James C. M. Ahern
Department of Anthropology, University of Wyoming, Laramie, Wyoming 82071-3431, U.S.A. E-mail: jahern@uwyo.edu

Sang-Hee Lee
Department of Anthropology, University of California, Riverside, California 92521-0418, U.S.A. E-mail: sang-hee.lee@ucr.edu

John D. Hawks
Department of Anthropology, University of Wisconsin, Madison, Madison, WI 53706, U.S.A. E-mail: jhawks@wisc.edu

Received 29 January 2002 Revision received 11 June 2002 and accepted 14 June 2002

Keywords: Vindija, Neandertal, Krapina, human evolution, South-central Europe, sample bias, bootstrap.

The late Neandertal supraorbital fossils from Vindija Cave, Croatia: a biased sample?

The late Neandertal sample from Vindija (Croatia) has been described as transitional between the earlier Central European Neandertals from Krapina (Croatia) and modern humans. However, the morphological differences indicating this transition may rather be the result of different sex and/or age compositions between the samples. This study tests the hypothesis that the metric differences between the Krapina and Vindija supraorbital samples are due to sampling bias. We focus upon the supraorbital region because past studies have posited this region as particularly indicative of the Vindija sample's transitional nature. Furthermore, the supraorbital region varies significantly with both age and sex.

We analyzed four chords and two derived indices of supraorbital torus form as defined by Smith & Ranyard (1980, *Am. J. phys. Anthrop.* **93**, pp. 589–610). For each variable, we analyzed relative sample bias of the Krapina and Vindija samples using three sampling methods. In order to test the hypothesis that the Vindija sample contains an over-representation of females and/or young while the Krapina sample is normal or also female/young biased, we determined the probability of drawing a sample of the same size as and with a mean equal to or less than Vindija's from a Krapina-based population. In order to test the hypothesis that the Vindija sample is female/young biased while the Krapina sample is male/old biased, we determined the probability of drawing a sample of the same size as and with a mean equal or less than Vindija's from a generated population whose mean is halfway between Krapina's and Vindija's. Finally, in order to test the hypothesis that the Vindija sample is normal while the Krapina sample contains an over-representation of males and/or old, we determined the probability of drawing a sample of the same size as and with a mean equal to or greater than Krapina's from a Vindija-based population. Unless we assume that the Vindija sample is female/young and the Krapina sample is male/old biased, our results falsify the hypothesis that the metric differences between the Krapina and Vindija samples are due to sample bias.

Journal of Human Evolution (2002) **43,** 419–432
doi:10.1006/jhev.2002.0586

Introduction

Northwestern Croatia has yielded many significant Paleolithic finds. Among these, the Neandertal fossils from Vindija Cave and Krapina Rockshelter have played a particularly important role in the debate surrounding Neandertal evolution and the origin of modern Europeans. Recent ESR and U-series analyses by Rink *et al.* (1995, 1999) date the Krapina deposits to approximately 130 ka B.P., while the Vindija fossils represent perhaps the most recent Neandertals in Central Europe. The uppermost Neandertal bearing level at Vindija, G_1, has two AMS radiocarbon dates of 28,020 ± 360

0047–2484/02/030419+14$35.00/0

and 29,080 ± 400 years B.P. (Smith *et al.*, 1999), although most of the Vindija fossils are from level G_3, which is 38,000 to 46,000 years old (Smith *et al.*, 1985).

Many morphological features place the Vindija fossils as intermediate between the Krapina Neandertals and modern Europeans (Smith & Raynard, 1980; Wolpoff *et al.*, 1981; Smith, 1982, 1984; Smith *et al.*, 1989; Ahern & Smith, 1993; Frayer *et al.*, 1993). These include reduced midfacial prognathism, reduced nasal breadth, thinner cranial vaults, development of incipient chins, and reduction and shape changes of the supraorbital torus in a manner that converges on the early modern European condition (Smith, 1984). Taken with the evidence for the persistence of Neandertal features in early modern Europeans (Szilvássy *et al.*, 1987; Frayer, 1992; Smith, 1992; Frayer *et al.*, 1993; Duarte *et al.*, 1999; Wolpoff *et al.*, 2001), the intermediacy of the Vindija sample between earlier Neandertals and early modern Europeans has been posited as evidence of Neandertal genetic contribution to early modern European populations. However, Stringer & Bräuer (1994; Bräuer, 1989) have reasonably argued that the more modern-like aspects of the Vindija specimens might be due to an over-representation of females and/or young individuals in the Vindija sample.

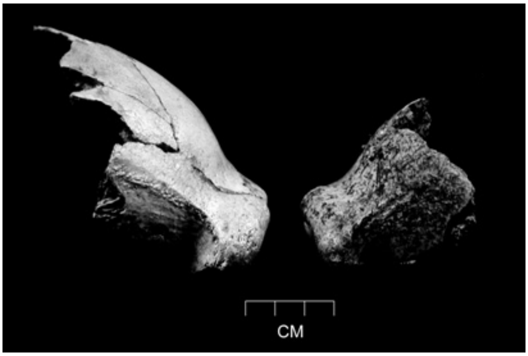

Figure 1. The fragmentary nature of the Krapina and Vindija remains. Krapina 4 (left) and Vindija 261 (right) are the most complete adult Neandertal frontal bones from their respective sites. Although fragmentary, the supraorbital fossils from Krapina represent between 7 and 14 individuals, and the supraorbital fossils from Vindija represent between 6 and 10 individuals. Photo courtesy M. H. Wolpoff.

Accurate sexing and aging of specimens is confounded by the fragmentary nature of the Vindija and Krapina remains (see Figure 1). Although all studies (e.g., Smith & Ranyard, 1980) comparing the Krapina and Vindija hominids have excluded clearly juvenile specimens, none has attempted to control for sex or adult age variation in the two samples. A frequently overlooked source of morphological variation is the continued remodeling that occurs during adulthood. Israel (1968, 1971, 1973, 1977). Behrents (1985), and others (e.g., Russell, 1983; Enlow, 1996) have demonstrated that cranial growth and remodeling continue throughout life. Russell (1983) established that among Australian aboriginal males, development of the supraorbital region continues after adulthood is reached. Thus, a sample of older adults can appear more robust than a sample of young individuals drawn from the same population.

Because anatomy can vary with sex and age, one must be sure that when asking phylogenetic questions one is comparing fossil samples with similar age and sex compositions. Determining which specimens and samples are comparable can be a difficult task even with relatively complete specimens, but it is especially daunting when the remains are fragmentary, such as in the case of the Krapina and Vindija cranial remains. Metric analysis of the Vindija supraorbitals places them between the earlier Krapina Neandertals and anatomically modern humans of the Central European early Upper Paleolithic (Smith & Ranyard, 1980). One possible explanation for the smaller size and different shape of the Vindija supraorbitals compared to those from Krapina is that the two samples represent two morphologically different

Table 1 Samples

Sample	*n*	Specimens
Krapina	13	3, 4, 6, 23, 37·1, 37·3, 37·4, 37·5, 37·6, 37·8, 37·9, 37·10, 37·11
Vindija	6	202, 260, 261, 262, 284, 305
Upper Paleolithic*	11	Brno 2, Dolní Vestonice 3, Kelsterbach, Mladeč 1, 2, and 5, Paderborn, Pavlov, Podbaba, Volgelherd 1 and 2
Neolithic*	50	Altendorf

*Not used in the analysis of sample bias.

populations. This is the interpretation supported by Smith & Ranyard (1980), who argued that the differences are the result of phyletic change within Neandertals. Alternatively, both samples represent the same population, and the differences between the two are the result of sample bias. In this study, we test the hypothesis that the metric differences between the Krapina and Vindija supraorbital samples are due to age and/or sex sample composition bias, rather than actual biological difference between populations. We do not limit our analysis to examining the suggestion that the Vindija sample is female and/or young biased (Bräuer, 1989; Stringer & Bräuer, 1994). We examine all of the possible ways that sample bias could explain Vindija's more modern-like anatomy.

Materials

We chose the supraorbital region for two reasons. First, supraorbitals are particularly well represented among the fossil samples. Descriptions of the Vindija and Krapina supraorbital samples can be found in Wolpoff *et al.* (1981), Smith *et al.* (1985), Radovčić *et al.* (1988), Smith & Ahern (1994), and Ahern (1998). There are a total of ten supraorbitals from Vindija. Three of these (Vi 224, Vi 227, and Vi 279) have been designated as juveniles based upon the gracility of the browbridge and lack of fine cancellous bone on the anterior wall of the frontal sinus (Wolpoff *et al.*, 1981; Ahern & Smith, in press). These juvenile specimens have been excluded from metric comparisons with the Krapina and Upper Paleolithic samples (Smith & Ranyard, 1980). Of the remaining seven adult supraorbitals, six preserve the metric points used in this analysis. There is a total of 18 supraorbital fossils from Krapina. Of these, 13 are adult and preserve the metric points used in this study. The second reason we chose the supraorbital region for our analysis is that it has been posited as especially indicative of the transitional nature of the Vindija hominids (Smith & Ranyard, 1980; Wolpoff *et al.*, 1981). Dr. Fred Smith generously let us use his original measurements of the Krapina, Vindija, Central European Upper Paleolithic, and Altendorf (Neolithic) supraorbitals. The Krapina and Vindija measurements were replicated successfully on the original fossils.

Methods

Metric analysis of the supraorbital samples

We first reexamined the metric differences between the Krapina, Vindija and Upper Paleolithic modern European supraorbitals, that were originally studied by Smith & Ranyard (1980). A reexamination was necessary in order to confirm that the Vindija supraorbital sample does actually differ from the Krapina sample. Furthermore, our analysis differed from that of Smith & Ranyard (1980) in that we included an additional specimen from Vindija (Vi

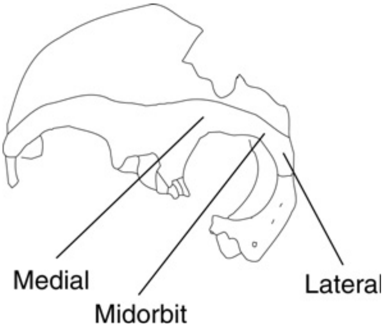

Figure 2. Metric landmarks of the supraorbital region as defined by Smith & Ranyard (1980). See text for definitions. Illustrated specimen is Krapina 6.

305) and an additional modern human sample (Neolithic from Altendorf). We examined measurements and derived indices based upon three landmarks of the supraorbital region as originally defined by Smith & Ranyard (1980). These landmarks are: medial, lateral, and midorbit. These points are illustrated in Figure 2. Rather than being actual points, the landmarks refer to definable parasagittal planes along the supraorbital region. Medial lies on the orbital segment of the supraorbital torus (or supercilliary arch) just lateral to the medial orbital margin. This landmark invariably corresponds to the thickest and highest point on the torus (Smith & Ranyard, 1980). Lateral corresponds to the thickest point on the lateral segment of the torus that is lateral to a parasagittal plane that passes through *frontotemporale*. Midorbit corresponds to the thinnest point on the torus between medial and lateral (Smith & Ranyard, 1980).

The supraorbital chords fall into two categories: (1) measures of thickness, and (2) measures of projection. Thickness measures for medial, midorbit and lateral follow those defined by Smith & Ranyard (1980). Measures of thickness are taken from the anterior margin of the orbit to the superior aspect of the torus. The superior aspect of the torus is usually easily defined at midorbit and lateral, where there tends to be more of an acute angle between the torus roof and the frontal squama. However, the browridge frequently grades more evenly with the squama at medial. Smith & Ranyard (1980) measured thickness at medial from the orbital margin to the deepest point within the sulcus between the supraorbital and the frontal squama.

Analysis of sample bias

The argument that the differences between the Vindija and the Krapina supraorbital samples are due to sample bias can be stated as a testable hypothesis: that both the Vindija and the Krapina samples come from the same population. Classical statistical tests would test whether they are both samples from the same population. However, classical tests rely on the assumed properties of the populations from which the samples are drawn. The most often used population is that of the normal distribution, with other common distribution, being t, F, χ^2 or γ. The assumption about the population is critical in determining the applicability of particular statistical tests. If it cannot be assumed that the sample is from a population with known statistical properties, one could still apply a wide range of statistical tests if the sample size is large enough by calling upon the central limit theorem, which states that a sufficiently large sample will converge upon a Gaussian normal distribution. Fossil samples do not satisfy either of those conditions: one can not make assumptions about the statistical properties of the population, and rare is the case when there is a fossil dataset of large enough sample size. Thus, a method that does not rely upon classical statistical tests and their assumptions is necessary for our analysis.

Given the established metric differences between the Krapina and Vindija supraorbital samples (Smith & Ranyard, 1980), the hypothesis, that the two samples come from the same population, can be rephrased as the null hypothesis that the differences between the Vindija and Krapina samples are due solely to sample bias rather than populational differences. To test this hypothesis, we employed a method that involved the generation of a normally distributed population followed by the resampling of a portion of this normal distribution in order to determine the probability that sample bias explains the observed differences between the Krapina and Vindija samples. This method makes two major assumptions, both of which reinforce the hypothesis that the Krapina–Vindija differences are due to sample composition and not anatomical difference between two populations. We assume that: (1) age and sex variation are patterned in the same way (e.g., a female biased sample cannot be distinguished from a young-biased sample) and (2) all of the variables that we analyzed vary to some degree with sex and/or age, and that population distributions are divisible by age and sex. These assumptions are obviously not "real-world" ones, since their actual reality is unclear [assumption (1)] or unknown [assumption (2)]. However, we make these assumptions precisely because they reinforce the null hypothesis and thus reduce the chance of Type I error.

Bräuer's (1989) and Stringer & Bräuer's (1994) criticism of the transitional nature of the Vindija late Neandertals focuses upon an over-representation of females and/or young in the Vindija sample. In terms of cranial osteology, we do not have a strong case to differentiate between females and young, since the young and female distributions overlap. However, if we assume that an osteological population consists of old and young, and the same population consists of the same number of males and females, we could posit that half of the variation for a trait that varies with age and sex would be due to females or young. Assuming that females on average are smaller than males on average, and that young on average are smaller than old on average, the hypothesis logically derived from Bräuer's (1989) criticism is that the Vindija sample is a sample from the lower half of the population (although other possibilities exist and are discussed below). If Krapina is assumed to be a random sample from the population, it follows that then Vindija is hypothesized to be a sample from the lower half of Krapina. Thus our assumption that age and sex variation are patterned in the same way both reinforces and is consistent with the null hypothesis that the Krapina–Vindija differences are due to sample composition bias.

Our method assumes that all of the variables that we examined vary, at least to some degree, with age and/or sex. If the variables do not vary with age or sex, then a female or young biased sample (as posited for Vindija) would not necessarily differ from a non-biased sample. Thus, the differences between the Krapina and Vindija supraorbital samples could not be attributed to age and/or sex sample bias. Our method even assumes that there is no overlap between females and males and between young and old, and that the lower half of the population distribution comprises females and young. This is, of course, unrealistic, but considerably strengthens the null hypothesis by discounting any males or old individuals that fall within the lower half of the population distribution and any females or young that fall in the upper half of the population distribution.

As illustrated in Figure 3, there are four ways in which sample bias could explain the differences between the Krapina and Vindija samples. (1) The Vindija sample has an over-representation of young and/or females, while the Krapina sample is normally distributed [Figure 3(b)]. (2) Both of the Krapina and Vindija samples have an over-representation of females and/or young, but the Vindija sample more so [Figure 3(c)]. (3) The Vindija sample has an over-representation of young and/or

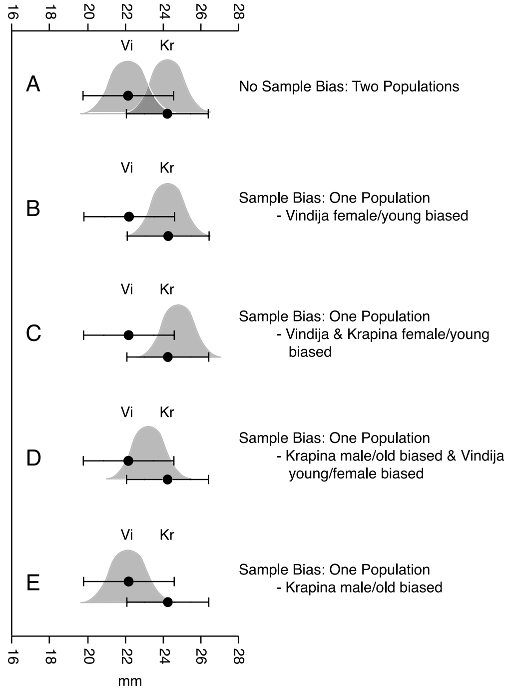

Figure 3. Possible scenarios that could result in the metric differences between the Vindija ("Vi") and Krapina ("Kr") supraorbital samples. For each plot, (a)–(e), the sample means and two standard deviation bars are given for the Vindija ("Vi") and Krapina ("Kr") samples. The density curve represents the population distribution in each scenario. The variable plotted is lateral projection.

females, while the Krapina sample would have an over-representation of old and/or males [Figure 3(d)]. (4) The Vindija sample is normally distributed, while the Krapina sample has an over-representation of males and/or old [Figure 3(e)]. The first three explanations are consistent with Bräuer's (1989) and Stringer & Bräuer's (1994) suggestion that females and/or young are over-represented in the Vindija sample. The fourth explanation is also a possibility, although it would indicate that early Neandertals in south Central Europe were much more similar to Upper Paleolithic modern humans than anyone has posited. We consider the third and fourth explanations unlikely, since it appears that the Krapina sample is not male and/or old biased. Wolpoff's (1979) analysis of the dental remains indicates that 88% of the dental individuals were less than 21 years old at death. Furthermore, recent analysis of the Krapina frontal sample indicates that it is female and young biased (Ahern, 1998, 1999). Nevertheless, in order to falsify the null hypothesis that the differences between the Vindija and Krapina samples are due to sample bias rather than populational differences, we must disprove all four of the sample bias scenarios illustrated in Figure 3(b)–(e). For each variable, we employed variants of our basic method in order to determine the probability of each of the sample bias scenarios. These variants differed in terms of the parameters for the normally distributed population that was generated as well as in terms of what portion of the generated population was resampled. The three variants are bootstraps of: (1) a Krapina-based population, (2) an intermediate-based population, and (3) a Vindija-based population.

Bootstrap of a Krapina-based population. This bootstrap method tests the hypothesis that the Vindija sample has an over-representation of females and/or young while the Krapina sample is either normally distributed or also has an over-representation of females and/or young. In order to determine the probability that the Vindija sample is female/young biased while the Krapina sample is either normal [Figure 3(b)] or female/young biased [Figure 3(c)], we assessed the probability that a sample of the same size as Vindija and with a mean equal to or less than Vindija's could be sampled from the lower half of a Krapina-based population for each variable. The Krapina-based population was generated as a normally distributed population (n= 10,000) based upon the Krapina mean and standard deviation. This method is clearly adequate for dealing with a case in which Vindija is female/young biased while

Krapina is normal. It is also applicable to a case in which both Vindija and Krapina are female/young biased, since the Vindija sample would be biased *relative* to the Krapina sample. If we were to generate the normal population using, for example, Krapina's mean plus one standard deviation for the population mean, we would significantly weaken the null hypothesis and increase the chance of Type I error.

Bootstrap of an intermediate-based population. The most extreme sample bias scenario is that the Vindija sample is female/young biased, while the Krapina sample is male/old biased [Figure 3(d)]. The bootstrap of an intermediate-based population method tests the hypothesis that this extreme bias scenario can explain the differences between the Krapina and Vindija samples. In order to determine the likeliness of this scenario, we assessed the probability that a sample of the same size as Vindija and with a mean equal to or less than Vindija's could be sampled from the lower half of a population ($n=10,000$) whose mean and standard deviation were averages of the Vindija and Krapina means and standard deviations. By using averages of the samples' means and standard deviations, the Vindija supraorbital metrics were small for the generated population, while the Krapina metrics were large for the generated population. Since this method uses a generated population whose mean lies in between the Vindija and Krapina samples, it makes the hypothesis more difficult to refute since there is greater overlap between the Vindija sample and the generated population than in the first bootstrapping method.

Bootstrap of a Vindija-based population. The last methodological variant that we employed dealt with a sample bias scenario in which the Vindija sample is normal, while the Krapina sample is male/old biased. The likeliness of this scenario was determined by assessing the probability that a sample of the same size as Krapina and with a mean equal to or greater than Krapina's would be sampled from the upper half of a Vindija-based population. The Vindija-based population was generated as a normally distributed population ($n=10,000$) based upon the Vindija mean and standard deviation.

The three bootstrap methods were written as scripts and executed using *Resampling Statistics* (Version 4.0.7, Resampling Stats, Inc.). The scripts are given in Appendix A. The null hypothesis that the Krapina–Vindija differences are due to sample bias was falsified if all three methodological variants yielded significant ($P<0{\cdot}05$) results for one or more variable.

Results

Metric comparisons

Summary statistics for the supraorbital samples are given in Table 2 and means and double standard deviations for the chords are plotted in Figure 4 and for the indices in Figure 5. Medial projection summary statistics are reported in Table 2 for Krapina and the modern samples, but are not plotted in Figure 4 since no Vindija supraorbitals preserve this measurement. For all but one of the measurements, the Vindija sample lies in between the earlier Neandertals from Krapina and the Upper Paleolithic and Neolithic modern humans. For thickness of the medial supraorbital, the single Vindija specimen (Vi 305) approximates the Upper Paleolithic mean. However, this measurement does not distinguish early Neandertals from recent Europeans, as demonstrated by the Altendorf sample's mean. For both indices, the Vindija mean is intermediate between the Krapina and Upper Paleolithic means. Because these indices measure shape, they are more demonstrative of evolutionary change. The differences observed for the chord measurements could possibly be explained allometrically. According to Smith & Ranyard (1980), the observed

Table 2 **Summary statistics for measurements and indices of the supraorbital region**

	Projection			Thickness			Indices	
	Lateral	Midorbit	Medial	Lateral	Midorbit	Medial	Mi/La Thick.	Mi/La Proj.
Krapina								
Mean	24·3	23·9	19·8	12·5	10·7	16·6	85·0	99·6
S.D.	1·4	1·2	2·3	1·6	1·8	4·2	7·7	6·9
n	8	11	5	11	13	4	11	8
Vindija								
Mean	22·1	18·9	—	10·6	8·9	18·9	80·7	86·5
S.D.	1·8	2·9	—	0·5	0·8	—	4·1	12·6
n	5	6	0	5	6	1	5	5
Upper Paleolithic*								
Mean	19·8	15·7	11·5	8·6	6·1	19·4	71·2	78·1
S.D.	3·2	4·2	3·7	1·5	1·1	3·7	6·5	11·2
n	8	8	5	11	11	9	11	8
Altendorf Neolithic*								
Mean	17·0	12·5	9·8	8·0	6·5	16·7	82·2	72·7
S.D.	3·1	3·5	2·8	1·5	1·3	2·6	13·9	11·2
n	32	32	31	50	50	49	50	31

*Used for comparison, only. Not used in the analysis of sample bias.

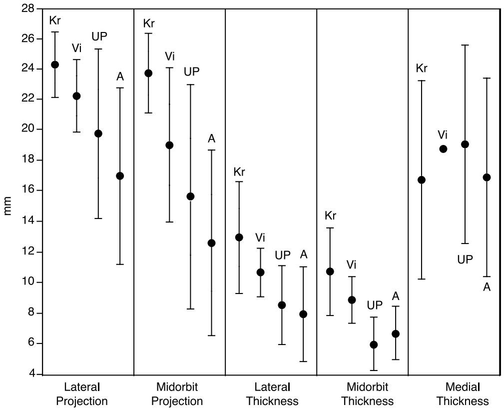

Figure 4. Plot of means and two standard deviations for supraorbital thickness and projection measurements. "Kr": Krapina, "Vi": Vindija, "UP": Central European Upper Paleolithic, "A": Neolithic sample from Altendorf, Germany. Note that the Vindija sample is intermediate between the earlier Krapina Neandertals and the Upper Paleolithic modern humans.

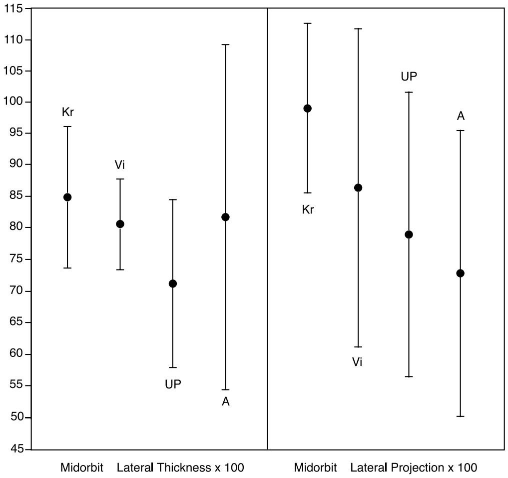

Figure 5. Plot of means and two standard deviations for indices or supraorbital shape. "Kr": Krapina, "Vi": Vindija, "UP": Central European Upper Paleolithic, "A": German Neolithic sample from Altendorf. Note that the Vindija sample lies intermediate between the earlier Krapina Neandertals and the Upper Paleolithic Central Europeans.

shape changes, as seen in Figure 5, are less likely to be related to allometry than the chords. An allometric argument for all of the observed differences between the Krapina and Vindija samples is further weakened by Trinkaus & Smith's (1995) demonstration that body size did not significantly differ between the two samples. The intermediacy

Table 3 Analysis of sample bias: resampling results

Variable or index	Bootstrap of a Krapina-based population*	Bootstrap of an intermediate-based population†	Bootstrap of a Vindija based population‡
Midorbit/lateral projection	0·0005§	0·0334	0·1304
Midorbit/lateral thickness	0·7938	0·9793	0·0956
Midorbit thickness	0·2013	0·7420	<0·0001
Midorbit projection	<0·0001	<0·0001	<0·0001
Lateral thickness	0·0924	0·5274	<0·0001
Lateral projection	0·0054	0·1740	0·0258

*Probability (*P*) of drawing a sample with an equal or smaller mean as the Vindija mean from the lower half of a Krapina-based population. This is the probability that the Vindija sample is female/young biased, while the Krapina sample is either normal or also female/young biased.

†Probability (*P*) of drawing a sample with an equal or smaller mean as the Vindija mean from the lower half of a population generated using the average of the Krapina and Vindija means and standard deviations for the population mean and standard deviation, respectively. This is the probability that the Vindija sample is female/young biased, while the Krapina sample is male/old biased.

‡Proability (*P*) of drawing a sample with an equal or greater mean as the Krapina mean from the upper half of a Vindija-based population. This is the probability that the Vindija sample is normal, while the Krapina sample is male/old biased.

§Degrees of freedom are not reported since they are not relevant to our statistical approach or most resampling approaches. These techniques do not rely upon traditional statistical distributions.

of the Vindija supraorbital sample was first reported by Smith & Ranyard (1980), and what we report here repeats their findings.

Analysis of sample bias

Only one Vindija specimen (Vi 305) preserves the medial point. Thus, medial measurements were not included in our analysis of sample bias. Of the 18 tests run on the six variables, nine were significant at $P<0{\cdot}05$ (see Table 3). Analysis of one variable, midorbit/lateral thickness, yielded no significant results, while analysis of another variable, midorbit projection, yielded significant results for all three bootstrap methods. Two variables yielded significant results for two out of three of the bootstrap methods, and two variables gave significant results for one out of three bootstrap methods.

Bootstrap of a Krapina-based population. Results for three of the six variables are significant at $P<0{\cdot}01$. Thus, for half of the variables, an over-representation of females and/or young among the Vindija sample while the Krapina sample is either normal or female/young biased is not likely.

Bootstrap of an intermediate-based population. Results for only one of the six variables is significant at $P<0{\cdot}01$, although there is an additional significant result at $P<0{\cdot}05$. Thus, for two out of six variables, the differences between the Krapina and Vindija are unlikely due to an over-representation of females and/or young in the Vindija sample while the Krapina sample has an over-representation of males and/or old.

Bootstrap of a Vindija-based population. Results for four of the six variables were significant at $P<0{\cdot}05$, and three of these were significant at $P<0{\cdot}0001$. Thus, it is highly unlikely that the Krapina sample is a male/old biased subset of a Vindija-based population.

Based upon the results of the analyses, scenario (e) in Figure 3 (Vindija is normal while Krapina is male/old biased) is the least likely explanation for the differences between the Krapina and Vindija supraorbital samples. Scenarios (b) and (c) in Figure 3 (Vindija is female/young biased while Krapina is either normal or female/young biased) are also unlikely. Only if the

Vindija sample is assumed to be female/young biased and the Krapina sample male/old biased [Figure 3(d)] does sample composition bias become a viable explanation for the differences between the Krapina and Vindija supraorbital samples. Furthermore, results of all three bootstrap methods were significant for one variable, midorbit thickness, indicating that sample bias cannot explain the differences between the Krapina and Vindija samples for this variable.

Discussion and conclusions

Our results falsify the hypothesis that the metric differences between the Krapina and Vindija supraorbital samples are due solely to sample composition bias. For half of the tests of significance, the probability that sample bias explains the metric differences is less than 5%. Yet, sample composition bias is not the only alternative to an evolutionary explanation. Klein (1989), Howell (1984), and others (e.g., Bräuer, 1989) claim that the transitional aspects of the Vindija hominids might be related to a reduction in robusticity and overall body size. Trinkaus & Smith (1995) have demonstrated that the Vindija hominids were not significantly smaller than the Krapina hominids or other Neandertals. Furthermore, mean robusticity measures of the Vindija mandibles are significantly greater than those of the Krapina mandibles (Ahern & Smith, 1993), indicating that the Vindija hominids, in at least one respect, were not less robust versions of the Krapina hominids. Our and previous findings support the original conclusion that the Krapina and Vindija hominids represent different populations (Smith & Ranyard, 1980; Wolpoff *et al.*, 1981; Smith *et al.*, 1989).

As Smith & Ranyard (1980) and others (e.g., Wolpoff *et al.*, 1981; Smith *et al.*, 1989) have argued, it is reasonable to posit that the Krapina and Vindija fossils represent two temporal samples of Neandertals and thus can be compared to analyze temporal trends within Neandertals. Not only do the Vindija hominids differ from the Krapina hominids, they differ in many ways that make them more like the early modern humans of the Upper Paleolithic. Our results regarding the supraorbital fossils support this conclusion. However, intermediacy of the Vindija hominids does not necessarily mean that the Vindija hominids, and thus Neandertals, evolved into, or even contributed any genes to, the people of the Upper Paleolithic. As Lieberman (1995) argues, temporal trends are not effective means to test phylogenetic hypotheses. Temporal trends may be indicative of actual evolutionary change and continuity, but they may also be the result of population replacement. For example, the Vindija hominids may be morphologically different not because they are the changed descendents of the Krapina hominids, but rather because they represent a completely different population that came from elsewhere. Although the temporal trend observed in the Krapina–Vindija–Upper Paleolithic sequence cannot falsify phylogenetic hypotheses, none the less it can be used to generate further phylogenetic hypotheses. The hypothesis that Neandertals and their successors in Europe were biologically conspecific would be supported by an evolutionary interpretation of the Krapina–Vindija–Upper Paleolithic sequence.

Summary

The late Neandertals from Vindija Cave have been posited as evolutionarily intermediates between the earlier Neandertals from Krapina and the modern humans of the Upper Paleolithic (e.g., Smith & Ranyard, 1980; Wolpoff *et al.*, 1981). However, some (e.g., Bräuer, 1989) have argued that Vindija's morphological intermediacy is due to an over-representation of females and/or young individuals in the sample, and that this bias is not easily detected because of the fragmentary nature of the remains. We re-examined the intermediacy of the Vindija supraorbital sample and tested the null hypothesis that the supraorbital metric differences between the Krapina and Vindija samples are due solely to sample bias. Our results indicate that there is a significantly low probability that sample bias accounts for the differences between these samples. Thus, the morphological intermediacy of the Vindija supraorbitals between those from Krapina and the Upper Paleolithic is not due to sample bias. This anatomical intermediacy may reflect evolutionary intermediacy, but alternative explanations of the temporal trends within the Krapina–Vindija–Upper Paleolithic sequence, such as population replacement, cannot be ruled out.

Acknowledgements

We thank the Croatian Academy of Sciences and Arts for continued access to the Vindija fossils and the Croatian Museum of Natural History for continued access to the Krapina fossils. The following were of particular assistance: Maja Paunović, Jakov Radovčić, Ivor Karavanić, Gordana Jambresić, and Tomislav Śokec. We thank Fred H. Smith for his helpful comments on an earlier version of this paper and for kindly allowing us to use his supraorbital data. We also appreciate the helpful comments of Milford Wolpoff, Brian Birch, Alison Hofbauer, Meredith McLaughlin, Terry Harrison, and three anonymous reviewers. Funding to one of the authors (JA) from the Department of Anthropology, the Rackham School of Graduate Studies, the International Institute of the University of Michigan, and the Office of Research, University of Wyoming supported portions of this project.

References

Ahern, J. C. M. (1998). Late Pleistocene frontals of the Krvatsko Zagorje: an analysis of intrapopulational variation among South Central European Neandertals. Ph.D. Dissertation, University of Michigan, Ann Arbor.

Ahern, J. C. M. (1999). Computational methods for addressing age and sex sample bias: a south-central European test case. *Am. J. phys. Anthrop.* (Suppl.) **27,** 81–82.

Ahern, J. C. M. & Smith, F. H. (1993). The transitional nature of the late Neanderthal mandibles from Vindija Cave, Croatia. *Am. J. phys. Anthrop.* (Suppl.) **16,** 47.

Ahern, J. C. M. & Smith, F. H. (in press). Aspects of supraorbital torus morphology in the Le Moustier 1 adolescent Neandertal. In (H. Ullrich, Ed.) *The Neandertal Adolescent Le Moustier 1—New Aspects, New Results*. Berlin: Museum für Vor- und Frühgeschichte.

Behrents, R. (1985). *Growth in the Aging Craniofacial Skeleton*. Ann Arbor, MI: Center for Human Growth and Development.

Bräuer, G. (1989). The evolution of modern humans: a comparison of the African and non-African evidence. In (P. Mellars & C. B. Stringer, Eds) *The Human Revolution*, pp. 121–154. Edinburgh: Edinburgh University Press.

Duarte, C., Maurício, J., Pettitt, P. B., Souto, P., Trinkaus, E., van der Plicht, H. & Zilhão, J. (1999). The early Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. *Proc. natl. Acad. Sci. USA* **96,** 7604–7609.

Enlow, D. H. (1996). *Essentials of Facial Growth*. Philadelphia: Saunders.

Frayer, D. (1992). The persistence of Neanderthal features in post-Neanderthal Europeans. In (G. Bräuer & F. Smith, Eds) *Continuity or Replacement: Controversies in* Homo sapiens *Evolution*, pp. 179–188. Rotterdam: Balkema.

Frayer, D. W., Wolpoff, M. H., Smith, F. H., Thorne, A. G. & Pope, G. G. (1993). The fossil evidence for modern human origins. *Am. Anthrop.* **95,** 14–50.

Hawks, J. & Wolpoff, M. (2001). Brief communication: paleoanthropology and the population genetics of ancient genes. *Am. J. phys. Anthrop.* **114,** 269–272.

Howell, F. (1984). Introduction. In (F. H. Smith & F. Spencer, Eds) *Origins of Modern Humans: A World Survey of the Fossil Evidence*, pp. xiii–xxii. New York: Alan R. Liss.

Israel, H. (1968). Continuing growth in the human cranial skeleton. *Arch. Oral Biol.* **13,** 133–137.

Israel, H. (1971). The impact of aging upon the adult craniofacial skeleton. Ph.D. Dissertation. University of Alabama, Birmingham.

Israel, H. (1973). Age factor and the pattern of change in craniofacial structures. *Am. J. phys. Anthrop.* **39,** 111–128.

Israel, H. (1977). The dichotomous pattern of craniofacial expansion during aging. *Am. J. phys. Anthrop.* **47,** 47–52.

Klein, R. (1989). *The Human Career*. Chicago: University of Chicago Press.

Lieberman, D. E. (1995). Testing hypotheses about recent human evolution from skulls: integrating morphology, function, development, and phylogeny. *Curr. Anthrop.* **36,** 159–198.

Radovčić, J., Smith, F. H., Trinkaus, E. & Wolpoff, M. H. (1988). *The Krapina Hominids: An Illustrated Catalog of the Skeletal Collection*. Zagreb: Mladost & the Croatian Natural History Museum.

Relethford, J. (1998). Genetics of modern human origins and diversity. *A. Rev. Anthropol.* **27,** 1–23.

Relethford, J. (2001). Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *Am. J. phys. Anthrop.* **115,** 95–98.

Rink, W., Schwarcz, H., Radovčić, J., Rukavina, D., Karavanić, I. & Richter, D. (1999). ESR dating at the Croatian Paleolithic sites of Krapina, Vindija, and Mujina Pécina. Paper presented at: *Krapina 1899–1999. The Krapina Neandertals and Human Evolution in Central Europe* (Zagreb, Croatia).

Rink, W., Schwarcz, H., Smith, F. & Radovčić, J. (1995). ESR ages for Krapina hominids. *Nature* **378,** 24.

Russell, M. (1983). Browridge development as a function of bending stress in the supraorbital region. *Am. J. phys. Anthrop.* (Suppl.) **60,** 248.

Smith, F. H. (1982). Upper Pleistocene hominid evolution in South-Central Europe: a review of the evidence and analysis of trends. *Curr. Anthrop.* **23,** 667–703.

Smith, F. H. (1984). Fossil hominids from the Upper Pleistocene of Central Europe and the origin of modern Europeans. In (F. H. Smith & F. Spencer, Eds) *The Origins of Modern Humans: a World Survey of the Fossil Evidence*, pp. 137–209. New York: Alan R. Liss.

Smith, F. H. (1992). The role of continuity in modern human origins. In (G. Bräuer & F. H. Smith, Eds) *Continuity or Replacement: Controversies in* Homo sapiens *Evolution*, pp. 145–155. Rotterdam: Balkema.

Smith, F. H. & Ahern, J. C. (1994). Additional cranial remains from Vindija Cave, Croatia. *Am. J. phys. Anthrop.* **93,** 275–280.

Smith, F. H., Boyd, D. C. & Malez, M. (1985). Additional Upper Pleistocene human remains from Vindija Cave, Croatia, Yugoslavia. *Am. J. phys. Anthrop.* **68,** 375–383.

Smith, F. H., Falsetti, A. B. & Donnelly, S. M. (1989). Modern human origins. *Yearb. phys. Anthrop.* **32,** 217–226.

Smith, F. H. & Ranyard, G. (1980). Evolution of the supraorbital region in upper Pleistocene fossil hominids from south-central Europe. *Am. J. phys. Anthrop.* **53,** 589–610.

Smith, F., Trinkaus, E., Pettitt, P. & Karavanić, I. (1999). Direct radiocarbon dates for Vindija G_1 and Velika Pećina Late Pleistocene hominid remains. *Proc. natl. Acad. Sci. (USA)* **96,** 12281–12286.

Stringer, C. & Bräuer, G. (1994). Methods, misreadings, and bias. *Am. Anthrop.* **96,** 416–424.

Stringer, C. B. & Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* **239,** 1263–1268.

Szilvássy, J., Kirtscher, H. & Vlček, E. (1987). Die Bedeutung rontgenologischer Methoden für anthropologische Untersuchung ur- und fruhgeschichtlicher Graberfelder. *Annls. Naturhist. Mus. Wien* **89,** 313–352.

Tattersall, I. & Schwartz, J. (1998). Morphology, paleoanthropology, and Neandertals. *Anat. Rec.* **253,** 113–117.

Trinkaus, E. & Smith, F. H. (1995). Body size of the Vindija Neandertals. *J. hum. Evol.* **28,** 201–208.

Wolpoff, M. (1979). The Krapina dental remains. *Am. J. phys. Anthrop.* **50,** 67–114.

Wolpoff, M. (1999). *Paleoanthropology*. 2nd edn. New York: McGraw-Hill.

Wolpoff, M. H., Hawks, J., Frayer, D. W. & Hunley, K. (2001). Modern human ancestry at the peripheries: a test of the replacement theory. *Science* **291,** 293–297.

Wolpoff, M. H., Smith, F. H., Malez, M., Radovčić, J. & Rukavina, D. (1981). Upper Pleistocene hominid remains from Vindija Cave, Croatia, Yugoslavia. *Am. J. phys. Anthrop.* **54,** 499–545.

Appendix 1: Resampling Stats (©Resampling Stats, Inc.) scripts for the three parametric bootstrap methods employed in this study

```
'BOOTSTRAP OF KRAPINA-BASED POPULATION
'A script for Resampling Stats (v.4.0.11, ©Resampling Stats, Inc.)
'Written 3/02 by
'Sang-Hee Lee (sang-hee.lee@ucr.edu) and Jim Ahern (jahern@uwyo.edu)
'Determines the probability that one sample (''Vindija'') could be drawn from the lower half
of a population based upon another sample (''Krapina'')
'
maxsize default 15000
copy (input) n   'input number of resampling
copy (input) Kmean   'input Krapina mean
```

```
copy (input) Ksd   ’input Krapina st dev
copy (input) Vmean   ’input Vindija mean
copy (input) Vn   ’input Vindija sample size
normal n Kmean Ksd K_normal ’K_normal is the generated Krapina as if normal
weed K_normal>Kmean ’K_lower K_lower is the lower half of the Kr distribution as if
normal repeat n
   sample Vn K_ lower FakeV
   mean FakeV FakeVbar
   score FakeVbar Vbar ’size of Vbar should equal the number of ’’REPEAT’’ line 3 end
histogram Vbar
count Vbar<=Vmean Surprise ’counting the number less than observed V mean
Divide Surprise n pmean
Print in Kmean Ksd Vmean Vn pmean
```

```
’BOOTSTRAP OF INTERMEDIATE-BASED POPULATION
’A script for Resampling Stats (v.4.0.11, ©Resampling Stats, Inc.)
’Written 3/02 by
’Jim Ahern (jahern@uwyo.edu) and Sang-Hee Lee (sang-hee.lee@ucr.edu)
’Determines the probability that sample A (’’Vindija’’) could be drawn from the lower half
of a population based upon the average mean and SD of samples A & B (’’Krapina’’)
’
maxsize default 15000
copy (input) n   ’input number of resampling
copy (input) Kmean   ’input Krapina mean
copy (input) Ksd   ’input Krapina st dev
copy (input) Vmean   ’input Vindija mean
copy (input) Vn   ’input Vindija sample size
copy (input) Vsd   ’input Vindija st dev
concat Kmean Vmean Means
Mean Means IntMean ’calculates average of Kr and Vi means
Concat Ksd Vsd SDs
Mean SDs IntSd   ’calculates average of Kr and Vi means
normal n Means SDs I_normal ’I_normal is the generated Intermediate as if normal
weed I_normal>IntMean I_lower ’I_lower is the lower half of the Int distribution as if
normal
repeat n
   same Vn I_lower FakeV
   mean FakeV FakeVbar
   score FakeVbar Vbar   ’size of Vbar should equal the number of ’’REPEAT’’ line 3
end
histogram Vbar
count Vbar<=Vmean Surprise ’counting the number less than observed V mean
Divide Surprise n pmean
Print in Kmean Ksd Vmean Vn Vsd pmean
```

```
’BOOTSTRAP OF A VINDIJA-BASED POPULATION
’A script for Resampling Stats (v.4.0.11, ©Resampling Stats, Inc.)
’Written 3/02 by
```

```
’Jim Ahern (jahern@uwyo.edu) and Sang-Hee Lee (sang-hee.lee@ucr.edu)
’Determines the probability that one sample (’’Krapina’’) could be drawn from the upper half of a population based upon another sample (’’Vindija’’)
’
maxsize default 15000
copy (input) n ’input number of resampling
copy (input) Kmean ’input Krapina mean
copy (input) Kn ’input Krapina sample size
copy (input) Vmean ’input Vindija mean
copy (input) Vsd ’input Vindija st dev
normal n Vmean Vsd V_normal ’V_normal is the generated Vindija as if normal
weed V_normal<Vmean V_upper ’V_upper is the upper half of the Vi distribution as if normal
repeat n
   sample Kn V_upper FakeK
   mean FakeK FakeKbar
   score FakeKbar Kbar ’size of Kbar should equal the number of ’’REPEAT’’ line 3
end
histogram Kbar
count Kbar>=Kmean Surprise ’counting the number greater than observed Kr mean
Divide Surprise n pmean
Print n Kmean Kn Vmean Vsd pmean
```