CrossMark

Age estimation from the acetabulum in South African black males

D. Botha[1] · S. Pretorius[2] · J. Myburgh[1] · M. Steyn[3]

Abstract Anthropologists are constantly seeking to improve methods for age estimation in the human skeleton. A new method was introduced about a decade ago that assesses the morphological changes that take place in the acetabulum as an individual ages. The pelvis is usually well preserved in forensic cases, which makes this method potentially valuable as an adult age indicator. This method employs seven variables, each with its own set of phases. To test the accuracy and reliability of this method, 100 black South African male acetabula from the Pretoria Bone Collection were assessed based on the criteria described in the original study. Box plots and transition curves were constructed to establish whether progression with age was visible and how it could possibly be modelled. Inter-observer reliability was also assessed by making use of Fleiss's Kappa statistic. Five specimens were used as out-of-sample examples for which maximum likelihood (point) estimates were calculated. The results demonstrated that middle and older individuals' age estimates were vastly underestimated. Inter-observer repeatability was poor, which suggested that the classification system most likely needs to be modified. A discussion and recommendation is given for improvement of reliability and repeatability of this method.

Keywords Adult · Age estimation · Acetabulum · Black males · Inter-observer

Introduction

Age estimation from adult skeletal remains is notoriously difficult as changes in the skeleton are slow and inconsistent once adulthood has been reached. This becomes even more difficult in older individuals, resulting in many osteologists simply reverting to estimates such as *older than 50* or *of advanced age* [1]. Age changes in the adult skeleton are complex and occur gradually, and levels of inter-individual variation are high.

Skeletal ageing depends on the individual's genetic make-up, lifestyle and nutrition. In younger adults a relatively accurate estimate can be obtained, but in the middle-aged groups and older adults estimates become more difficult [1–3]. Many authors have noted that existing methods tend to overestimate age of young adults and underestimate that of old individuals, implying that many methods/skeletal features do not show much progression with age. Konigsberg and Frankenberg [4] noted that the observed low incidence of older adults in archaeological populations is most probably due to our poor ability to age them correctly, rather than it being a true reflection of what is happening in that specific population. It is thus clear that more research is needed on adult age estimation, and that we need to continue to improve on existing methods and attempt to find new methods to accurately age individuals. These methods and the standards it provide should be based on modern skeletal material in order to apply it in a forensic context.

Recently, the statistics applied in age estimation models have been the centre of many debates in the literature [1,

✉ D. Botha
deona.botha@gmail.com

1   Department of Anatomy, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria 0007, South Africa

2   Department of Mathematics and Applied Mathematics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

3   School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

5–7], particularly when it comes to multifactorial age estimation. Confidence intervals are particularly problematic when multifactorial methods are used, and it is not always clear exactly how it should be calculated [7]. Issues such as prior distributions of age of the reference sample [1, 8] and various statistical approaches such as Bayesian statistics [9, 10] and transition analysis [1, 11] are discussed at length. Although wide age ranges are sometimes obtained by using transition analysis, this approach has succeeded in giving improved accuracy and specificity in terms of age estimation and also provides a 95 % confidence interval [12, 13].

For many years, gross morphological age estimations of the adult skeleton have focused mainly on the cranium, ribs and pelvis. Relatively few new methods have been added in the past two decades, although many studies have been published that test the most popular methods in terms of accuracy and inter-observer repeatability. The development of population-specific standards and the use of multifactorial analysis has also enjoyed much attention.

In 2006, Rissech et al. [14, 15] made known a method that described morphological age changes in the acetabulum that are potentially valuable in estimating age at death of adults, although this has not been widely tested. As the os coxa is usually well preserved in forensic cases, it serves as a widespread and accessible age indicator. This method [14, 15] used seven variables of the acetabulum in their assessment: the acetabular groove, acetabular rim shape, acetabular rim porosity, apex activity, activity on the outer edge of the acetabular fossa, activity of the acetabular fossa and porosities of the acetabular fossa. These authors examined os coxae of males from Portugal and Bayesian inference was used as a statistical tool. Their results showed significant correlation of each trait with age with small inter- and intra-observer error. Differences between known and estimated ages were within a 20-year range for 89 % of the specimens. Testing this on other samples, good results were found, although estimates became less accurate as geographical distance increased.

Calce and Rogers [16] used a Canadian sample to test the precision of the Rissech et al. scoring techniques, evaluate the age estimates for individuals over 40 and compare the results obtained by using different reference populations (i.e., test the impact of choosing other reference samples). They found that the technique tended to underestimate age, but overall, more than 80 % of their estimates were within 12 years of known age. In a follow-up study, Calce [17] suggested a simplified version the method, using the three features that gave the most accurate results in the previous study [16] (acetabular groove, osteophyte development and apex growth), and found improved results when assigning individuals to one of three general age groups (young adults 17–39, middle adults 40–64, old adults 65+).

Rougé-Maillart et al. [18, 19] endeavoured to develop a more user-friendly method than what was proposed by the original authors. They combined the auricular surface and acetabular criteria and used four auricular surface and three acetabular (rim—scored from 1 to 5; fossa—scored from 1 to 4; apical activity—scored from 1 to 3) traits. A composite score is then obtained by adding the values of all seven variables, which may be read from a table. Results showed that the criteria used to score the acetabulum correlated fairly well with age, but when the acetabular scores were combined with that of the auricular surface, the effective results were improved. However, inter-observer variability was low.

The aim of this study was to assess degenerative changes in the acetabula of South African black males using the original Rissech method in order to estimate age. As the changes are subtle and complex to score, a strong emphasis was placed on the ability of three observers to consistently score the changes, in order to ascertain which changes can be assessed with a high degree of reliability. The accuracy and repeatability of this method was tested in order to provide recommendations for using the acetabulum in the estimation of age in black South African populations.

Materials and methods

Skeletons used in this study were randomly selected from the Pretoria Bone Collection [20]. This cadaver-based collection houses skeletons of modern, known individuals. As the collection is still growing, these individuals represent the currently living population. One hundred skeletons of African males were included in the study, with a roughly uniform prior distribution with ages ranging from 16 to 96 years (Table 1). Specimens showing pathology (for example fused sacro-iliac joints) were excluded.

All assessments were done blindly (i.e., without knowing the age of the specimen). Only the left os coxae of the individuals were placed in random order on a table, and observers were instructed not to view the pubis. Three observers scored all 100 acetabula. The seven features, as outlined by Rissech

Table 1 Sample sizes for the different age groups used in age estimation of black African males using the acetabulum

Age group (years)	Number of specimens
15–19	2
20–29	14
30–39	14
40–49	13
50–59	14
60–69	16
70–79	13
80–89	12
90–99	2
Total	100

et al. [14] were scored and are summarised in Table 2. The detailed descriptions of each phase can be found in Rissech et al. [14]. All three observers are skilled osteologists with experience in skeletal analysis.

In the statistical analysis, box plots were firstly constructed for all seven variables to show the age range, mean, standard error and standard deviation for each feature. Inter-observer reliability agreement was then tested by means of Fleiss's Generalised Kappa statistical measure [21, 22].

A generalised linear model, as described by Boldsen et al. [1], was subsequently used to determine the likelihood function of an attribute being in a specific stage of classification. Of the sample, 95 skeletons were used to fit the model and five were excluded to be used as out-of-sample tests for a post-assessment investigation. By assuming that the prior distribution of age at death is uniform between the ages under investigation, a posterior density function was determined for the age at death, given that the observed variable is in a specific phase.

Variables and their associated phases (as used in the original application of this method) were used by the three observers to classify the acetabula, therefore a generalised linear model could be used without modification of the assumptions made by Boldsen et al. [1]. We therefore fitted a discrete time proportional hazards model or continuation ratio model.

Age progression was assessed by means of transition analysis—firstly for individual traits and then for traits in combination. Transition curves for all seven attributes and the three observers were constructed. This indicated the age-specific probability that a variable will transition from one stage to the next.

In order to develop a likelihood curve for the age at death for an individual, all traits need to be combined. This is done by multiplying the individual likelihood functions for each of the variables with one another, which yields a large number of possible combinations. The age that maximises the resulting likelihood function is then known as the maximum likelihood of age at death. Although not all combinations are likely to occur, it is necessary to provide age at death estimates for a few specimens to validate the method. Out-of-sample observations were considered and combined likelihood curves plotted. Five observations from each observer were not included in the fitted model and were instead used as the out-of-sample tests to demonstrate how multiple traits may be combined.

Results

Box plots drawn up for each of the seven variables as scored by each of the three observers were used to establish whether the attributes were, in fact, progressing with age. The box plots are shown in Fig. 1a–g. It was found that in most cases there was some progression with age. Progression with age was not seen for the acetabular rim porosity (observers B and C) and for activity of the acetabular fossa (observer C). This could possibly indicate that these indicators/features do not behave biologically the same in African individuals as in their European counterparts, i.e., porosity of the acetabular fossa may not develop in African populations.

The overall results of the Fleiss's Generalised kappa statistics testing for inter-observer repeatability, with the corresponding p values, are summarised in Table 3. The results did not show high levels of agreement between the observers. Fair agreement (according to interpretation of Kappa values given by Landis and Koch [23]) between observers was seen in variables 2 and 4. The full set of results, including confidence intervals and the results for specific categories (phases) is given in Appendix A. From these results, it was interesting to see that most agreement (highest Kappa value for each variable) occurred in the highest possible category (phase)
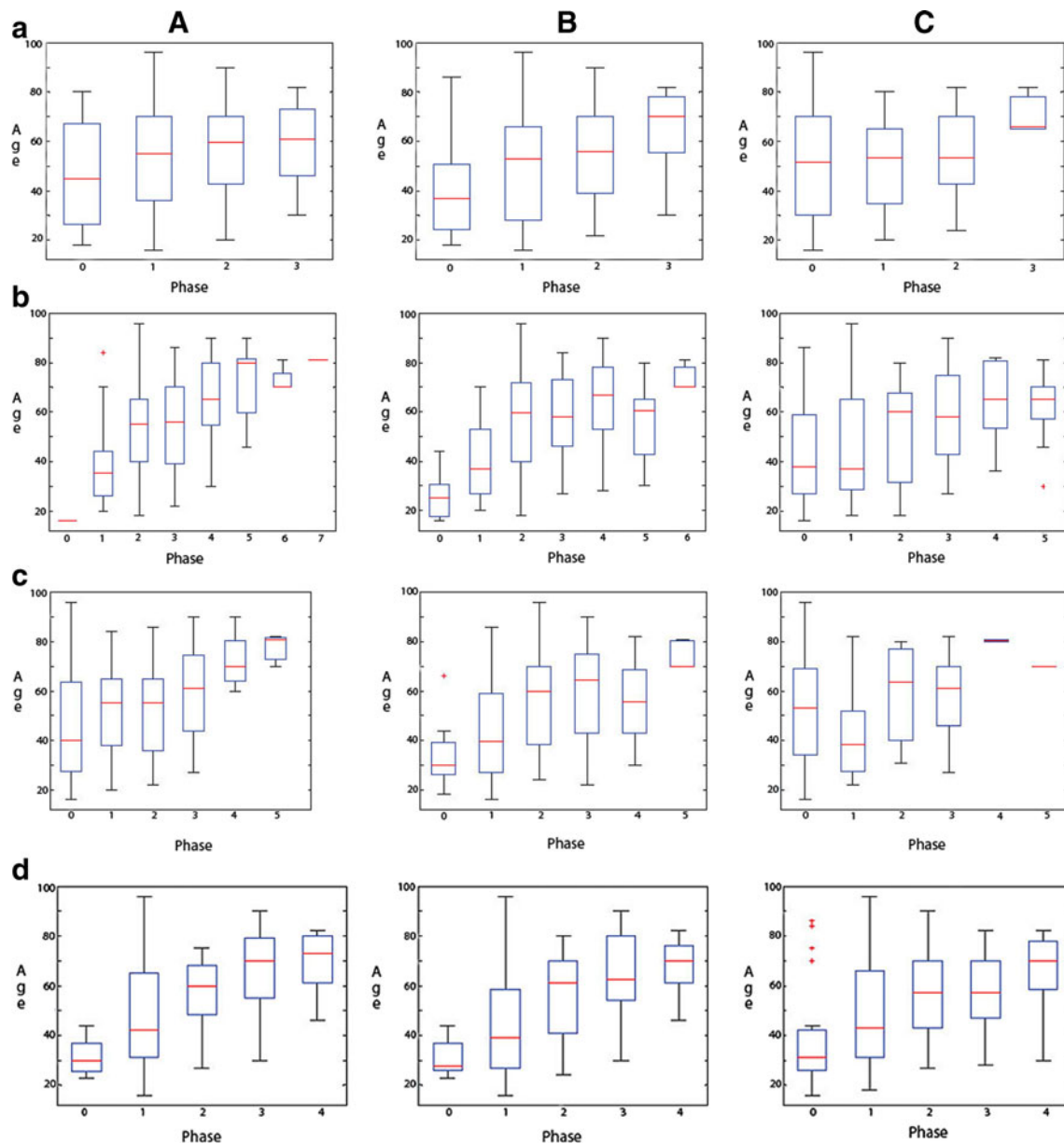
Table 2 Seven variables observed on the acetabulum, with possible score categories [14]

Variable	Description
Acetabular groove 0–3	Groove below and around margin of acetabular rim. Changes from no groove to very pronounced groove
Acetabular rim shape 0–6	Acetabular rim changes from round and smooth, to narrow and irregular, eventually forming a high crest due to osteophyte formation
Acetabular rim porosity 0–5	Acetabular rim changes from smooth without porosities, to having micro- and then macro-porosities
Apex activity 0–4	Apex of the posterior horn of the lunate surface changes from round, to having a small osteophyte and then a large osteophyte
Activity on outer edge of acetabular fossa 0–5	Osteophytic formation occurs as mini crest from outer edge of acetabular fossa towards lunate surface. This edge changes from smooth, to having an observable crest to a partially covered fossa as the outer edge is completely destructed
Activity of the acetabular fossa 0–5	Initially the lunate surface is level with the acetabular fossa, with age the fossa becomes deeper. The fossa gradually becomes filled with new bone formation
Porosities of the acetabular fossa 0–6	Acetabular fossa changes from dense, to having micro-porosities and then macro-porosities. Eventually bone proliferation obliterates fossa

Fig. 1 Boxplots for all variables (**a**–**g**) assessed by observers A, B and C. **a** Acetabular groove. **b** Acetabular rim shape. **c** Acetabular rim porosity. **d** Apex activity. **e** Activity on outer edge of acetabular fossa. **f** Activity of the acetabular fossa. **g** Porosities of the acetabular fossa

for each variable. This may possibly indicate that morphological characteristics seen in older individuals are clearer and more easily classified than age-related changes seen in the middle-aged adult group.

The transition results for observer B was used as this was the most experienced participant in the study (Fig. 2). The curves behaved as expected, since we assume that progression from one state to the next can only happen in a forward manner. The transition curves for variables 1, 3, 5, 6, and 7 seem to show slow morphological change as age progresses. Variables 2 and 4, however, displayed a clearer age progression pattern.

Five skeletons were excluded from the fitted sample in order to provide out-of-sample data that could be used to

preliminarily assess the fit of the model. Each combination of traits was used to derive a likelihood function for each observer's fitted model. The associated likelihood functions are outlined in Fig. 3a–e.

The maximum likelihood ages (point estimates) are outlined in Table 4. Save for out-of-sample number 3, which offers fairly accurate age estimates, the remaining observed samples are poorly estimated. Observer C seems to have the maximum likelihood age estimates closest to the actual ages, whereas the estimates of observers A and B are far off the mark. There are likely several reasons for the discrepancy. When we look at the classifications used for these skeletons, some classifications seem very unlikely if we consider the
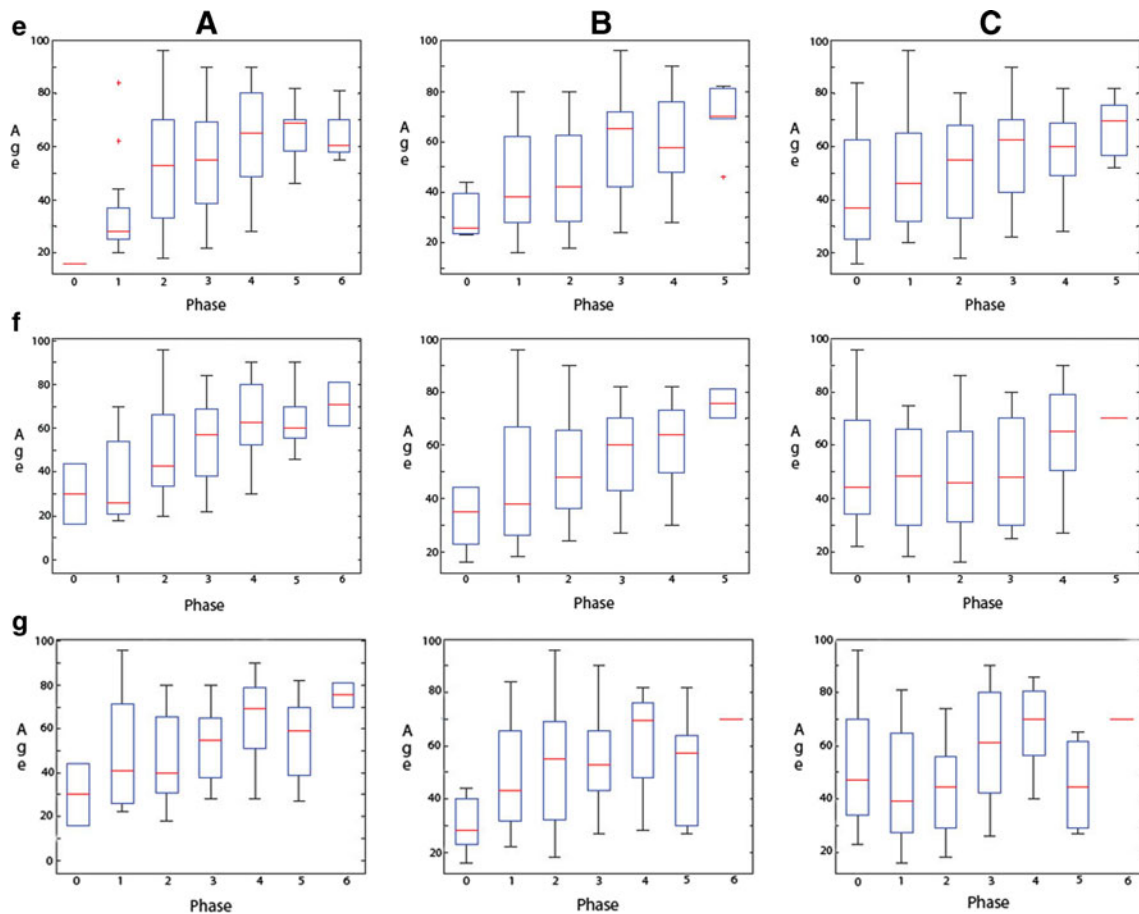
Fig. 1 (continued)

assumption that the stages of indicators progress with age. For example, when considering Observer A's results (Table 5), skeleton 1 was classified in stage 1 for both variables 4 and 5. We would expect an individual of age 65 to be in a higher stage of classification for both of these variables.

The method, does however, seem to be more accurate when used in younger individuals. For example, out-of-sample skeleton 3 is of young age and the classification indicates this well. This may suggest that, in this population group, very few of the individuals actually progress towards a stage where high scores are observed, especially in those variables where increased porosity is expected (V3 and V7). A possible

explanation for this observation may be linked to the differences in bone microstructure between black and white individuals. Black individuals have higher bone turnover rates than whites and are less prone to bone failure due to fatigue, which suggests that black individuals have a stronger bone structure and are less prone to developing bone porosity [24].

Discussion

In this study, the changes observed in the acetabulum for estimation of age as described by Rissech et al. [14] were tested for its accuracy and repeatability in black South African males. In general, poor results were obtained.

Previous publications investigating this method have reported that there are a number of conditions which should be met before reliable results can be obtained when using the acetabulum in age estimation. Rissech et al. [14] stated that firstly, all seven traits must be intact for assessment and that damaged acetabula should not be used for estimating age. Secondly, the data sample for a specific population should be large and from the same geographical area in order to establish the relationship between the different phases of each

Table 3 Inter-observer results

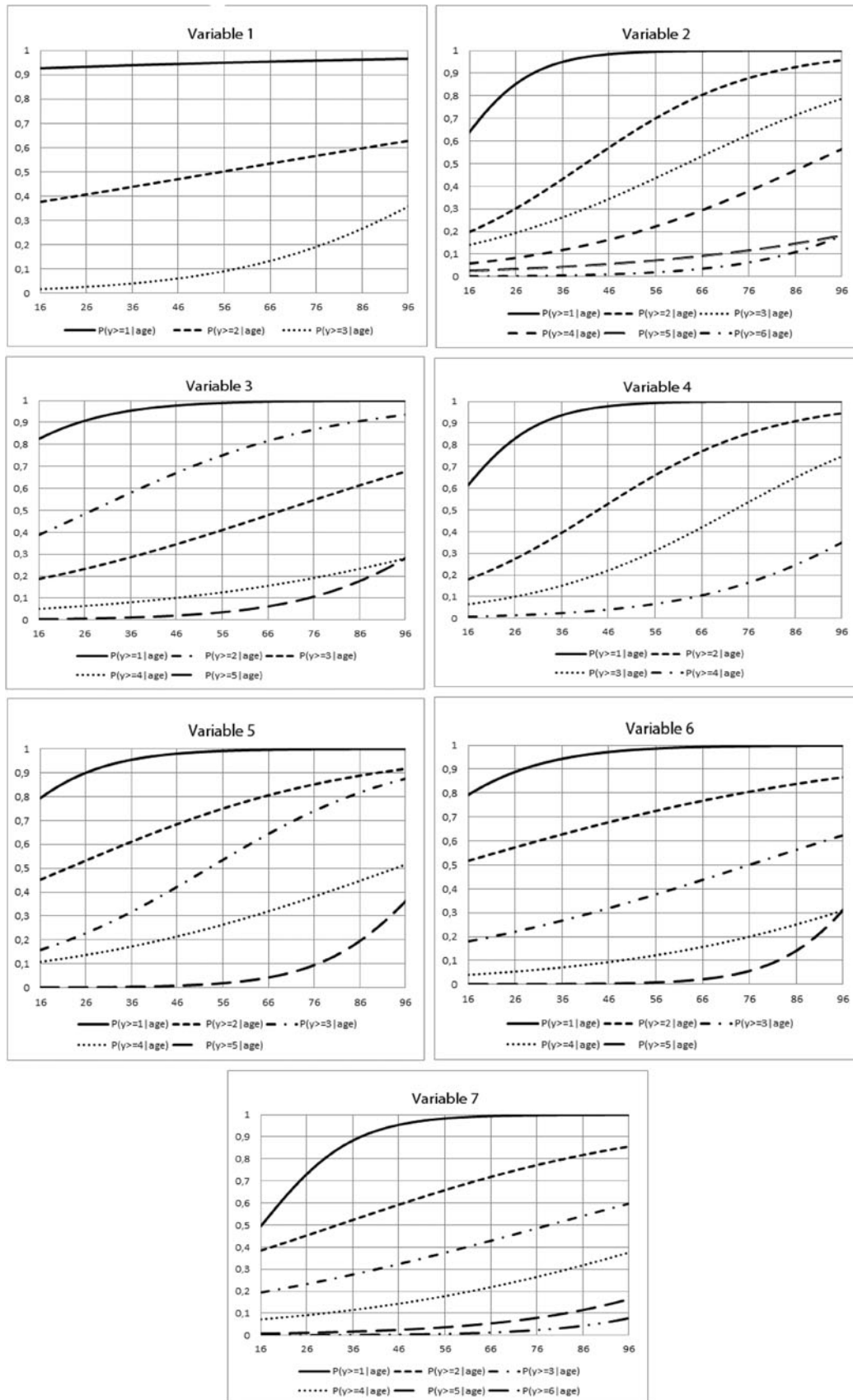| Variable	Kappa	p value
V1	−0.0391	0.9943
V2	0.2302	0.0000
V3	−0.0105	0.6523
V4	0.3021	0.0000
V5	0.0683	0.0010
V6	0.1084	0.0000
V7	0.1658	0.0000

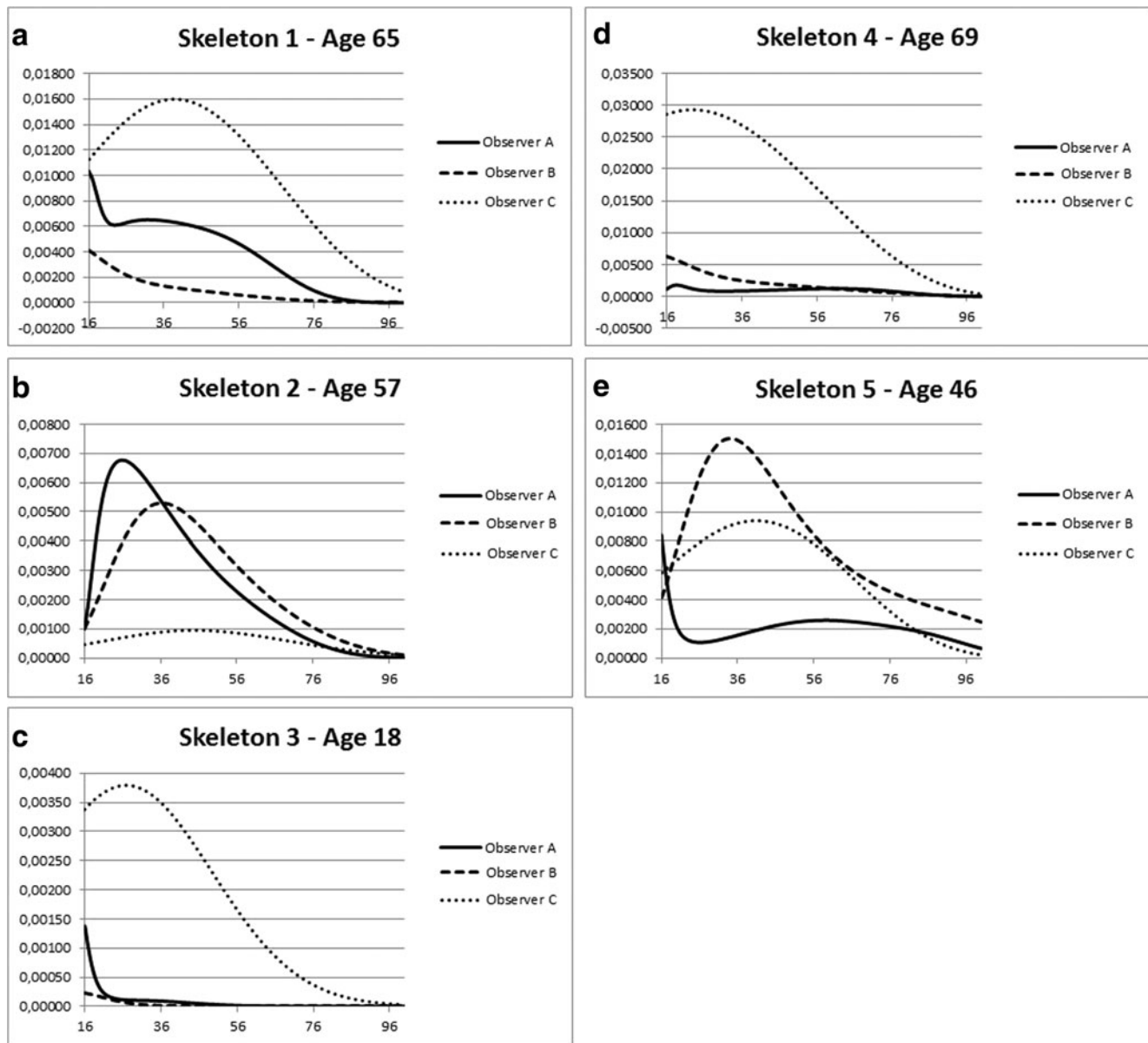**Fig. 2** Transition curves for variables 1 to 7 (Observer B)

Fig. 3 Likelihood curves for out-of-sample specimens 1 to 5 (**a–e**)

variable with known age at death. Calce and Rogers [16] mentioned that it is imperative that traits be scored in a consistent manner to yield accurate results. The first two conditions were met, but results from this study suggested that

Table 4 Point estimates for five out-of-sample specimens

Out-of-sample number	Actual age	Observer A: Estimated age	Observer B: Estimated age	Observer C: Estimated age
1	65	16	16	40
2	57	26	36	45
3	18	16	16	27
4	69	19	16	23
5	46	16	33	41

scoring in a consistent/repeatable manner may not be possible. Although this method is potentially valuable in estimating age at death, it thus presents with several complications.

Scoring consistency, as well as inter-observer repeatability and reliability are, in general, a major problem in all adult age estimation methods, since most methods are qualitative and open to interpretation. The various features all change gradually, and the transition from one stage to the next is not always clear or exact. Detailed descriptions, drawings and casts have been used, for example, in sternal rib end and pubic symphysis analysis to help the observer, but it seems that there is a wide variation in how an individual case is scored, even amongst experienced observers [25]. The sternal rib and pubic symphysis methods have proven to be relatively accurate in terms of inter-observer reliability [26–28] which has allowed for it to

Table 5 Observer A's classifications for the five out-of-sample skeletons

Out-of-sample number	Actual age	Estimated age (Maximum likelihood)	V1	V2	V3	V4	V5	V6	V7
1	65	16	2	4	2	1	1	2	2
2	57	36	2	5	2	3	3	2	1
3	18	16	0	0	1	1	1	1	0
4	69	16	2	4	1	3	3	2	2
5	46	33	3	3	3	3	5	2	3

become established methods of age estimation for forensic anthropologists, albeit it with wide ranges. The current method, however, showed poor inter-observer repeatability and thus needs some refinement before it can be recognised as an accurate method for age estimation.

From a statistical point of view, the main critique of the method employed here probably lies not with the model but with the classification system used for the attributes of each skeleton. Almost all of the evidence points to the system being too subjective, or possibly unsuitable for non-European skeletons. This is evidenced by the box plots indicating that the stages of each variable are not all progressing significantly with age and is supported by the lack of agreement between the observers as indicated by the Kappa statistics. The generalised linear models fitted do not offer good maximum likelihood estimates of age when the classifications are inconsistent with the actual age. In this study only point estimates were used, but it can be assumed that the age ranges (confidence intervals) would have been equally unreliable. It is also interesting to note that in a recent South African study testing the original Boldsen et al. [1] transition analysis method that include cranial sutures, pubic symphyses and auricular surfaces, equally poor results were found [29]. In this study it was found that the age ranges generated by the programme were so wide that it encompassed the entire adult age range, making them practically useless. It can be suggested that the skeletons of black South Africans, representative of populations who are not prone to develop osteoporosis, react differently when it comes to changes with age, and may show less porosity of bone with advancing age compared to European populations.

The results of this study indicated that variables one (acetabular groove) and four (apex activity) gave the most reliable age estimates and were found to be the most user-friendly. The other variables should be refined according to the degree of variability and observable transition of the specific trait in the population, the number of traits assessed reduced, or complete exclusion should be considered. Since population standards are essential for this method, it is necessary to adapt the number of variables and its corresponding phases to the morphological characteristics of the acetabulum observed in a specific population. Calce [17] also simplified the method by reducing the number of variables to three (acetabular groove,

osteophyte development of the rim and apex growth). All variables incorporating the fossa were excluded. They broadly classified the changes that develop with age into young, middle and old adult groups. The results of the current study are in general agreement with at from Calce [17].

Rougé-Maillart and colleagues [18, 19] mentioned that age-related changes in the acetabulum are slow to develop and morphological maturity is most likely reached at a late stage in life. This seems particularly true for our sample, where many of the older individuals were vastly underestimated. It seems that in this population, some of the individuals may simply not develop the described old age characteristics. This is something that needs to be followed up in future studies.

Although initial results for this study are poor, it is suggested that some adjustments may improve this method for age estimation in black African males. We recommend that the number of variables be decreased to four, namely the acetabular groove (variable 1), acetabular rim shape and porosity (variable 2; as a combination of variable 2 and 3 as described by Rissech et al. [14]), apex activity (variable 3; previously variable 4) and activity of the acetabular fossa and its outer edge (variable 4; combination of variable 5, 6 and 7 of the Rissech method). By means of re-assessment of the sample and transition analysis more accurate age estimations and inter-observer reliability could possibly be reached. Re-assessment with an expanded sample containing both sexes is the ideal and the goal is to establish a higher correlation of age with refinement of criteria.

References

1. Boldsen JL, Milner GR, Konigsberg LW, Wood JW (2002) Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW (eds) Paleodemography. Age distributions from skeletal samples. Cambridge University Press, Cambridge, pp 73–106

2. Saunders SR, Fitzgerald C, Rogers T, Dudar C, Mckillop H (1992) A test of several methods of skeletal age estimations using a documented archaeological sample. Can Soc Forensic Sci 25:97–118

3. Loth SR, İşcan MY (2000) Morphological age esitimation. In: Siegel JA, Saukko PJ, Knupfer GC (eds) Encyclopedia of forensic science. Academic, San Diego, pp 242–252

4. Konigsberg LW, Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am J Phys Anthropol 89:235–256

5. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH (2008) Estimation and evidence in forensic anthropology: age-at-death. J Forensic Sci 53:541–557

6. Rogers T (2009) Skeletal age estimation. In: Blau S, Ubelaker DH (eds) Handbook of forensic anthropology and archaeology. Left Coast Press, California, pp 208–221

7. Garvin HM, Passalacqua NV, Uhl NM, Gipson DR, Overbury RS, Cabo LL (2012) Developments in forensic anthropology: age-at-death estimation. In: Dirkmaat DC (ed) A companion to forensic anthropology. Wiley-Blackwell, West Sussex, pp 202–223

8. Bocquet-Appel JP, Masset C (1982) Farewell to the paleodemography. J Hum Evol 12:321–333

9. Lucy D, Aykroyd RG, Pollard AM, Solheim T (1996) A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J Forensic Sci 41:189–194

10. Phillips LD (1973) Bayesian statistics for social scientists. Nelson, London

11. Milner GR, Boldsen JL (2012) Transition analysis: a validation study with known-age modern American skeletons. Am J Phys Anthropol 148:98–110

12. Godde K, Hens S (2012) Age-at-death estimation in an Italian historic sample: a test of the Suchey-Brooks and transition analysis methods. Am J Phys Anthropol 138:31–38

13. Godde K, Hens S (2015) Modeling senescence changes of the pubic symphysis in historic Italian populations: comparison of the Rostock and forensic approaches to aging using transition analysis. Am J Phys Anthropol 156:466–473

14. Rissech C, Estabrook GF, Cunha E, Malgosa A (2006) Using the acetabulum to estimate age-at-death of adult males. J Forensic Sci 51:213–229

15. Rissech C, Estabrook GF, Cunha E, Malgosa A (2007) Estimation of age-at-death for adult males using the acetabulum, applied to four western European populations. J Forensic Sci 52:774–778

16. Calce SE, Rogers TL (2011) Evaluation of age estimation technique: testing traits of the acetabulum to estimate age at death in adult males. J Forensic Sci 56:302–311

17. Calce SE (2012) A new method to estimate adult age-at-death using the acetabulum. Am J Phys Anthropol 148:11–23

18. Rougé-Maillart C, Jousset N, Vielle B, Gaudin A, Telmon N (2007) Contribution of the study of acetabulum for the estimation of adult subjects. Forensic Sci Int 171:103–110

19. Rougé-Maillart C, Vielle B, Jousset N, Chappard D, Telmon N, Cunha E (2009) Development of a method to estimate skeletal age at death in adults using the acetabulum and the auricular surface on a Portuguese population. Forensic Sci Int 188:91–95

20. L'Abbé EN, Loots M, Meiring JH (2005) The Pretoria Bone Collection: a modern South African skeletal sample. HOMO – J Comp Hum Biol 56:197–205

21. Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76:378–382

22. King J (2004) Software solutions for obtaining a kappa-type statistic for use with multiple raters. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas.

23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

24. Schnetler CM, Pettifor JM, Mesquita JM, Bird MDT, Schnaid E, Smyth AE (1993) Histomorphometry of iliac crest bone in 346 normal black and white South African adults. Bone Miner 10: 183–199

25. Kimmerle EH, Konigsberg LW, Jantz RL, Baraybar JP (2008) Analysis of age at death estimation through the use of pubic symphyseal data. J Forensic Sci 53:558–568

26. Brooks S, Suchey JM (1990) Skeletal age determination based on the os publis—a comparison of the Ascade-Nemeskeri and Suchey-Brooks methods. Hum Evol 5:227–238

27. Oettlé AC, Steyn M (2000) Age estimation from the sternal ends of ribs by phase analysis in South African blacks. J Forensic Sci 45: 1071–1079

28. Iscan MY, Steyn M (2013) The human skeleton in forensic medicine, 3rd edn. Charles C Thomas, Springfield

29. Jooste N, L'Abbé EN, Pretorius S, Steyn M (submitted) Validation of transition analysis as a method of age estimation in a modern South African sample. Med Sci Law.