Assessing exact randomization-based methods for determining the taxonomic significance of variability in the human fossil record

L.C. Aiello[a*], M. Collard[a], J.F. Thackeray[b] and B.A. Wood[c]

The identification of fossil hominin species is one of the principal challenges in human palaeontology. Here, we use metrical data from extant primate species and Plio-Pleistocene hominin specimens to evaluate four exact randomization-based methods for assessing specific diversity in the fossil record. The first method is based on size dimorphism, the second on average taxonomic distance, the third on size-corrected average taxonomic distance, and the fourth on the standard error of the slope in least squares regression. The study examines how the methods compare with respect to the size and/or shape information they recover, and how their conclusions regarding the specific affinities of hominin specimens are affected by different comparative samples. The study also investigates how the methods are affected by any overlap between intra- and interspecific variability within the reference samples, and how they compare with respect to the fossil specimens they suggest should be considered to be conspecific. The outcome of these analyses suggests that the results of recent exact randomization analyses of Plio-Pleistocene hominin specific diversity should be interpreted with caution.

The reconstruction of human evolutionary history must be based on reliable hypotheses about the origin, nature and fate of species groups.[1-3] As such, the identification of species in the human fossil record is central to palaeoanthropology. Conventionally, the specific composition of fossil hominin[†] samples has been assessed in relation to the *possibility* of observing comparable size and shape variation in extant reference species that are closely related to the fossils, most notably *Homo sapiens, Pan troglodytes, Pongo pygmaeus* and *Gorilla gorilla.*[3-6] Researchers, in other words, have sought to determine whether the size and shape discrepancies exhibited by the fossils in a sample are ever observed in what are believed to be appropriate extant comparator species, and have assigned the fossils to one or more species accordingly.

The possibility-orientated methods have been supplemented, in recent years, with methods that use exact randomization to establish the *probability* of sampling in extant reference species the size and/or shape differences observed among the fossil specimens.[7-13] These methods proceed by recording several measurements on a sample of extant primates and a number of fossil specimens. Next, a distance statistic (e.g., average taxonomic distance) is computed for every possible pair of extant conspecifics, and the resulting distances plotted as a frequency histogram. The distance between every possible pair of fossil specimens is then calculated, and these distances are superimposed on the histogram for the extant species. Lastly, the conspecificity, or otherwise, of the fossil specimens is considered in relation to the distances recorded for the extant species. Conventionally, two fossils are considered to be conspecific if the distance between them is less than, or equal to, those recorded for 95% of the randomly sampled pairs of extant specimens.

Applications of exact randomization methods in palaeoanthropology have, to date, yielded results that are difficult to interpret in relation to each other, and to those obtained using the possibility-orientated methods. Some are consistent with the consensus taxonomy for the Plio-Pleistocene hominins.[9,10] Others have suggested that taxonomic differentiation should probably be increased.[8] Still others have supported a significant reduction in the number of species recognized in the early hominin fossil record.[13]

In this paper, we use data from extant primates species and Plio-Pleistocene fossil hominin specimens to assess four exact randomization methods. The questions addressed in the study are:

1. How do the methods compare with respect to the size and/or shape information they recover?
2. How are their conclusions regarding the specific affinities of hominin specimens affected by different comparative samples?
3. How are the methods affected by any overlap between the intra- and interspecific variability within their reference samples?
4. How do they compare with respect to the fossil specimens they suggest should be considered to be conspecific?

Materials and methods

The study is based on 20 craniodental variables recorded on eight extant non-human primate species and seven fossil hominin specimens. The measurements are listed in Table 1. The non-human primate data are taken from Collard.[14] The non-human primate sample comprises 19 *Colobus badius,* 8 *Erythrocebus patas,* 25 *Macaca fascicularis,* 26 *Mandrillus leucophaeus,* 15 *Pan troglodytes,* 15 *Papio anubis,* 14 *Papio cynocephalus* and 24 *Theropithecus gelada,* making a total of 146 specimens.

The fossil hominin data are taken from Wood[5]. The hominin sample consists of two crania that are usually assigned to *Paranthropus boisei,* KNM-ER 406 and OH 5, one *Australopithecus africanus* cranium, Sts 5, and two specimens that are assigned either to *Homo habilis*[15] or to *Australopithecus habilis,*[16,17] KNM-ER 1813 and OH 24. The fossil sample also includes KNM-ER 3733

[a]Department of Anthropology, University College London, Gower Street, London WC1E 6BT, U.K. E-mail: l.aiello@ucl.ac.uk, m.collard@ucl.ac.uk

[b]Department of Palaeontology, Transvaal Museum, P.O. Box 413, Pretoria, 0001 South Africa. E-mail: mrsples@global.co.za

[c]Henry R. Luce Professor of Human Origins, Department of Anthropology, George Washington University, 2110 G Street NW, Washington DC 20052, U.S.A. E-mail: bwood@gwu.edu

*Author for correspondence. E-mail: l.aiello@ucl.ac.uk

[†]Hominin is used here in preference to hominid to denote the tribe comprising anatomically modern humans and the fossil species that are more closely related to them than to any other species with extant members.

Table 1. Measurements used in the study.

Measurement	Reference number in Wood[5]
Glabella-opisthocranion	1
Minimum frontal breadth	8
Biporionic breadth	11
Parietal sagittal chord	25
Occipital sagittal length	39
Superior facial height	43
Superior facial breadth	49
Anterior interorbital breadth	55
Orbital height	57
Minimum malar height	59
Maximum nasal aperture width	68
Nasal height	69
Sagittal length of nasal bones	71
Superior breadth of the nasal bones	73
Inferior breadth of the nasal bones	74
Maxillo-alveolar breadth	87
Palatal height at M^1	95
Upper premolar alveolar length	96
Interalveolar distance at P^4	100
Interalveolar distance at M^2	101

(early African *Homo erectus* or *Homo ergaster*[16,17]) and KNM-ER 1470 (*Homo rudolfensis* or *Australopithecus rudolfensis*[16,17])‡.

Taxonomic methods

The methods we examine are based on four measures of morphological similarity: size dimorphism ratio (DIM), average taxonomic distance (ATD), size-corrected average taxonomic distance (GATD), and the standard error of the slope in least squares regression (SEEB). DIM is the ratio of the geometric mean of one specimen to that of another.[9] The geometric mean is one of the Mosimann family of size variables, and is calculated as the nth root of the product of n measurements. ATD is equivalent to the actual distance by which two specimens would be separated if they were plotted in a multidimensional space in which each variable is represented by an axis perpendicular to the axes representing the other variables.[9] It is calculated as the square root of the mean of the sum of the squares of the differences between the values for n characters. GATD is calculated in the same way as the ATD except that the values for each variable are first divided by the geometric mean of all the variables for the appropriate specimen.[9] SEEB is derived from the values of n measurements recorded on a pair of specimens when they are regressed against each other using least squares regression.[13] The $\log_{10}$ transformed standard error is used in the analysis.

Distances based on these measures are computed for every possible pair of reference specimens (both conspecifics and members of different species) and every possible pair of fossil specimens. Histograms are derived from the four sets of distances, with all intraspecific comparisons presented by one set of bars (black) and all interspecific comparisons by another (white) (Figs 1–4). The 95% confidence limits (CLs) for both the intra- and interspecific variation are determined empirically. Lastly, distances for the pairs of fossil specimens are compared to the CLs for the reference species.

Size and shape information recovered

The size-corrected average taxonomic distance and the standard error of the slope in least squares regression are highly

‡Copies of the dataset are available on request from M.C.

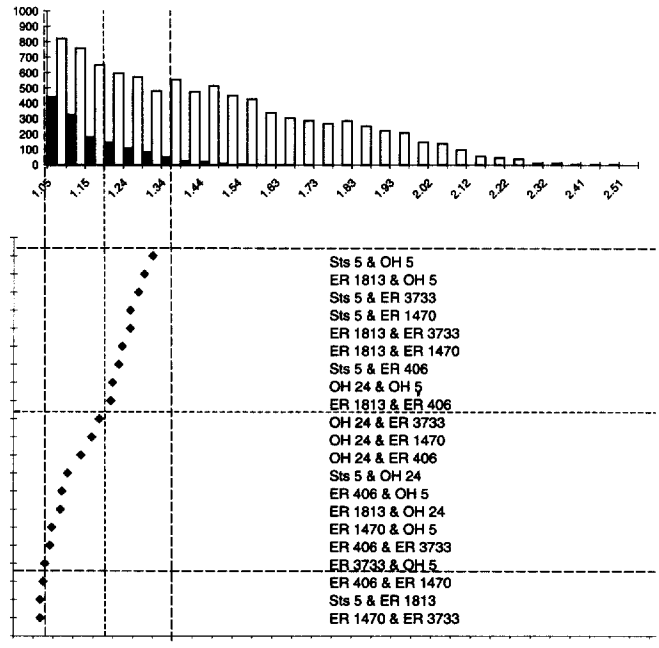

Fig. 1. Size dimorphism (DIM). The upper graph represents the size dimorphism distribution obtained from the pair-wise comparisons of the reference sample (black bars = intraspecific comparisons, white bars = interspecific comparisons). The lower graph shows the range of size dimorphism obtained from the pair-wise comparisons of the seven fossil hominin specimens. The dotted line represents the upper 95% CL for variation in *P. troglodytes*. The dashed line on the right represents the upper 95% CL for intraspecific variation in the total reference sample, while the dashed line on the left represents the lower 95% CL for interspecific variation. The fossil pairs whose size dimorphism differences are greater than would be expected for *P. troglodytes* are found to the right of the dotted line, while those pairs whose disparities are greater than would be expected for any single species comparison in the entire reference sample fall to the right of the right-hand dashed line. Those pairs whose disparities are less than would be expected for any interspecific pair in the combined reference sample fall to the left of the left-hand dashed line. (ER = KNM-ER.)

correlated with each other, as are size dimorphism and average taxonomic distance (Table 2). The strong correlation between these pairs of distance measures and the weaker cross-correlation between measures from each pair indicate that they are measuring essentially separate aspects of cranial metric variability (see also ref. 9). The DIM method is based on one of the Mosimann family of size variables and the strong correlation between the distances obtained using the DIM and ATD methods indicates that the latter is biased towards the recovery of information about size rather than shape. Conversely, the GATD method adjusts for the effects of size difference between specimens. By dividing the individual measurements by the geometric mean of all measurements, crania that are geometrically larger or smaller than each other will appear identical.[18] In this sense GATD removes size variation from the analysis and compares crania on the basis of their geometric shape similarity to each other.

The strong correlation between the GATD and SEEB methods suggests that SEEB also retrieves mainly shape information. The correlation is less than perfect, however. The SEEB method reflects both geometric and allometric shape similarity between crania rather than just geometric shape similarity. The statistic reflects the average deviation of the individual measurements from a least squares regression line fit to the bivariate scatter of pair-wise measurements of one specimen plotted against the measurements from a second specimen. If the crania are either geometrically or allometrically identical the points will lie on a

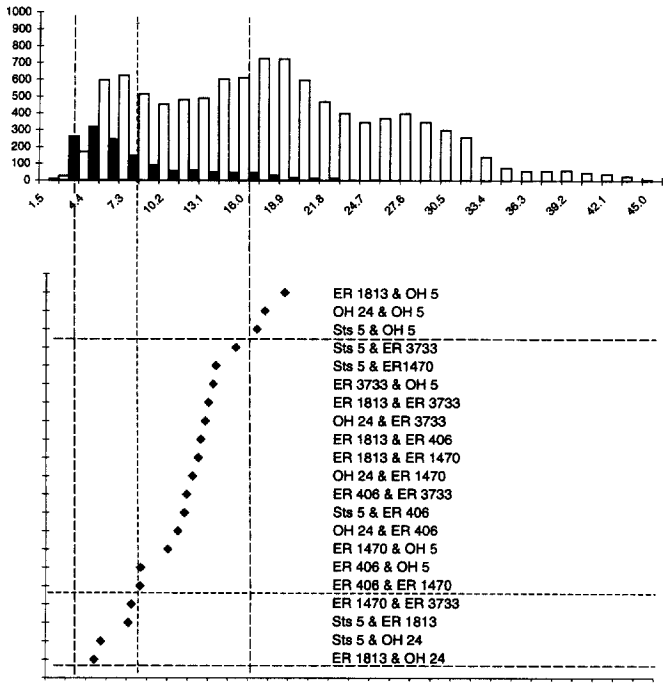

Fig. 2. Average taxonomic distance (ATD). The upper graph represents the ATD distribution obtained from the pair-wise comparisons of the reference sample. The lower graph shows the range of ATD obtained from the pair-wise comparisons of the seven fossil hominin specimens. Further explanation as given in Fig. 1.

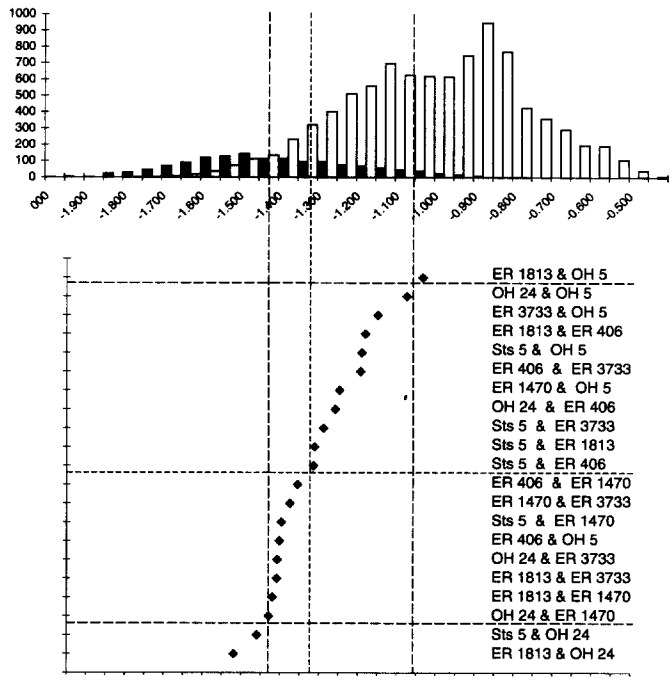

Fig. 4. Log (base 10) standard error of the least squares slope (SEEB). The upper graph represents the SEEB distribution obtained from the pair-wise comparisons of the reference sample. The lower graph shows the range of SEEB obtained from the pair-wise comparisons of the seven fossil hominin specimens. Further explanation as in Fig. 1.

straight line and there will be no deviation from the regression line. The greater the geometric or allometric shape difference between the crania, the further the points will lie from the regression line and the larger the SEEB value.

Specific affinities of hominin specimens

In Figs 1–4, the taxonomic hypotheses for all pairs of fossil hominin specimens based on the pattern and degree of variation in the *Pan troglodytes* sample using the four taxonomic distance methods are compared with the equivalent hypotheses based on the variation in the total reference sample. It is clear from these figures that assessments of the probability of fossil conspecificity are heavily dependent on the composition of the reference sample. For example, when *P. troglodytes* is used as the comparator, the SEEB method (Fig. 4) suggests that 10 of the 21 hominin fossil pairs are compatible with the single-species hypothesis. However, when the total reference sample is used as the comparison, the SEEB method suggests that 20 of the fossil comparisons are consistent with the single-species hypothesis. Likewise, the impact of changing the composition of the reference sample differs from one method to another. Thus, when the full comparative sample is used, the number of pairs of fossil hominins that are compatible with the single-species hypothesis increases from 4 to 18 with the ATD method, from 10 to 20 with the SEEB method, from 12 to 21 with the DIM method and from 13 to 19 with the GATD method.

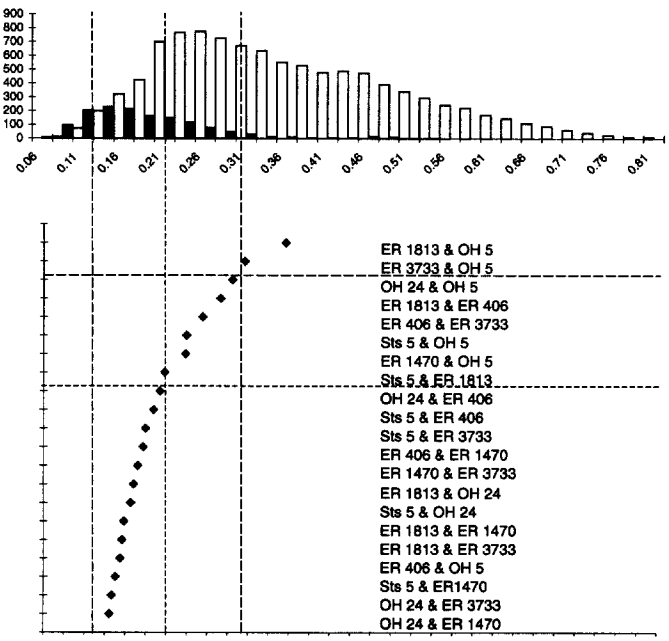

Fig. 3. Size-corrected average taxonomic distance (GATD). The upper graph represents the GATD distribution obtained from the pair-wise comparisons of the reference sample. The lower graph shows the range of GATD obtained from the pair-wise comparisons of the seven fossil hominin specimens. Further explanation as given in Fig. 1.

Intra- and interspecific variability in reference sample

The distances between the fossil hominin pairs are shown in Figs 1–4, together with the distributions of the intra- and

Table 2. Pearson correlation coefficients for the four exact randomization methods.

	GATD	ATD	DIM	SEEB
GATD	1.000	0.523*	0.084	0.954**
ATD	0.692**	1.000	0.758**	0.698**
DIM	0.441**	0.884**	1.000	0.224
SEEB	0.926**	0.838**	0.651**	1.000

Values below the diagonal are correlations based on 10 440 pair-wise comparisons between the reference specimens. Those above the diagonal are correlations based on 21 pair-wise comparisons between the fossil specimens. GATD = size-corrected average taxonomic distance. ATD = average taxonomic distance. DIM = size dimorphism ratio. SEEB = standard error of the slope in least squares regression.
*P = 0.05 (two-tailed). **P = 0.01 (two-tailed).

Table 3. Summary of the results of assessing the status of pairs of fossil hominin crania.

	DIM	ATD	GATD	SEEB
Sts 5 and KNM-ER 1470				
KNM-ER 1813 and KNM-ER 1470				
OH 24 and KNM-ER 1470				
KNM-ER 1813 and OH 5		Different	Different	Different
OH 24 and OH 5		Different		
KNM-ER 3733 and OH 5			Different	
Sts 5 and OH 5		Different		
Sts 5 and KNM-ER 3733				
KNM-ER 1813 & KNM-ER 3733				
Sts 5 and KNM-ER 406				
KNM-ER 1813 and KNM-ER 406				
OH 24 and KNM-ER 3733				
OH 24 and KNM-ER 406				
Sts 5 and OH 24				Conspecific
KNM-ER 406 and OH 5				
KNM-ER 1813 and OH 24				Conspecific
KNM-ER 1470 and OH 5				
KNM-ER 406 and KNM-ER 3733				
KNM-ER 406 and KNM-ER 1470	Conspecific			
Sts 5 and KNM-ER 1813	Conspecific			
KNM-ER 1470 and KNM-ER 3733	Conspecific			

Conspecific = unambiguous conspecific status, i.e., outside the 95% CLs for interspecific variation based on the combined catarrine sample. Different = unambiguous different species, i.e., outside the 95% CLs for intraspecific variation based on the combined reference sample. Blank cells = status cannot be determined unambiguously. KNM-ER = Kenya National Museums, East Rudolf. OH = Olduvai hominid. Sts = Sterkfontein.

interspecific distances within and among the comparative samples. The area between the two dashed vertical lines in the lower graphs delineates the zone of overlap between the upper 95% CL for the intraspecific comparisons and the lower 95% CL for the interspecific comparisons in the total reference sample. Points falling to the left of this area represent pairs of hominin fossils that are unambiguously compatible with the single-species hypothesis, while points falling to the right of the area identify fossil pairs that are unambiguously compatible with the multiple species hypothesis. It is evident from the figures that none of the methods provides a clear separation between intra- and interspecific variation in the comparative sample. No matter which of the four methods is used, most of the 21 fossil hominin comparisons fall in the central zone. Within this zone the fossil pairs are as likely to be conspecific as they are to belong to separate species. While the methods are not equally affected by the overlap between intra- and interspecific variability among the reference species, the differences are marginal. For example, the percentage of the fossil comparisons falling in the zone of intraspecific/interspecific overlap is 90% for the GATD method, and 86% using the ATD method.

Conspecificity of fossil specimens

It is evident from Figs 1–4 that there is little agreement among the methods as to which fossil specimens should probably be considered conspecific and which should probably be assigned to separate species; these results are summarized in Table 3. Only one pair of fossils, KNM-ER 1813 (*H.* or *A. habilis*) and OH 5 (*P. boisei*), are recognized as most probably representative of two species by more than one of the methods, and no fossil pair is identified as most probably conspecific by more than one method. Consequently, these results indicate that, for the most part, any taxonomic conclusions are 'method specific'. The results also indicate that, even when the techniques are highly correlated (such as SEEB and GATD, and DIM and ATD), they do not necessarily support the same taxonomic hypotheses. This point is illustrated in Fig. 5, which shows the highly correlated GATD and SEEB distance statistics for the fossil sample. In spite of the high correlation between these distance measures (r = 0.93 for the fossil sample and r = 0.95 for the extant sample, $P <$ 0.01), the 95% CLs for each measure cut the correlated scatter at different points, thereby producing different hypotheses of conspecificity for the pairs of hominin fossils.

Discussion

The four methods for assessing specific diversity in the hominin fossil record can be divided into those that primarily recover size information (DIM and ATD) and those that mainly retrieve shape information (SEEB and GATD). It is worthy of note that the SEEB (standard error of the slope in least squares regression) method, which has recently been advocated by Thackeray and his co-workers,[13] is significantly correlated with the GATD (size-corrected average taxonomic distance) method favoured by Richmond and Jungers.[9] It is also noteworthy that,

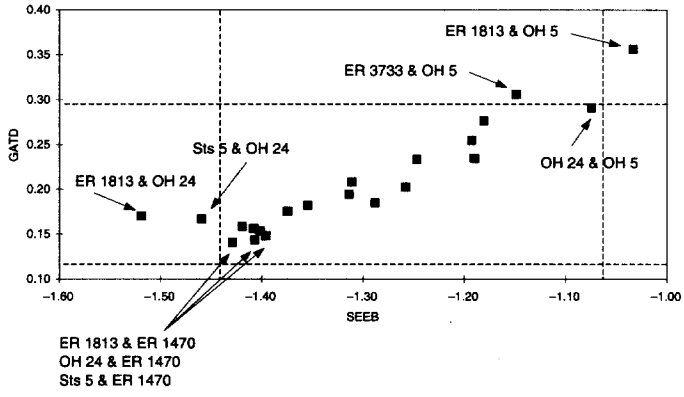

Fig. 5. The size-corrected average taxonomic distance (GATD) is strongly correlated with the log (base 10) standard error of the least squares regression slope (SEEB) (r = 0.93 for the fossil sample, n = 21). The vertical lines represent the lower 95% CL for the intraspecific comparisons and the upper 95% CL for the interspecific comparisons for SEEB. The horizontal lines represent the same confidence limits for GATD. Note that the 95% CLs cut the correlated scatter at different points, thereby producing different hypotheses of conspecificity for the fossil pairs.

in spite of this strong correlation, the hypotheses of hominin conspecificity suggested by these methods are substantially different. This is because the two methods reflect shape similarity in different ways. The GATD method only recognizes geometric shape similarity, whereas the SEEB method recognizes both geometric and allometric shape similarity. These results leave no doubt that taxonomic hypotheses based on exact randomization procedures are specific to the distance measure employed. As such, the use of a particular distance measure should be justified before an analysis is carried out.

Three additional implications emerge from this study. First, taxonomic hypotheses arising from the application of any of the distance measures are sensitive to the reference species used. It is usual to compare the size and shape variation observed among the early hominins with the intraspecific size and shape variation exhibited by their closest relatives, the extant hominoids. However, is propinquity the only relevant criterion for selecting extant comparators in taxonomic analyses of the early hominins? It could be argued that the restricted geographic distribution of the extant great apes adversely affects their appropriateness as models for early hominin species, given that early hominin fossils have been found as far apart as central and southern Africa.[19] Might a more widely distributed species, such as the savanna baboon, *Papio hamadryas*, provide a more appropriate model?

Second, it is clear that there is considerable overlap in the ranges of intra- and interspecific variation in the reference sample. This overlap occurs irrespective of the method employed for assessing variation. In the past, it has frequently been assumed that the null hypothesis for analyses of the number of species represented in a fossil sample should be that the specimens are conspecific.[4,13,20–23] However, given that some extant species do not differ morphologically while others are markedly polymorphic, it is perhaps more appropriate to employ T.C. Chamberlain's[24] method of multiple working hypotheses and also test the hypothesis that two fossil specimens belong to different species.

Third, many of the fossil pairs for which the hypothesis of conspecificity cannot be rejected are traditionally recognized as members of different species. For example, based on the currently preferred technique of size-corrected average taxonomic distance (GATD), the hypothesis of conspecificity cannot be rejected for either KNM-ER 1813 (*H.* or *A. habilis*) and KNM-ER 3733 (*H. ergaster*) or KNM-ER 3733 and OH 24 (*H.* or *A. habilis*). Likewise, the single-species hypothesis cannot be rejected for KNM-ER 3733 by comparison with Sts 5 (*A. africanus*). This suggests that the metrical data providing the basis for these analyses are not sensitive to the morphological features that have been used for traditional taxonomic assessment. There is an urgent need to devise methods for determining taxonomic affinity that do have the appropriate level of sensitivity.[25]

In summary, this study highlights several previously unrecognized shortcomings in the way exact randomization methods have been used to assess the specific diversity of samples of early fossil hominins. Lockwood and his co-workers[26] have criticized exact randomization methods in relation to the theory underpinning the sampling procedure. This study demonstrates that exact randomization results are dependent on both the distance statistic used and the taxon, or taxa, used as the referent. Furthermore, it emphasizes the necessity to test the hypothesis that two fossil specimens are conspecific and the postulate that the specimens belong to different species. The study also casts doubt on the sensitivity of the metrical data in relation to its ability accurately to reflect actual taxonomic differences. Thus, the findings of recent studies based on these methods should be interpreted with caution. There remains considerable scope for refining the data and methods that form the basis of attempts to assess hominin specific diversity.

For access to the specimens in their care, we thank the staff of the Anthropologisches Institut und Museum der Universität Zürich, Switzerland, the Department of Human Anatomy and Cell Biology, the University of Liverpool, U.K., the Muséum d'Histoire Naturelle, Geneva, Switzerland, the Muséum d'Histoire Naturelle, Paris, France, the Museum für Naturkunde, Humboldt-Universität zu Berlin, Germany, and the Natural History Museum, London. We also thank A.T. Chamberlain for allowing us to use his data, and C. Lockwood and several anonymous reviewers for commenting on earlier versions of the manuscript. This research was supported, in part, by The Wellcome Trust (M.C.) and The Henry Luce Foundation (B.A.W.). The ideas discussed in this paper were first presented as a poster at the Dual Congress of the International Association for the Study of Human Palaeontology and the International Association of Human Biologists held at Sun City, South Africa, 28 June–4 July, 1998.

Received 15 September 1999. Accepted 9 March 2000.

1. Tattersall I. (1986). Species recognition in human paleontology. *J. Hum. Evol.* **15**, 165–175.
2. Tattersall I. (1992). Species concepts and species identification in human evolution. *J. Hum. Evol.* **22**, 341–349.
3. Wood B.A. (1992). Early hominid species and speciation. *J. Hum. Evol.* **22**, 351–365.
4. Kimbel W.H. and White T.D. (1988). Variation, sexual dimorphism and the taxonomy of *Australopithecus*. In *Evolutionary History of the 'Robust' Australopithecines*, ed. F.E. Grine, pp. 175–192. Aldine de Gruyter, New York.
5. Wood B.A. (1991). *Koobi Fora Research Project Volume 4: Hominid Cranial Remains*. Clarendon Press, Oxford.
6. Kimbel W.H., Johanson D.C. and Rak Y. (1994). The first skull and other new fossil discoveries of *Australopithecus afarensis* at Hadar, Ethiopia. *Nature* **368**, 449–451.
7. Grine F.E., Demes B., Jungers W.L. and Cole T.M. (1993). Taxonomic affinity of the early *Homo* cranium from Swartkrans, South Africa. *Am. J. Phys. Anthropol.* **92**, 411–426.
8. Grine F.E., Jungers W.L. and Schultz J. (1996). Phenetic affinities among early *Homo* crania from East and South Africa. *J. Hum. Evol.* **30**, 189–225.
9. Richmond B.G. and Jungers W.L. (1995). Size variation and sexual dimorphism in *Australopithecus afarensis* and living hominoids. *J. Hum. Evol.* 29, 229–245.
10. Kramer A., Donnelly S.M., Kidder J.H., Ousley S.D. and Olah S.M. (1995). Craniometric variation in large-bodied hominoids: testing the single-species hypothesis for *Homo habilis*. *J. Hum. Evol.* **29**, 443–462.
11. Lague M.R. and Jungers W.L. (1996). Morphometric variation in Plio-Pleistocene hominid distal humeri. *Am. J. Phys. Anthropol.* **101**, 401–427.
12. Thackeray J.F. (1997). Probabilities of conspecificity. *Nature* **390**, 30–31.
13. Thackeray J.F., Bellamy C.L., Bellars D., Bronner G., Bronner L., Chimimba C., Fourie H., Kemp A., Kruger M., Plug I., Prinsloo S., Toms R., Van Zyl A.J. and Whiting M.J. (1997). Probabilities of conspecificity: application of a morphometric technique to modern taxa and fossil specimens attributed to *Australopithecus* and *Homo*. *S. Afr. J. Sci.* **93**, 195–196.
14. Collard M. (1998). *Morphological evolution of the hominoids and papionins: implications for palaeoanthropological cladistics*. Ph.D. thesis, University of Liverpool.
15. Tobias P.V. (1991). *Olduvai Gorge: the Skulls, Endocasts and Teeth of* Homo habilis. Cambridge University Press, Cambridge.
16. Wood B.A. and Collard M. (1999). The human genus. *Science* **284**, 65–71.
17. Wood B.A. and Collard M. (1999). The changing face of genus *Homo*. *Evol. Anthropol.* **8**, 195–207.
18. Jungers W.L., Falsetti B. and Wall C.E. (1995). Shape, relative size, and size-adjustments in morphometrics. *Ybk Phys. Anthropol.* **38**, 137–162.
19. Brunet M., Beauvilain A., Coppens Y., Heintz E., Moutaye A.H.E. and Pilbeam D. (1995). The first australopithecine 2500 kilometres west of the Rift Valley (Chad). *Nature* **378**, 273–275.
20. Martin L.B. and Andrews P. (1984). The phyletic position of *Graecopithecus freybergi* Koenigswald. *Cour. Forsch. Inst. Senckenberg* **69**, 25–40.
21. Martin L.B. and Andrews P. (1993). Species recognition in Middle Miocene hominoids. In *Species, Species Concepts and Primate Evolution*, eds W.H. Kimbel and L.B. Martin, pp. 393–427. Plenum Press, New York..
22. Wood B.A., Li Y. and Willoughby C. (1991). Intraspecific variation and sexual dimorphism in cranial and dental variables among higher primates and their bearing on the hominid fossil record. *J. Hum. Evol.* **174**, 185–205.
23. Kramer A. (1993). Human taxonomic diversity in the Pleistocene: does *Homo erectus* represent multiple hominid species? *Am. J. Phys. Anthropol.* **91**, 161–171.
24. Chamberlain T.C. (1965). The method of multiple working hypotheses. *J. Geology* **5**, 837–848.
25. Aiello L.C., Wood B.A., Key C. and Lewis M. (1999). Morphological and taxonomic affinities of the Olduvai ulna (OH 36). *Am. J. Phys. Anthropol.* **109**, 89–110.
26. Lockwood C.A., Richmond B.G., Jungers W.L. and Kimbel W.H. (1996). Randomization procedures and sexual dimorphism in *Australopithecus afarensis*. *J. Hum. Evol.* **31**, 537–548.