum ossification (L)	0.314	0.753	Retained
Iliac crest tuberculum ossification (R)	0.359	0.720	Retained
Ilium anterior inferior Iliac spine exostoses (L)	-1.104	0.270	Retained
Ilium anterior inferior iliac spine exostoses (R)	0.091	0.927	Retained
Acetabulum posterior margin lipping (L)	1.022	0.307	Retained
Acetabulum posterior margin lipping (R)	1.462	0.144	Retained
Acetabulum posterior articular surface extra bone (L)	-0.759	0.448	Retained
Acetabulum posterior articular surface extra bone (R)	-0.118	0.906	Retained
Acetabulum inferior joint lipping (L)	-0.272	0.786	Retained
Acetabulum inferior joint lipping (R)	0.065	0.949	Retained
Ischial tuberosity superior margin spur (L)	0.453	0.650	Retained
Ischial tuberosity superior margin spur (R)	0.700	0.484	Retained
Ischial tuberosity medial spur (L)	NA	NA	NA
Ischial tuberosity medial spur (R)	NA	NA	NA
Ischial tuberosity bumps (L)	-1.463	0.143	Retained
Ischial tuberosity bumps (R)	-1.176	0.240	Retained
Auricular surface inferior surface porosity (L)	-0.041	0.967	Retained
Auricular surface inferior surface porosity (R)	-0.862	0.389	Retained
Auricular surface superior posterior iliac exostoses (L)	0.484	0.628	Retained
Auricular surface superior posterior iliac exostoses (R)	0.656	0.512	Retained
Auricular surface inferior posterior iliac exostoses (L)	0.141	0.888	Retained
Auricular surface inferior posterior iliac exostoses (R)	-0.360	0.719	Retained

Table C.6 continued

Trait	Fisher's Z	p-value	H₀
Auricular surface posterior exostoses (L)	0.986	0.324	Retained
Auricular surface posterior exostoses (R)	0.442	0.658	Retained
Pubic symphyseal collar (L)	1.922	0.055	Retained
Pubic symphyseal collar (R)	NA	NA	NA
Pubic symphyseal relief (L)	-0.043	0.966	Retained
Pubic symphyseal relief (R)	-0.174	0.862	Retained
Pubic superior protuberance/apex (L)	-0.964	0.335	Retained
Pubic superior protuberance/apex (R)	-0.865	0.387	Retained
Pubic symphysis ventral margin (L)	-1.162	0.245	Retained
Pubic symphysis ventral margin (R)	0.229	0.819	Retained
Pubic symphysis dorsal margin (L)	0.868	0.386	Retained
Pubic symphysis dorsal margin (R)	1.283	0.199	Retained
C1 lipping	0.799	0.424	Retained
C1 eburnation	0.072	0.943	Retained
L1 spinous process flattening	-0.912	0.362	Retained
L1 superior epiphyseal ring fusion	1.262	0.207	Retained
L1 inferior epiphyseal ring fusion	1.370	0.171	Retained
L1 superior surface morphology	1.272	0.203	Retained
L1 inferior surface morphology	1.598	0.110	Retained
L5 superior epiphyseal fusion	1.327	0.185	Retained
L5 inferior epiphyseal fusion	0.638	0.524	Retained
L5 superior surface morphology	0.451	0.652	Retained
L5 inferior surface morphology	0.976	0.329	Retained
L5 superior margin shape	1.191	0.234	Retained
L5 inferior margin shape	0.886	0.376	Retained
S1 margin shape	-0.286	0.775	Retained
S1-2 fusion	-0.241	0.809	Retained
Sacral elbow	1.184	0.237	Retained
C lipping	-0.034	0.973	Retained
C candlewax	0.158	0.874	Retained
T lipping	-0.916	0.360	Retained
T candlewax	0.234	0.815	Retained
L lipping	0.496	0.620	Retained
L candlewax	0.920	0.357	Retained
DISH	0.647	0.518	Retained
R1 fusion (L)	-0.234	0.815	Retained
R1 fusion (R)	NA	NA	NA
Sternum central dorsal ridge	NA	NA	NA
R2 rim profile (L)	-0.503	0.615	Retained
R2 rim profile (R)	0.862	0.389	Retained
R2 ossification pattern (L)	0.472	0.637	Retained
R2 ossification pattern (R)	2.393	0.018	Rejected
R3-10 rim profile (L)	1.761	0.078	Retained
R3-10 rim profile (R)	1.774	0.076	Retained
R3-10 ossification pattern (L)	2.868	0.004	Rejected
R3-10 ossification pattern (R)	2.156	0.031	Rejected

Table C.6 continued

Trait	Fisher's Z	p-value	H₀
R2-10 shingle ribs (L)	NA	NA	NA
R2-10 shingle ribs (R)	NA	NA	NA
L Parietal depression	0.346	0.729	Retained
R Parietal depression	0.161	0.872	Retained
Spheno-occipital synchondrosis	0.902	0.367	Retained
L Occipital condyle lipping	0.224	0.823	Retained
R Occipital condyle lipping	2.430	0.015	Rejected

Table C.7 Mann-Whitney U ($\alpha = 0.05$) test for male (n = 45) vs female (n = 44) trait scores with H₀ = no difference in distribution

Trait	p-value	H₀
Femur fovea margin lipping (L)	0.013	Rejected
Femur fovea margin lipping (R)	0.080	Retained
Femur head surface extra bone (L)	<0.001	Rejected
Femur head surface extra bone (R)	0.006	Rejected
Femur greater trochanter roughening (L)	0.024	Rejected
Femur greater trochanter roughening (R)	0.117	Retained
Femur trochanteric fossa exostoses (L)	0.287	Retained
Femur trochanteric fossa exostoses (R)	0.892	Retained
Femur trochanteric medial exostoses (L)	0.450	Retained
Femur trochanteric medial exostoses (R)	0.641	Retained
Tibia relative weight (L)	0.097	Retained
Tibia relative weight (R)	0.005	Rejected
Tibia actual weight (L)	<0.001	Rejected
Tibia actual weight (R)	<0.001	Rejected
Tibia length (L)	<0.001	Rejected
Tibia length (R)	<0.001	Rejected
Fibula wings (L)	0.333	Retained
Fibula wings (R)	0.701	Retained
Calcaneus relative weight (L)	<0.001	Rejected
Calcaneus relative weight (R)	<0.001	Rejected
Calcaneus actual weight (L)	<0.001	Rejected
Calcaneus actual weight (R)	<0.001	Rejected
Calcaneus length (L)	<0.001	Rejected
Calcaneus length (R)	<0.001	Rejected
Clavicle medial epiphysis fusion (L)	0.574	Retained
Clavicle medial epiphysis fusion (R)	1.000	Retained
Clavicle medial epiphysis gravel (L)	0.313	Retained
Clavicle medial epiphysis gravel (R)	0.337	Retained
Clavicle medial macroporosity (L)	0.012	Rejected
Clavicle medial macroporosity (R)	0.071	Retained
Clavicle lateral macroporosity (L)	0.191	Retained
Clavicle lateral macroporosity (R)	0.573	Retained
Scapula glenoid fossa lipping (L)	0.517	Retained
Scapula glenoid fossa lipping (R)	0.350	Retained

Table C.7 continued

Trait	p-value	H₀
Humerus relative weight (L)	<0.001	Rejected
Humerus relative weight (R)	<0.001	Rejected
Humerus actual weight (L)	<0.001	Rejected
Humerus actual weight (R)	<0.001	Rejected
Humerus length (L)	<0.001	Rejected
Humerus length (R)	<0.001	Rejected
Humerus lesser tubercle anterior surface bumps (L)	0.231	Retained
Humerus lesser tubercle anterior surface bumps (R)	0.219	Retained
Humerus lesser tubercle margin shape (L)	1.000	Retained
Humerus lesser tubercle margin shape (R)	0.786	Retained
Humerus greater tubercle pits (L)	0.820	Retained
Humerus greater tubercle pits (R)	0.490	Retained
Humerus medial epicondyle (L)	0.115	Retained
Humerus medial epicondyle (R)	0.032	Rejected
Humerus lateral epicondyle (L)	0.586	Retained
Humerus lateral epicondyle (R)	0.520	Retained
Radius tuberosity medial crest (L)	0.033	Rejected
Radius tuberosity medial crest (R)	0.102	Retained
Ulna olecranon spur (L)	0.064	Retained
Ulna olecranon spur (R)	0.455	Retained
Trapezium lipping (L)	0.960	Retained
Trapezium lipping (R)	0.487	Retained
Innominate relative weight (L)	0.001	Rejected
Innominate relative weight (R)	0.001	Rejected
Innominate actual weight (L)	<0.001	Rejected
Innominate actual weight (R)	<0.001	Rejected
Innominate length (L)	<0.001	Rejected
Innominate length (R)	<0.001	Rejected
Sacroiliac joint fusion (L)	0.176	Retained
Sacroiliac joint fusion (R)	1.000	Retained
Iliac crest fusion (L)	0.801	Retained
Iliac crest fusion (R)	0.460	Retained
Iliac crest tuberculum ossification (L)	0.939	Retained
Iliac crest tuberculum ossification (R)	0.147	Retained
Ilium anterior inferior Iliac spine exostoses (L)	0.303	Retained
Ilium anterior inferior iliac spine exostoses (R)	0.326	Retained
Acetabulum posterior margin lipping (L)	0.527	Retained
Acetabulum posterior margin lipping (R)	0.091	Retained
Acetabulum posterior articular surface extra bone (L)	0.111	Retained
Acetabulum posterior articular surface extra bone (R)	1.000	Retained
Acetabulum inferior joint lipping (L)	0.877	Retained
Acetabulum inferior joint lipping (R)	0.240	Retained
Ischial tuberosity superior margin spur (L)	0.100	Retained
Ischial tuberosity superior margin spur (R)	0.314	Retained
Ischial tuberosity medial spur (L)	0.165	Retained
Ischial tuberosity medial spur (R)	0.082	Retained

Table C.7 continued

Trait	p-value	H ₀
Ischial tuberosity bumps (L)	0.454	Retained
Ischial tuberosity bumps (R)	0.310	Retained
Auricular surface inferior surface porosity (L)	0.675	Retained
Auricular surface inferior surface porosity (R)	1.000	Retained
Auricular surface superior posterior iliac exostoses (L)	0.372	Retained
Auricular surface superior posterior iliac exostoses (R)	0.243	Retained
Auricular surface inferior posterior iliac exostoses (L)	0.879	Retained
Auricular surface inferior posterior iliac exostoses (R)	0.822	Retained
Auricular surface posterior exostoses (L)	0.917	Retained
Auricular surface posterior exostoses (R)	0.099	Retained
Pubic symphyseal collar (L)	0.957	Retained
Pubic symphyseal collar (R)	0.343	Retained
Pubic symphyseal relief (L)	0.170	Retained
Pubic symphyseal relief (R)	0.592	Retained
Pubic superior protuberance/apex (L)	0.601	Retained
Pubic superior protuberance/apex (R)	0.792	Retained
Pubic symphysis ventral margin (L)	0.172	Retained
Pubic symphysis ventral margin (R)	0.485	Retained
Pubic symphysis dorsal margin (L)	0.230	Retained
Pubic symphysis dorsal margin (R)	0.385	Retained
C1 lipping	0.924	Retained
C1 eburnation	0.871	Retained
L1 spinous process flattening	0.963	Retained
L1 superior epiphyseal ring fusion	0.180	Retained
L1 inferior epiphyseal ring fusion	0.727	Retained
L1 superior surface morphology	0.266	Retained
L1 inferior surface morphology	0.321	Retained
L5 superior epiphyseal fusion	0.155	Retained
L5 inferior epiphyseal fusion	0.466	Retained
L5 superior surface morphology	0.571	Retained
L5 inferior surface morphology	0.133	Retained
L5 superior margin shape	0.177	Retained
L5 inferior margin shape	0.085	Retained
S1 margin shape	0.202	Retained
S1-2 fusion	0.448	Retained
Sacral elbow	0.046	Rejected
C lipping	0.054	Retained
C candlewax	0.027	Rejected
T lipping	0.818	Retained
T candlewax	0.163	Retained
L lipping	0.031	Rejected
L candlewax	0.010	Rejected
DISH	0.089	Retained
R1 fusion (L)	0.971	Retained
R1 fusion (R)	0.155	Retained
Sternum central dorsal ridge	0.134	Retained

Table C.7 continued

Trait	p-value	H₀
R2 rim profile (L)	0.273	Retained
R2 rim profile (R)	0.507	Retained
R2 ossification pattern (L)	0.839	Retained
R2 ossification pattern (R)	1.000	Retained
R3-10 rim profile (L)	0.188	Retained
R3-10 rim profile (R)	0.372	Retained
R3-10 ossification pattern (L)	0.065	Retained
R3-10 ossification pattern (R)	0.289	Retained
R2-10 shingle ribs (L)	0.364	Retained
R2-10 shingle ribs (R)	NA	NA
L Parietal depression	0.884	Retained
R Parietal depression	0.457	Retained
Spheno-occipital synchondrosis	0.597	Retained
L Occipital condyle lipping	0.124	Retained
R Occipital condyle lipping	0.330	Retained

C.4 Differences between the left and right sides

Table C.8 Percentile bootstrap on a Spearman's rho test ($\alpha = 0.05$, $n = 89$, $B = 2000$) comparing the correlation of trait-age between the left and right side

Trait	p-value	H₀
Femur fovea margin lipping	0.774	Retained
Femur head surface extra bone	0.501	Retained
Femur greater trochanter roughening	0.386	Retained
Femur trochanteric fossa exostoses	0.071	Retained
Femur trochanteric medial exostoses	0.958	Retained
Tibia relative weight	0.467	Retained
Tibia actual weight	0.555	Retained
Tibia length	0.652	Retained
Fibula wings	0.151	Retained
Calcaneus relative weight	0.132	Retained
Calcaneus actual weight	0.999	Retained
Calcaneus length	0.044	Rejected
Clavicle medial epiphysis fusion	0.996	Retained
Clavicle medial epiphysis gravel	0.412	Retained
Clavicle medial macroporosity	0.123	Retained
Clavicle lateral macroporosity	0.532	Retained
Scapula glenoid fossa lipping	0.163	Retained
Humerus relative weight	0.035	Rejected
Humerus actual weight	0.022	Rejected
Humerus length	0.264	Retained
Humerus lesser tubercle anterior surface bumps	0.214	Retained
Humerus lesser tubercle margin shape	0.868	Retained
Humerus greater tubercle pits	0.929	Retained
Humerus medial epicondyle	0.911	Retained
Humerus lateral epicondyle	0.252	Retained

Table C.8 continued

Trait	p-value	H₀
Radius tuberosity medial crest	0.252	Retained
Ulna olecranon spur	0.260	Retained
Trapezium lipping	0.454	Retained
Innominate relative weight	0.266	Retained
Innominate actual weight	0.724	Retained
Innominate length	0.942	Retained
Sacroiliac joint fusion	0.071	Retained
Iliac crest fusion	0.953	Retained
Iliac crest tuberculum ossification	0.167	Retained
Ilium anterior inferior iliac spine exostoses	0.213	Retained
Acetabulum posterior margin lipping	0.629	Retained
Acetabulum posterior articular surface extra bone	0.786	Retained
Acetabulum inferior joint lipping	0.139	Retained
Ischial tuberosity superior margin spur	0.765	Retained
Ischial tuberosity medial spur	0.372	Retained
Ischial tuberosity bumps	0.435	Retained
Auricular surface inferior surface porosity	0.252	Retained
Auricular surface superior posterior iliac exostoses	0.662	Retained
Auricular surface inferior posterior iliac exostoses	0.799	Retained
Auricular surface posterior exostoses	0.076	Retained
Pubic symphyseal collar	<0.001	Rejected
Pubic symphyseal relief	<0.001	Rejected
Pubic superior protuberance/apex	0.452	Retained
Pubic symphysis ventral margin	0.224	Retained
Pubic symphysis dorsal margin	0.028	Rejected
R1 fusion	0.579	Retained
R2 rim profile	0.989	Retained
R2 ossification pattern	0.816	Retained
R3-10 rim profile	0.401	Retained
R3-10 ossification pattern	0.130	Retained
R2-10 shingle ribs	NA	NA
Parietal depression	0.921	Retained
Occipital condyle lipping	0.993	Retained

Table C.9 Wilcoxon signed rank test (paired, n = 89, $\alpha = 0.05$) for left vs right trait scores

Trait	Z	p-value	H₀
Femur fovea margin lipping	-0.905	0.366	Retained
Femur head surface extra bone	0.850	0.396	Retained
Femur greater trochanter roughening	-0.509	0.611	Retained
Femur trochanteric fossa exostoses	1.387	0.166	Retained
Femur trochanteric medial exostoses	-1.732	0.083	Retained
Tibia relative weight	<0.001	1.000	Retained
Tibia actual weight	-2.437	0.015	Rejected
Tibia length	0.249	0.803	Retained
Fibula wings	<0.001	1.000	Retained
Calcaneus relative weight	NA	NA	NA
Calcaneus actual weight	-0.925	0.355	Retained
Calcaneus length	0.143	0.887	Retained
Clavicle medial epiphysis fusion	0.272	0.786	Retained
Clavicle medial epiphysis gravel	-0.707	0.480	Retained
Clavicle medial macroporosity	1.015	0.310	Retained
Clavicle lateral macroporosity	0.258	0.7963	Retained
Scapula glenoid fossa lipping	-2.846	0.004	Rejected
Humerus relative weight	-1.414	0.157	Retained
Humerus actual weight	-5.607	<0.001	Rejected
Humerus length	-6.035	<0.001	Rejected
Humerus lesser tubercle anterior surface bumps	<0.001	1.000	Retained
Humerus lesser tubercle margin shape	0.218	0.827	Retained
Humerus greater tubercle pits	-0.905	0.366	Retained
Humerus medial epicondyle	<0.001	1.000	Retained
Humerus lateral epicondyle	-0.378	0.706	Retained
Radius tuberosity medial crest	1.000	0.317	Retained
Ulna olecranon spur	-1.508	0.132	Retained
Trapezium lipping	-0.378	0.706	Retained
Innominate relative weight	1.000	0.317	Retained
Innominate actual weight	-1.275	0.202	Retained
Innominate length	1.335	0.182	Retained
Sacroiliac joint fusion	1.336	0.182	Retained
Iliac crest fusion	1.642	0.101	Retained
Iliac crest tuberculum ossification	0.545	0.586	Retained
Ilium anterior inferior Iliac spine exostoses	<0.001	1.000	Retained
Acetabulum posterior margin lipping	0.535	0.593	Retained
Acetabulum posterior articular surface extra bone	-0.333	0.739	Retained
Acetabulum inferior joint lipping	-1.897	0.058	Retained
Ischial tuberosity superior margin spur	1.000	0.317	Retained
Ischial tuberosity medial spur	-1.000	0.317	Retained
Ischial tuberosity bumps	0.500	0.617	Retained
Auricular surface inferior surface porosity	1.134	0.259	Retained
Auricular surface superior posterior iliac exostoses	<0.001	1.000	Retained
Auricular surface inferior posterior iliac exostoses	-0.333	0.739	Retained
Auricular surface posterior exostoses	-0.243	0.808	Retained

Table C.9 continued

Trait	Z	p-value	H₀
Pubic symphyseal collar	0.577	0.564	Retained
Pubic symphyseal relief	-1.000	0.317	Retained
Pubic superior protuberance/apex	0.277	0.782	Retained
Pubic symphysis ventral margin	1.269	0.205	Retained
Pubic symphysis dorsal margin	1.342	0.180	Retained
R1 fusion	<0.001	1.000	Retained
R2 rim profile	-0.447	0.655	Retained
R2 ossification pattern	0.022	0.982	Retained
R3-10 rim profile	1.000	0.317	Retained
R3-10 ossification pattern	1.890	0.059	Retained
R2-10 shingle ribs	1.000	0.317	Retained
Parietal depression	-0.378	0.706	Retained
Occipital condyle lipping	1.000	0.317	Retained

Appendix D Similarity report

fulldissertation-FINAL.docx			
ORIGINALITY REPORT			
15%	9%	12%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	hdl.handle.net Internet Source	1%	
2	Submitted to University of Witwatersrand Student Paper	1%	
3	etheses.whiterose.ac.uk Internet Source	<1%	
4	www.statemachine.net Internet Source	<1%	
5	Jooste, Nicolene. "Validating the Accuracy and Repeatability of Transition Analysis for Age Estimation in South Africa", University of Pretoria (South Africa), 2023 Publication	<1%	
6	Getz, Sara Marie. "Improved Skeletal Age-at-Death Estimation and Its Impact on Archaeological Analyses.", The Pennsylvania State University, 2019 Publication	<1%	
7	api.research-repository.uwa.edu.au Internet Source	<1%	