JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2010, Vol. 55, No. 2
doi: 10.1111/j.1556-4029.2009.01279.x
Available online at: interscience.wiley.com

TECHNICAL NOTE

ANTHROPOLOGY

Laurent Fanton,[1,3] M.D.; Marie-Paule Gustin,[2] Ph.D.; Ulysse Paultre,[2] M.S.; Bettina Schrag,[3] M.D.; and Daniel Malicier,[1] Ph.D.

Critical Study of Observation of the Sternal End of the Right 4th Rib

ABSTRACT: Studies of the method of estimating age at death by the 4th rib exclusively concerned the phase method without fundamentally challenging the method as such. The present study analyzed observation of the variables on which the İşcan method is based. Ten observers made two assessments of the stage of pit depth, pit shape, rim and wall configurations of 59 right 4th ribs harvested from males (mean age: 49 years; range: 47–94 years). Observation showed poor reproducibility and repeatability for all three variables (Wilcoxon test, κ-coefficient). Analysis of problem ribs revealed difficulty in measuring and imprecision in describing pit depth and failure to take account of continual aging for the other two variables. Despite these results, İşcan's variables provide objective information on age at death. It is recommended that the method be improved by better description of the variables and use of multivariate statistical analysis.

KEYWORDS: forensic science, forensic anthropology, age at death, 4th rib, reproducibility, repeatability, relationship to age

The validity of scientific proof and expert testimony in forensic science has been an issue in recent years (1). More demanding legal requirements have meant that approximate scientific observations have given way to results obtained by precise reproducible methods that can be routinely applied. Forensic genetics has led the field in this transition, establishing quality assurance procedures, notably to validate analysis techniques (2). Laboratory approval procedures forced the pace in this development (3), which was also facilitated by the fact that forensic and medical genetics techniques are closely related, so that quality assurance procedures can easily be transposed from the latter to the former. Forensic anthropology faces the same new quality assurance demands (4,5) but, while there has been some methodological reflection (6,7), there are as yet no quality assurance guidelines in this field. Biological anthropology will therefore have to seek inspiration in methods employed in other fields but transposable to the forensic context (8,9).

The present study concerned one of the current reference methods for estimating age at death (10,11): observation of the sternal end of the 4th rib. This method, derived from metrology, was first proposed by İşcan in the 1980s. It consists of evaluating one quantitative variable (pit depth) and two ordinal qualitative variables (pit shape, and rim and wall configurations) (12), on which the phase method is based (13–16). These variables are each attributed six stages (0–5), and integrate assessment of wall solidity and bone quality as of stage 2 of the qualitative variables. Assessment studies have highlighted a number of problems—bias (17,18), poor reproducibility (18), and doubts regarding transposition across populations (16,19–21)—although without fundamentally challenging the method. Exclusively concerned with phase method and age estimation, they provided no analysis of the variables being used (19,22,23). A critical study of the İşcan method and its variables is reported, with a quality assurance approach. Several indications for future improvement are suggested.

Method

A forensic sample of 59 4th ribs from males of French ethnicity (mean age: 49 years; range: 47–94 years) (Fig. 1) was collected at the Institute of Forensic Medicine of Lyon during year 2004. In accordance with French law, the Public Prosecutor approved sampling during forensic autopsy, and families were informed of the nature and objectives of the study.

Ribs were macerated in water until the soft tissue and cartilage were easily detached, and then heated to 80°C in a bain-marie of water until they were completely removed.

Each rib was examined twice (sessions 1 and 2) at 2 weeks' interval, by 10 observers: one anthropologist, five forensic physicians, and four non-forensic observers. Pit depth, pit shape, and rim and wall configurations were estimated. After an explanation of the method, each observation was performed with the help of the photographs and commentaries from İşcan's original article (12). Depth was measured with calipers; the observers estimated the corresponding stage for pit shape and rim and wall configuration or, in case of doubt, the nearest two stages so as to define an intermediate stage.

Analysis first assessed reproducibility and repeatability. A given observer's reproducibility rating was determined in terms of

[1]Université de Lyon, Université Lyon 1, Institut Universitaire de Médecine Légale, Lyon F-69008, France.
[2]Université de Lyon, Université Lyon 1, Institut des Sciences Pharmaceutiques, Lyon F-69008, France.
[3]Centre Universitaire Romand de Médecine Légale, site de Lausanne, Lausanne 1011, Switzerland.

Received 30 July 2008; and in revised form 1 Dec. 2008; accepted 3 Feb. 2009.

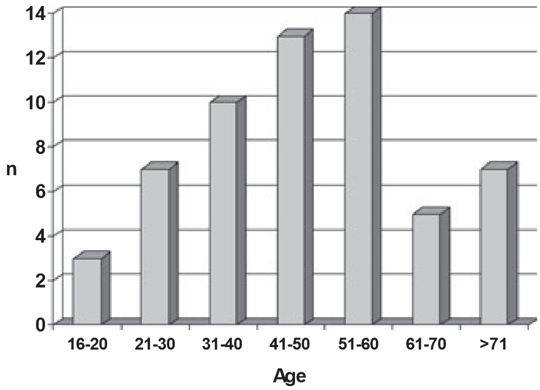

FIG. 1—*Histogram of study population (*n = *59).*

concordance with the İşcan expert's assessment, and repeatability by the agreement between sessions. Difficulties of description concerning certain particular ribs were then taken into account before going on to analyze the method's predictive value for age at death.

Statistics

Analysis of the quantitative variable (pit depth) was performed by a paired non-parametric Wilcoxon test. Qualitative variables (pit shape and rim and wall configuration) were analyzed using the kappa coefficient (κ). The relation between age and variables was studied by the Pearson correlation coefficient.

Results

Reproducibility and Repeatability of Pit Depth Assessment

Pit depth assessment reproducibility was calculated for each observer as the difference from the expert's measurements in either session. A paired non-parametric Wilcoxon test assessed the significance of these differences. Pit depth observations were found to show poor reproducibility: only two observers' estimates agreed with the expert's values at a 5% α-risk on the first session, and four on the second.

Pit depth assessment repeatability was calculated for each observer as the difference between session values, on the same statistical test. With the same significance criterion, only five observers showed sufficient inter-session repeatability. These poor results could not have been due to the measuring instrument itself, which was precise to the nearest 0.10 mm. Rather, they were mainly due to the difficulty of choosing the measurement location.

Reproducibility and Repeatability of Pit Shape and of Rim and Wall Configuration Assessment

The reproducibility and repeatability of the pit shape and rim and wall configuration observations were assessed for each observer against the expert values using the κ-coefficient on either session. This provided the following scale of agreement: 0.00–0.20, poor; 0.21–0.40, slight; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, perfect or nearly perfect (24) (Table 1).

Observations on both variables showed poor to slight reproducibility, although there was some improvement in second-session estimates for more than half the observers. Reproducibility was also found to be independent of academic status.

Inter-session repeatability was also poor to slight for all observers—and was again found to be independent of academic status.

Study of Rib Factor

Pit Depth—The scatter of pit depth readings per rib allowed certain problem ribs to be identified. Scatter was assessed by calculating the interquartile interval for all depth measurements excluding that made by the expert. Figure 2 shows that nine ribs presented a large interquartile interval (>3 mm).

On individual examination, it emerged that the difficulty of measuring the depth of these cavities was secondary to problems of identifying the point of measurement, and/or of performing the measurement. These problems consisted of: (i) the presence of several possible measurement points (Fig. 3); (ii) irregularity of the pit or rim (Fig. 4); or (iii) the presence of projections or bone windows (Fig. 5). Cartilage remaining in the bed of the joint and loss of substance due to iterative measurement may also have impaired assessment. The difficulty in determining the measurement location may itself have been dependent on individual differences in recognition skill.

Pit Shape and Rim and Wall Configuration—Agreement on pit shape and rim and wall configurations was assessed by the κ-coefficient for all measurements of each, excluding the expert's values (n = 18). Figure 6*a,b* present inter-observer agreement per rib as a function of the median of all 18 observations. The median rather than the expert's value was chosen as reference as it provided better consensus between the 18 observations.

Figure 6*a* shows poor agreement (κ < 0.20) on pit shape for 20 ribs, most of which (17 out of 20) presented median values between 3 and 4.

Figure 6*b* shows poor inter-observer agreement on rim and wall configurations in many ribs, although less than for pit shape. Again, most ribs presented median values between 3 and 4. As of stage 3, the higher the median, the greater the inter-observer agreement.

TABLE 1—*Reproducibility and repeatability of pit shape and of rim and wall configurations.*

	Pit Shape			Rim and Wall Configurations		
	Obs vs. Exp			Obs vs. Exp		
	Session 1	Session 2	Session 1 vs. Session 2	Session 1	Session 2	Session 1 vs. Session 2
Kappa min	0.06	0.12	0.11	0.16	0.14	0.10
Kappa max	0.25	0.32	0.47	0.22	0.34	0.47

Reproducibility: each observer's agreement with the İşcan expert's pit shape and rim and wall configurations values were assessed by κ-coefficient on both sessions (obs vs. exp, Session 1, Session 2). Repeatability: each observer's inter-session discrepancies in pit shape and rim and wall configurations were assessed by κ-coefficient (Session 1 vs. Session 2). Kappa min and Kappa max: minimum and maximum κ-coefficients for all observers.

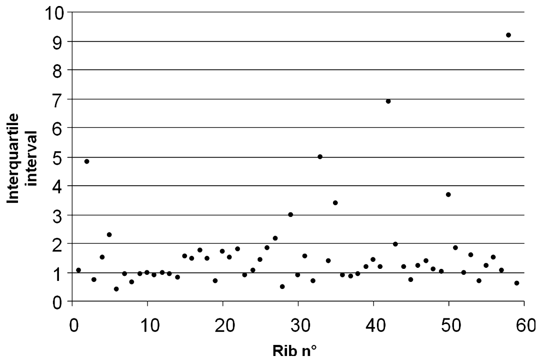

FIG. 2—*Pit depth value scatter per rib (*n = 59*). For each rib (1–59), the scatter of pit depth readings was given by the corresponding interquartile interval.*

FIG. 3—*Pit bottom irregularity.*

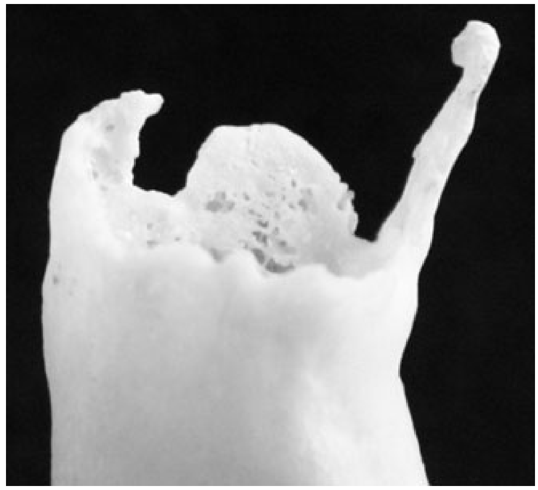

FIG. 4—*Rim irregularity.*

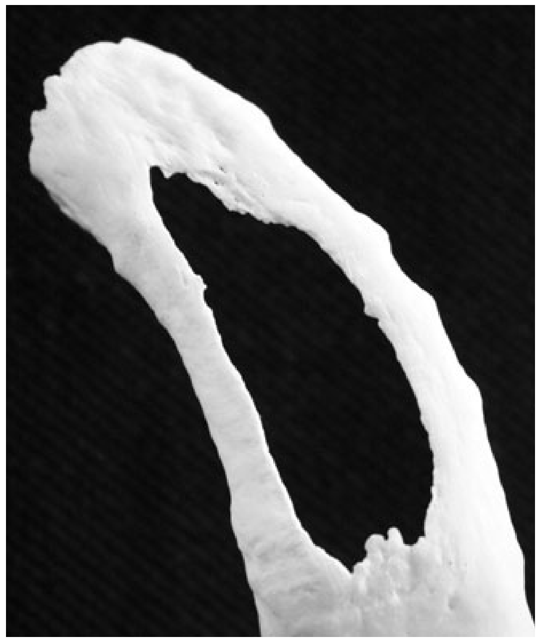

FIG. 5—*Bone windows.*

Finally, Fig. 7, while failing to disclose any statistical trend as such, rather suggests that the greater the agreement on pit form, the greater that on rim and wall configuration.

Analyzing ribs showing poor agreement ($\kappa < 0.20$) revealed ambiguities in the observation of pit shape and rim and wall configurations. Difficulty in recognizing pit shape was secondary to the imprecise descriptions in the protocol, mixing up subjective judgment of pit shape with pit depth and additional factors such as wall quality or bone texture. Our tactic of taking account of intermediate stages in case of hesitation between two stages proved insufficient, as precise specific criteria for each intermediate stage were lacking; this was especially true for stages 3 and 4. Paradoxically, the expected difficulty in recognizing stage 5 pit shape, due to the interference of bone projections, did not in fact impair consensus—perhaps because most observers, having easily attributed stage 5 to the wall configuration, simply worked on the assumption that the pit shape had to be stage 5 too. This morphological analysis also disclosed certain aspects of pit shape that could be taken account of in a more precise description: e.g., an asymmetric posterior flaring (Fig. 8).

The same kinds of difficulty as found for pit shape were encountered in recognizing rim and wall configuration. Difficulty in distinguishing between regular and irregular indentation hampered differentiation between stages 2, 3, and 4 if no other discriminatory factor was present (Fig. 9). This difficulty was aggravated by taking into account additional subjective factors shared with pit shape and by the existence of intermediate cases. In contrast, nascent posterior and/or anterior bone projections ("beaks") seemed to be discriminatory factors inducing wide consensus. It was also found that the anterior rim did not always provide the same indication as the posterior rim.

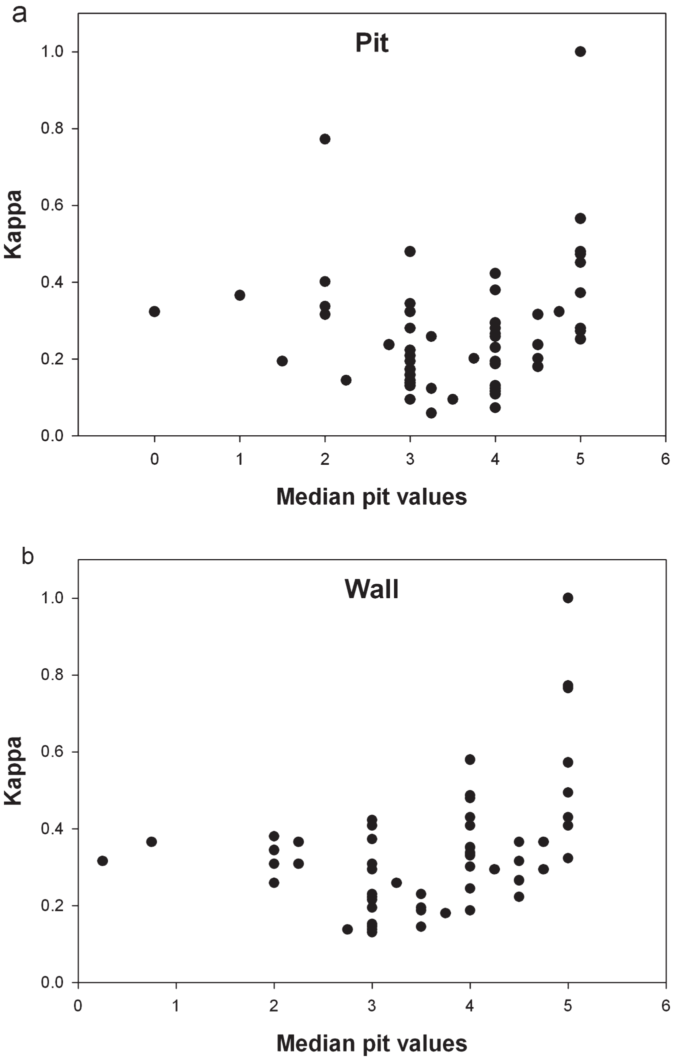

FIG. 6—*Inter-observer agreement on pit shape and rim and wall configurations. Agreement per rib (*n = *59) is given by the κ-coefficient for the 18 observations, (a) for pit shape, and (b) for rim and wall configurations. The median of the observations is taken as a priori reference value.*

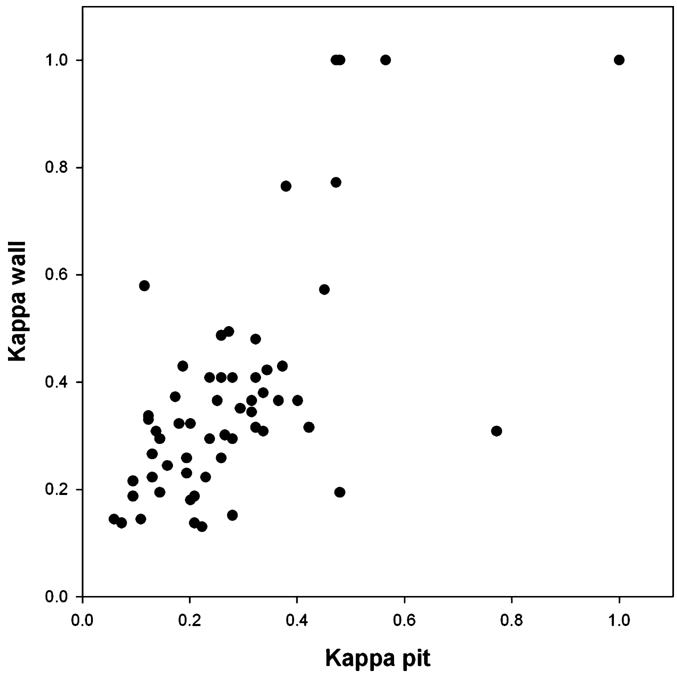

FIG. 7—*Relation between agreement on pit shape and on rim and wall configurations, per rib. Relation between inter-observer agreement on pit shape and on rim and wall configuration, per rib (*n = *59): κ-coefficients for pit shape (*x*-axis) and for rim and wall configuration (*y*-axis).*

Relationship of the Variables to Age

After examining how the method can be implemented, it seemed appropriate to look at its principles: i.e., how age relates to the variables—an essential issue if, despite their shortcomings in terms of quality, morphological observations made according to the İşcan method nevertheless provide useful information as to age at death. The method lacks precision (on an acceptability range defined by a standard deviation of 5–7 years) as of the first point of the description (5).

To assess the correlation, various analytic models (linear, second order polynomial and exponential) were implemented for different situations: first or second session expert data ($n = 59$) and the 10 observers' first or second session data ($n = 590$). As the pit shape and rim and wall configurations were ordinal and had enough possible values, they were treated as continuous. Whatever the variable, the best correlations were always found for the logarithm of age. This gave the classical result of predictive value diminishing with age. The most significant correlations emerged when observed values of the 10 observers for the second sessions were taken into account. Table 2 shows the age correlations and p-values for each variable. Pit depth consistently gave the weakest correlations.

Finally, the interest of using all three variables simultaneously was evaluated by studying their inter-relations. Taken 2 by 2, they proved dependent, partially confirming the bias due to confusion between cavity and wall descriptions. These findings prompted us to undertake a further study (presently underway) based on multivariate analysis.

Discussion

To our knowledge, this is the first study of the İşcan method inspired by a quality assurance approach. Objective analysis of the variables found reproducibility and repeatability to be poor. The difficulty of measuring pit depth was such that, as İşcan himself suggested, this variable should not be included in future studies (11). Difficulties in observing pit shape and rim and wall configurations were due to vagueness in the protocol description and the fact that continual aging was not taken into account. The problems especially affected stages 3 and 4, which partially explains the method's imprecision as of these stages. Furthermore, in case of hesitation between two stage attributions, observers tended to attribute to both variables the stage of the one which was less ambiguous, which to some extent made up for the failings of the method but at the cost of the independence of the variables.

Overall, the present findings raise the question of how best to use everyday language to convey an expert's implicit knowledge. Despite the poor results, İşcan's variables did prove to contain objective information on age at death. We therefore intend to continue this work with a view to improve the method of estimating age at death from the 4th rib. Firstly, the pit shape and rim and wall configurations need to be able to be analyzed in greater detail, and new variables need to be found by means of modern

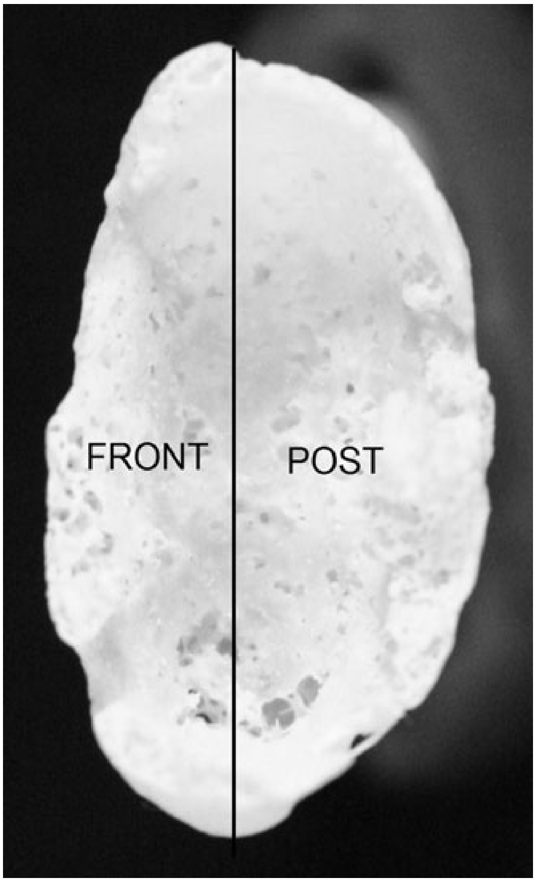

FIG. 8—*Asymmetric posterior flaring of the shape.*

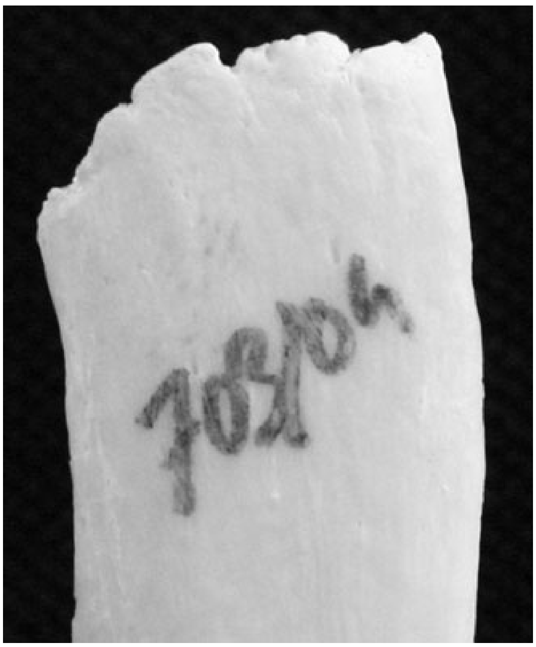

FIG. 9—*Indentation patterns.*

image-processing tools. Secondly, multivariate statistical analysis should be implemented to reveal relations between variables and between groups of variables and age, enabling not only a Bayesian but a multiple regression model. This work is underway, but first requires building up a large sample of ribs.

TABLE 2—*Correlations between variables and age at death.*

	Correlation	*p*-value
Cavity depth	0.53	$<10^{-42}$
Cavity pattern	0.71	$<10^{-90}$
Rim and wall configuration	0.73	$<10^{-99}$

Correlations were assessed by Pearson coefficients between each variable and the logarithm of age, over 10 observations and 59 ribs (n = 590). Significance was assessed by a Student's test of non-correlation.

Acknowledgments

We thank Pr. Christian Zaccharia Paultre who contributed intellectually, and Fabien Bévalot, Fernanda Freitas, Claire Desbois, Sabine Lalliard, Habdelhamid Grait, and Aissa Boudaba for data collection.

References

1. Davies J. Admissibility of scientific evidence in courts. Med Law 2005;24(2):243–57.
2. Nishimi RY, O'Connor KW, Gwin HL, Anderson MA. Genetic witness: forensic uses of DNA tests. J Int Bioethique 1991;2(1):29–32.
3. Hallenberg C, Morling N. A report of the 1997, 1998 and 1999 Paternity Testing Workshops of the English Speaking Working Group of the International Society for Forensic Genetics. Forensic Sci Int 2001; 116(1):23–33.
4. Givas CR, Komar DA. Kumho, Daubert, and the nature of scientific inquiry: implications for forensic anthropology. J Forensic Sci 2008; 53(4):771–6.
5. Ritz S, Cattaneo C, Collins MJ, Waite ER, Schütz HW, Kaatsch HJ, et al. Age estimation: the state of the art in relation to the specific demands of forensic practice. Int J Legal Med 2000;113(3):129–36.
6. Rogers TL, Allard TT. Expert testimony and positive identification of human remains through cranial suture patterns. J Forensic Sci 2004;49(2):203–7.
7. Christensen AM. The impact of Daubert: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). J Forensic Sci 2004;49(3):427–30.
8. Ritz-Timme S, Rochholz G, Schütz HW, Collins MJ, Waite ER, Cattaneo C, et al. Quality assurance in age estimation based on aspartic acid racemisation. Int J Legal Med 2000;114(1–2):83–6.
9. Schmeling A, Geserick G, Reisinger W, Olze A. Age estimation. Forensic Sci Int 2007;165(2–3):178–81.
10. Rösing FW, Graw M, Marré B, Ritz-Timme S, Rothschild MA, Rötzscher K, et al. Recommendations for the forensic diagnosis of sex and age from skeletons. Homo 2007;58(1):75–89.
11. Ubelaker DH. Issue in the global applications of methodology in forensic anthropology. J Forensic Sci 2008;53(3):606–7.
12. İşcan MY, Loth SR, Wright RK. Metamorphosis at the sternal rib end: a new method to estimate age at death in white males. Am J Phys Anthropol 1984;65(2):147–56.
13. İşcan MY, Loth SR, Wright RK. Age estimation from the rib by phase analysis: white males. J Forensic Sci 1984;29(4):1094–104.
14. İşcan MY, Loth SR. Age estimation from the rib by phase analysis: white females. J Forensic Sci 1985;30(3):853–63.
15. İşcan MY, Loth SR. Determination of age from the sternal rib in white males: a test of the phase method. J Forensic Sci 1986;31(1):122–32.
16. İşcan MY, Loth SR. Determination of age from the sternal rib in white females: a test of the phase method. J Forensic Sci 1986;31(3):990–9.
17. Russell KF, Simpson SW, Genovese J, Kinkel MD, Meindl RS, Lovejoy CO. Independent test of the fourth rib aging technique. Am J Phys Anthropol 1993;92(1):53–62.
18. Dudar JC, Pfeiffer S, Saunders SR. Evaluation of morphological and histological adult skeleton age-at-death estimation techniques using ribs. J Forensic Sci 1993;38(3):677–85.
19. Kimmerle EH, Prince DA, Berg GE. Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth. J Forensic Sci 2008;53(3):594–600.

20. İşcan M, Loth SR, Wright RK. Racial variation in the sternal extremity of the rib and its effect on age determination. J Forensic Sci 1987;32(2):452–66.
21. Oettlé AC, Steyn MS. Age estimation from sternal ends of ribs by phase analysis in South African Blacks. J Forensic Sci 2000;45(5):1071–9.
22. Baccino E, Ubelaker DH, Hayek LAC, Zerilli Z. Evaluation of seven methods of estimating age at death from mature human skeletal remains. J Forensic Sci 1999;44(5):931–6.
23. Martrille L, Ubelaker DH, Cattaneo C, Seguret F, Tremblay M, Baccino E. Comparison of four skeletal methods for the estimation of age at death on white and black adults. J Forensic Sci 2007;52(2):302–7.
24. Kapucuoglu N, Bulbul D, Tulunay G, Temels MA. Reproductibility of grading systems for endometrial endometrioid carcinoma and their relation with pathologic prognostic parameters. Int J Gynecol Cancer 2008;18(4):790–6.

Additional information and reprint requests:
Laurent Fanton, M.D.
Institut Universitaire de Médecine Légale
de Lyon Université Claude-Bernard Lyon 1
12 avenue Rockefeller
Lyon 69008
France
E-mail: laurent.fanton@chu-lyon.fr