Assigned Resampling Method:
A new method to estimate size sexual dimorphism in samples of unknown sex

Sang-Hee Lee

Department of Anthropology, University of California at Riverside,
Riverside, CA 92521-0418, USA; E-mail: sang-hee.lee@ucr.edu

ABSTRACT   This paper proposes a new method, Assigned Resampling Method (ARM), to estimate the degree of size sexual dimorphism in samples of unknown sex. ARM resamples with replacement pairs of observation from the mixed-sex data and after applying a filter, generates a distribution of dimorphism estimates by transforming the pair values into ratios. The mean of the distribution is proposed as the ARM estimate of sexual dimorphism. Using 40 metric variables from comparative data sets of known sex (91 humans, 46 chimpanzees, and 56 gorillas), the ARM estimates were compared with the observed sexual dimorphism. Results show that the difference between the ARM estimates and the observed sexual dimorphism is within 5% for most of the variables examined. ARM is shown to perform reliably under various conditions of unequal sex ratios and small sample sizes. By directly comparing fragmentary materials without estimating body size, and by using an algorithm that does not rely on sex diagnosis of individual specimens, ARM addresses the challenges of studying sexual dimorphism with fossil samples.

KEY WORDS   sexual dimorphism, Assigned Resampling Method (ARM), data resampling, bootstrapping, fossil samples

Introduction

In living species of higher vertebrates, high sexual dimorphism in body size is correlated with ecological factors such as sex ratio and mating systems. Although clear correlation is yet to be shown in fossil species [PLAVCAN 2000], it is generally understood that sexual dimorphism is one of the major causes of biological variation in a population. Accordingly, documenting its change through time is a fundamental topic that needs to be addressed by examining the fossil evidence. However, several characteristics of fossils impose limitations on analyzing sexual dimorphism.

Most importantly, the sex of a fossil specimen of an extinct species is often controversial: debates continue regarding the sex of important fossil finds, such as AL 288-1 [HÄUSLER and SCHMID 1995; TAGUE and LOVEJOY 1998], Sts 5 and Taung [LOTH and HENNEBERG 1996]. When the skeletal sample comes from a living population, the problem of unknown sex is not an insurmountable one. Several methods have been shown to be reliable in accurate identification of sex from skeletal materials (sex diagnosis is beyond the scope of this paper: for comprehensive compilations of sex identification methods, see BUIKSTRA and UBELAKER [1994], KATZENBERG and SAUNDERS [2000], KROGMAN and İŞCAN [1986], SCHWARTZ [1995]). However, sex identification studies have limited applicability for studies of evolutionary changes in sexual dimorphism, where the primary data come from fossils: it is not clear if the same traits that are used to identify sex in extant populations can be applied with equal reliability to ancestral hominids, which may have been under different selection pressure [HAGER 1989, 1991, 1996]. Furthermore, skeletal elements necessary for applying the methods may be absent because of the fragmentary nature of preservation. The fragmentary nature of fossil data also limits comparisons of anatomical elements, since an element may be present in some specimens and absent in others. This paper reviews the methods so far proposed to estimate the degree of sexual dimorphism in fossil samples and proposes an alternative that addresses the limitations in the previ-ously proposed methods.

Previous efforts to solve the problem

Several attempts have been made to solve the problem of estimating sexual dimorphism in a fossil sample. Since methods of assigning sex to individual specimens as well as those of testing for differences between males and females[1] have only a limited use for fossil specimens, other methods of estimating sexual dimorphism have been proposed that do not rely on the accurate assessment of the sex of individual specimens. These methods employ a wide range of approaches and vary in their robustness depending upon different circumstances.

A commonly used estimate of sexual dimorphism uses the largest and the smallest data points in a sample, i.e., maxima and minima [ARSUAGA et al. 1997; JOHANSON and WHITE 1979; JUNGERS 1988; LEE 1995; MCHENRY 1986; RICHMOND and JUNGERS 1995; ZIHLMAN 1985]. This approach is designed to reflect the maximum possible difference between males and females of a sample. The attraction of this method may lie in its simplicity but more likely in the common situation where the fossil sample consists of only a couple of measurements, which narrows down the list of possible options in statistical methods. As the maximum and minimum define the range of a sample, they can be compared to the ranges of comparative samples [JOHANSON and WHITE 1979]. Alternatively, they can be

---

[1] This paper does not review the large body of literature on these sex diagnosis methods as well as on the causal models of sexual dimorphism, as they are not directly relevant. Interested readers are advised to refer to LEE [1999].

compared to male and female means of comparative samples [MCHENRY 1986; ZIHLMAN 1985]. Since the estimate is the maximum possible sexual dimorphism in a sample, its comparison is conservative, unlikely to make the mistake of rejecting a true null hypothesis (low Type I error). However, it may be too conservative a method, so that it is unlikely to reject the null hypothesis even if the hypothesis is false (high Type II error). More pertinent to fossil studies is the problem that the method is sensitive to new data points and outliers [HAMILTON 1982].

When the sex samples are distinct enough from each other, the mixed-sex distribution will be a bimodal distribution. The two peaks of the bimodal distribution, then, can be assumed to represent male and female samples. A bimodal distribution has been used as evidence for sexual dimorphism in extinct mammalian species [KURTÉN 1969] and in hominids [WOLPOFF 1975, 1976a,b]. However, using bimodality to indicate sexual dimorphism has several problems. First, the appearance of bimodality may be less than obvious and it is difficult to apply an objective criterion for determining bimodality. For example, one person's bimodal distribution [WOLPOFF 1976b] may be another's highly skewed distribution [TRINKAUS 1976]. In addition, the shape of a distribution can vary significantly depending on the size and the starting points of the histogram intervals [OXNARD et al. 1985; SIEGEL 1976]. Second, there is no reason to assume a one-to-one relationship between bimodal distribution and the degree of sexual dimorphism: bimodality may not reflect sexual dimor-

phism, and sexual dimorphism may not show a bimodal distribution. Factors other than sexual dimorphism, such as taxa, population, ecological strategies, can result in a bimodal distribution. Even within the same taxon, a non-dimorphic species can exhibit a bimodal distribution when samples from different geographical sites are pooled [GODFREY et al. 1993]. Sexually dimorphic species can exhibit unimodal distribution when the sexes substantially overlap, when only a small distance separates the sexes as in humans, or when the sample size is small as often in fossil samples.

If it can be assumed that males and females do not overlap in distribution, or that they overlap to an insignificant degree, dividing the mixed-sex distribution at the point where the two sex samples meet would result in a reasonable approximation of the sex samples themselves. The division point can be the mean, which may correspond to the point of intersection of two samples [BENNETT 1981], or the median, under the condition of equal sex ratios. Hereafter, these are referred to as the "mean method" or the "median method," following GODFREY et al. [1993]. Sexual dimorphism can be calculated from the means of the divided samples.

The mean method, and the median method to a lesser degree, has received favorable reviews when compared to other methods such as the coefficient of variation or finite mixture analysis (discussed below) [GODFREY et al. 1993; PLAVCAN 1994]. The method is robust: that is, it is not affected by violations of assumptions, and the estimate consistently maintains its accuracy

under varying degrees of sexual dimorphism and sex ratios. When sex ratios are balanced and intrasexual variability is low, this method provides the most accurate estimates, although the median method is more sensitive to uneven sex ratio and increased intrasexual variability than the mean method. This method does not make assumptions of normality or homoskedasticity (equal variance), which are usually required in many conventional statistical tests. However, the sexes must be assumed not to overlap, an assumption that is rarely if ever met in human samples. Hence, the estimates are expected to increasingly deviate from the true sexual dimorphism in proportion to the degree of actual overlap. In addition, this method is less reliable with small sample size [COPE and LACY 1995].

The coefficient of variation has been used in many studies to gauge sexual dimorphism in fossil data [FLEAGLE et al. 1980; KAY 1982; KIMBEL and WHITE 1988; LEUTENEGGER and SHELL 1987; LOCKWOOD et al. 1996; MCHENRY 1991; PLAVCAN 1994; WOOD 1976]. A coefficient of variation is calculated by dividing the standard deviation with the sample mean and therefore increases when the standard deviation increases. The coefficient of variation method assumes a linear relationship between the variability of the mixed-sex sample and the variability of sex samples that are its components. When sexual dimorphism increases, it results in the increase in the difference between male and female means. This in turn results in the increase in the standard deviation of the mixed-sex sample, which will in turn result in the increase

in the coefficient of variation. Consequently, the coefficient of variation is expected to show a high correlation with sexual dimorphism. In turn, the sexual dimorphism of a sample can be predicted from the coefficient of variation using regression equations [KAY 1982; LEUTENEGGER and SHELL 1987]. The coefficient of variation method can be as accurate as the mean method, but it is more sensitive to uneven sex ratios and higher intrasexual variability [PLAVCAN 1994].

As it uses a standardized unit of variability, the method of the coefficient of variation is more reliable than range-based methods [COPE and LACY 1995]. However, criticism of the coefficient of variation method focuses on the validity of the critical assumption that the variability increases when two samples are combined in a mixed-sex sample. The method has been undermined by the observation that the coefficient of variation of a mixed-sex sample is not necessarily higher than that of each sex sample [ALBRECHT and MILLER 1997; MARTIN 1983; MARTIN et al. 1994; VITZHUM 1990].

In some cases, the extent of overlapping may be so large as to render the mixed-sex distribution unimodal. The method of finite mixture analysis explores what the sample statistics would be if two normal distributions were assumed to be embedded in one unimodal distribution. This method can be applied to sexual dimorphism if the observed distribution is made up of two component distributions of sexes. Theoretically, the component distributions can be derived by applying the method of moments [PEARSON 1894]. Moments

around the mean of the mixed-sex sample are used to estimate the mean and standard deviations of the two underlying distributions. However, solving for the moments requires evaluating polynomial equations of high degree. In the early years, practical applications of the method had to make simplifications by making additional assumptions that certain parameters are known or restrained [COHEN 1967]. For example, the means of the sex subsamples could be mathematically derived, if the standard deviations of the subsamples and the mixing proportion (sex ratio) were known. The finite mixture analysis method became more widely used with the advancements in computers, which made the intensive computing accessible [DONG 1997; GODFREY et al. 1993; JOSEPHSON et al. 1996]. However, additional assumptions are needed for applying the method to fossil samples. Examples of such assumptions include that the sex samples have normal distributions, of equal variance, and of equal mixture in the component distributions of sex. Means of the two distributions have to be separated by two standard deviations before the mixed distribution exhibits bimodality [ROBERTSON and FRYER 1969; TITTERINGTON et al. 1985]. For example, the means of the sex samples could be separated by 47% of the total observed range for small sample size (n = 10), or by 28% for large sample size (n > 100) in an apparently unimodal distribution [GODFREY et al. 1993]. However, this hinges on the assumption of equal proportion and equal variances for the samples. The method of finite mixture analysis has the advantage that it works whether there is minimal or substantial overlap between the two sex samples, and that the resulting estimate of sexual dimorphism is conservative [JOSEPHSON et al. 1996]. However, this method is less reliable when the sample shows high sexual dimorphism [GODFREY et al. 1993; PLAVCAN 1994]. This is not surprising as the finite mixture analysis is most useful for assessing the maximum possible sexual dimorphism in a unimodal distribution. In addition to the assumptions required as discussed above, sensitivity to small sample size [DONG 1997] further limits the usefulness of the method in fossil studies.

One way to measure sexual dimorphism is by eliminating the overlap so that only the ones whose sex can be unambiguously known remain in the mixed-sex distribution [BENNETT 1981; CHAKRABORTY and MAJUMDER 1982]. The area under the mixed-sex distribution curve is calculated, after the area overlapped by both sexes is excluded. Then, an estimate of sexual dimorphism is mathematically derived from the non-overlapping area. This approach has not been applied in any other situations beyond the original proposal and its discussion is at the theoretical level. The theoretical background of this method requires stringent assumptions. First, both sex samples have to be normal without a strong skewness or kurtosis so that the area under the curve can be calculated. Second, samples sizes should be large enough to accurately estimate a variance [BENNETT 1981]. As these assumptions are not likely to be met in a fossil sample, application of this method to fossil data has foreseeable limitations.

The sexual dimorphism estimate proposed by Lovejoy and colleagues also divides the mixed-sex distribution into two samples [LOVEJOY et al. 1989]. Coined "technique dimorphism," their method takes all possible ways of dividing the mixed-sex distribution into two samples. This starts with taking the smallest data point as a female and the rest as males. Sex means are calculated from the divided samples. In the next step, the two smallest data points are considered females and the rest males. Another pair of sex means is calculated. In the following steps, the observed distribution is repeatedly subdivided with the female sample increasing its size by one and the male sample decreasing its size by one. This is repeated until the observed distribution is divided into the largest data point of male and the rest of females. The average for each sex sample is calculated and is used in the ratio of sexual dimorphism. To a degree, this method takes the approach of jackknife [EFRON and TIBSHIRANI 1993]. Each hypothetical case of subdivisions is weighted by the binomial probability, which are then pooled to calculate a total estimate of sexual dimorphism. The method of technique dimorphism is similar to the exact randomization method, where all possible divisions of the mixed-sex distribution form the basis of the comparison. As was the case for the mean and the median methods, the technique dimorphism method assumes that sex distributions do not overlap. Although the method has not been explored in detail, limitations are expected from the assumption of no overlap.

Data resampling relies on the assumption that the observed data adequately represent the population [EFRON and TIBSHIRANI 1993]. In paleoanthropology, a small number of studies have applied data resampling in studying sexual dimorphism in fossil samples when the sample consists of two specimens [LIEBERMAN et al. 1988; RICHMOND and JUNGERS 1995]. The ratio formed by the two specimens is compared against a distribution generated by computing ratios from all possible pairs in a sample distribution, usually an extant ape species. The approach of data resampling has the merit of addressing characteristics of a fossil sample that limit the applicability of established statistical methods. However, the studies that so far have applied data resampling to examine sexual dimorphism in a fossil sample are subject to the same kind of limitations coming from using the traditional estimates of sexual dimorphism (discussed above).

The methods to estimate sexual dimorphism in fossil samples that have been proposed so far are in actuality limited in their applicability to fossils in particular, and in any other samples. One limitation is that choosing what to use among the proposed methods requires some prior knowledge about the sexual dimorphism of the sample to be analyzed. There is no *a priori* reason to use one and not the other, unless assumptions can be made about the amount of overlap between the male and the female distributions and the amount of sexual dimorphism itself. Unless a living analog is used, these assumptions cannot be made with fossil data alone. There is a need for an estimate of sexual dimorphism that does not hinge on such assumptions.

Assigned Resampling Method

The Assigned Resampling Method (ARM) estimates the degree of size sexual dimorphism from a distribution of resampled ratios derived from sampling two values with replacement from a mixed-sex distribution. Each pair that has met the criterion set by a filter (discussed below) is transformed into a ratio, from dividing the larger value of the pair by the smaller. Of the distribution of the ratios thus generated, ARM defines the mean as the estimate of sexual dimorphism (see Appendix 1 for the algorithm). The proposed method is designed to solve two problems in estimating sexual dimorphism in fossil samples. First, because the method does not use sex sample properties such as sex mean or variance, sex of individual specimens becomes irrelevant. Second, the associated standard error of the mean furnishes the variance measure of the ARM estimate, allowing for comparisons between different samples.

Assigned Resampling Method does not posit that all males are bigger than all females. It is not necessary to assume that there is no overlap between the male and female distributions. The method does not require that the observed data have a reliable identification of sex for individual specimens, nor does the method assign sex to a particular individual specimen. An observed value is "assigned" male or female only in relation to the other value in the pair of that particular drawing event. A value may be assigned a male in one event if drawn with an observed value that is smaller, while it will be assigned a female if drawn again, but this time with a larger

value. ARM draws many pairs of values with replacement, which renders the resulting distribution a random sample.

ARM relies on an assumption that when two data points are sampled from a mixed-sex distribution, the larger value is likely to be a male value and the smaller value a female. There are two components in this assumption: males are bigger on average than females in the species studied; and a randomly drawn pair consists of one male and one female. The former component is reasonable if it could be assumed that fossil hominids did not deviate from the general pattern of sexual dimorphism in all higher primates, where males on average are larger than females. The latter only applies to some of the times in actual resampling. There are four possible types of pairs when a pair of values is drawn from a mixed-sex distribution: male-female, female-male, male-male, and female-female. Except for drawing a pair of male-female, the other three cases are erroneous. By assigning a male value to the larger value of the two, the method mitigates the effect of drawing a female-male pair, as it in effect flips the female-male pair into a male-female pair. An effect not mitigated is the deviation from drawing a same sex pair, either a pair of males or a pair of females. Assuming a mixed-sex distribution with an equal number of males and females, there is a probability of 0.5 of drawing a pair of same sex. A filter that would have the effect of excluding erroneous pairs is implemented to solve this problem. As the method does not assign sex to a particular observation, it is impossible to know which pair is of the same sex. Consequently, it is impossible

to decide which pairs should be excluded. However, it is most likely that the pair is of the same sex when both values of the pair are sampled from either the larger end or the smaller end of the range of the mixed-sex distribution. Since the ratio of the larger to the smaller value would be closer to 1 when both values are close to each other, the effect of drawing such a pair would be to decrease the mean of the ratio distribution. Therefore, sampling both values from either the larger end or the smaller end would underestimate the observed sexual dimorphism.

A filter that excludes the pairs that have both the values drawn from either end of the range of values of the mixed-sex distribution can reduce the amount of underestimation. If the probability of drawing a pair of the same sex is 0.5, in other words, 50% (as discussed above), the goal of the filter would be to exclude 50%. However, since there is no *a priori* reason to assume that 50% of the sampled pairs are indeed erroneous, the filter implemented is defined arbitrarily. The filter incorporated in ARM excludes pairs beyond the range of more than ±0.5 standard deviation from the mean. When various filters ranging from mean ±0.25 SD to mean ±2 SD were compared using the data sets described below, the filter proposed above in fact performed the best (discussed below).

Using ratios in statistics has been controversial, as ratios do not make a good statistic when used as a scaled expression [SMITH 1980, 1996; WELSH *et al.* 1988]. This is in part because the numerators are often correlated with the denominators: for example, ratio between teeth size and body size. How-

ever, such criticism does not apply to the ratios in ARM because the numerator is not a part of the denominator.

The observed sexual dimorphism is a ratio of means, while the ARM estimate is a mean of ratios. Because the method takes the mean of the ratios to approximate a ratio of the means, differences between the two measures are expected. However, if the denominators do not vary, the effect is the same as dividing each value by a constant; therefore, the mean of ratios converges to the ratio of means. When the denominators of the ratios converge to a constant (the female mean in this case), the mean of the ratios would have a similar effect as dividing all the values by the same number. This condition cannot be met in ARM, as the denominators and the numerators often may change places. However, this paper argues that this weakness is considered compensated for by the advantage of ARM of not relying on sex assessments of individual specimens, and the behavior of ARM estimates is examined empirically.

Materials

Using skeletal samples that have records of sex can test the degree to which ARM provides reliable measures of sexual dimorphism, and the sensitivity of ARM in varying conditions. Three living species most closely related to fossil hominids were examined: modern humans, chimpanzees, and gorillas. All data were collected by the author in the Hamann-Todd Collection housed at the Cleveland Museum of Natural History, which includes a large selection of non-human primates and humans. Only com-

plete, adult individuals with fully erupted permanent dentition were used. For humans, only those individuals between 20 and 50 years of age were included in the data set. The final data set consists of 91 modern humans (51 males, 40 females), 46 chimpanzees (18 males, 28 females), and 56 gorillas (31 males, 25 females). Using digital sliding calipers, spreading calipers, tape measure, and osteometric board, measurements were taken on the left side of an individual. When the left side was absent, the right side was measured in substitution. Only cases with reasonable bilateral symmetry were included in this study (see LEE [1999] for the original measurements).

Forty variables were measured for each individual specimen (listed in Appendix 2). The variables were selected on the basis of two criteria: many of the measurements are those used in standard osteological practices reflecting the general size of a skeletal element (length and breadth), following Martin who defined the measurements and the measurement points [MARTIN and SALLER 1957; LEE 1999]. The rest of the variables were selected because they allow comparison with hominid fossil data. It is expected that each variable would show a variation in the amount of sexual dimorphism and in the sex sample properties.

Methods

For each variable of each species sample, ARM estimates were calculated as if sex were not known. The ARM estimates were then compared to the observed sexual dimorphism of the male mean divided by the female mean for each variable according to the recorded

sex in four aspects: (1) how closely the ARM estimates approximate the observed values; (2) whether the performance of ARM changes under different filters; (3) whether the performance of ARM changes under conditions of unequal sex ratios; and (4) whether the performance of ARM changes with respect to small sample size.

ARM was evaluated by comparing the ARM estimates to the observed sexual dimorphism by dividing an ARM estimate by the corresponding observed sexual dimorphism, expressed as percentage. An approximation of 100% indicates a perfect replication of ARM estimate of the observed value. Lower than 100% indicates an underestimation, while higher than 100% indicates an overestimation by the ARM estimate. This study adopts the criterion that a deviation up to 5% is acceptable, and considers unacceptable if an ARM estimate underestimates or overestimates the true sexual dimorphism by more than 5%.

Several sets of filters were compared to examine if the implemented filter (0.5 standard deviation) is most effective. All forty variables of three samples were analyzed without any filters, then with four sets of filters: 0.5 SD, 0.75 SD, 1 SD, and 2 SD. For each species, variables were grouped into three classes of approximation: (1) group of variables of which ARM estimates overestimate by more than 5%; (2) group of variables of which ARM estimates are within the acceptable range; and (3) group of variables of which ARM estimates underestimate by more than 5%. The number of variables that fall within the acceptable range of deviation was examined with respect to changes in filter settings.

To gauge the sensitivity of ARM to biased sex ratios, ratios of males to females were arbitrarily varied from 1:1 to 4:1 in both directions of males and females by randomly excluding a portion of a sample. For each sample of arbitrarily varied sex ratio, the ARM estimate of sexual dimorphism was compared to the observed sexual dimorphism. By randomly excluding specimens of known sex, samples of different sex ratios were generated: three samples had more males (two, three, and four times the number of females), and three samples had more females (two, three, and four times the number of males). For each of the samples, ARM estimates were compared to the respective observed sexual dimorphism. Changes in the degree of approximation between ARM estimates and the observed sexual dimorphism were examined.

As fossil data sample sizes are often small, it is a matter of importance to examine the effect of small sample size on ARM. To gauge the sensitivity of ARM to smaller sample size, specimens from the human sample were randomly excluded to generate samples of smaller size. The human sample was selected because it has the largest sample size, which allows for the widest range of possible sample sizes that can be manipulated. Because the original sample consists of unequal number of sexes (51 males and 40 females), it posed an added uncertainty of maintaining the sex ratio while varying the sample size. Therefore, the sex ratio was made to be equal (40 males, 40 females) by randomly excluding 11 males from the sample. Then the sample size was reduced into various sizes (80, 70, 60,

50, 40, 30, 20, 10, 8, 6, 4), while maintaining the equal sex ratio for each sample. For each of the samples, ARM estimate of sexual dimorphism was compared to the observed sexual dimorphism. Changes in the degree of approximation between ARM estimates and the observed sexual dimorphism were examined.

Results

The results of the analysis examining whether ARM provides reliable measures of sexual dimorphism are summarized in Figure 1. For each graph, the x-axis expresses the degree of approximation. In the human sample, 38 out of 40 variables fall within the acceptable range. The two variables that fall outside the acceptable range are mandibular canine socket height and breadth, both of which are overestimated to the extent of 6%. In the chimpanzee sample, 37 out of 40 variables fall within the acceptable range. The three variables that fall outside the acceptable range are mandibular canine socket breadth, midpoint shaft circumference of the tibia, and the minimum shaft circumference of the tibia, all of which are overestimated to the extent of 8%. In the gorilla sample, 39 out of 40 variables fall within the acceptable range. The variable that falls outside the acceptable range is mandibular corpus height at symphysis, which is underestimated to the extent of 6%. The ARM estimates of sexual dimorphism show acceptable approximation to the observed sexual dimorphism for the majority of 40 variables for all three samples, within 5% of deviation from the observed values.
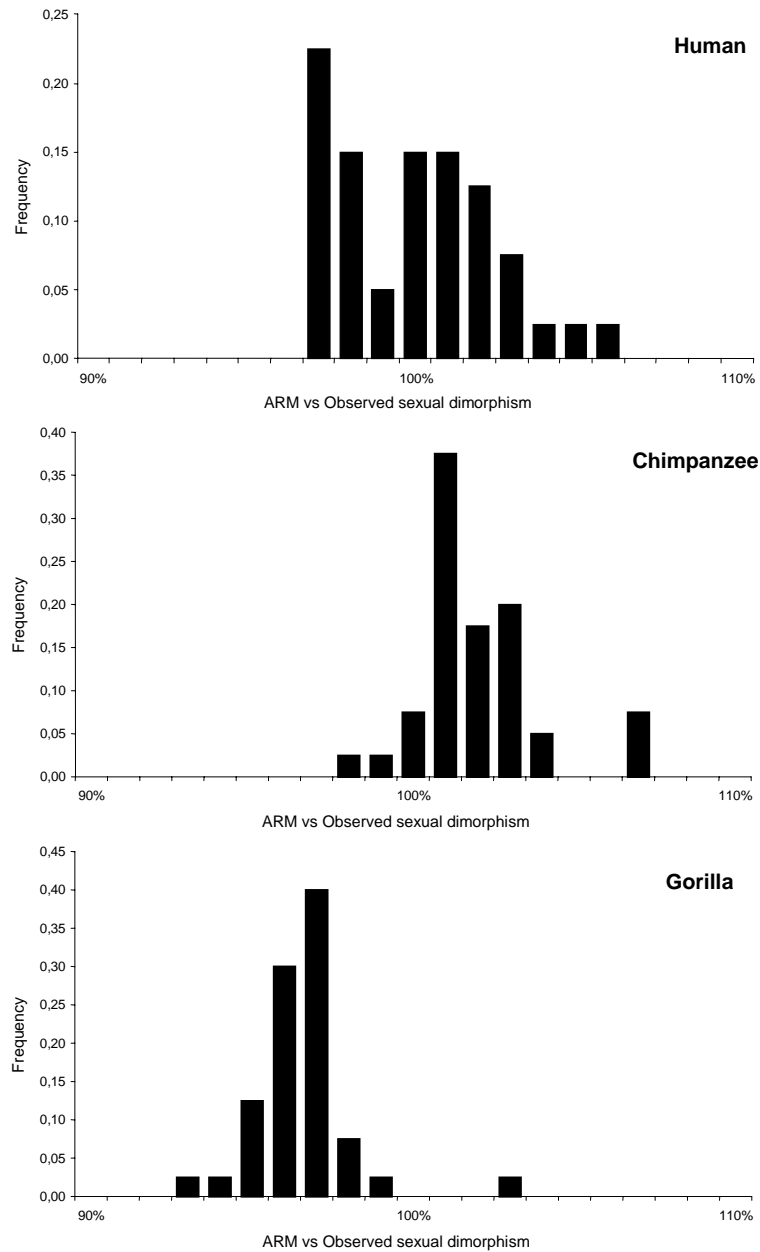
Fig. 1. Performance of ARM on three samples: human (n = 91), chimpanzee (n = 46), and gorilla (n = 56). The x-axes represent the degree of approximation expressed as ARM estimate value in percentage of observed sexual dimorphism. Values equal to 100% denote an exact replication; values less than 100% denote ARM estimate under-estimating the observed; values greater than 100% denote ARM estimate overestimating the observed. The y-axes represent the frequency of variables for each approximation interval.

The standard errors of the ARM estimates ranged from 0.00027 to 0.01584 in the gorilla sample, from 0.00004 to 0.009 in the chimpanzee sample, and from 0.00008 to 0.00576 in the human sample. Since the observed sexual dimorphism in the form of a ratio between the male mean and the female mean is not associated with a variance measure, a comparison was not possible between the ARM estimates and the observed.

For the human sample and for the chimpanzee sample, there was no difference in the performance of ARM among the different filters (data not shown). However, ARM performance changed with different filters when applied to the gorilla sample (Figure 2). When no filter was implemented for the gorilla sample, only six variables were within the acceptable range of 5% deviation, and 34 out of 40 variables are underestimated by more than 5%. When a filter with 0.75 SD (Filter 75) was used, 26 variables were within a 5% deviation, and 14 out of 40 variables were underestimated by more than 5%. The number of variables that are underestimated by more than 5%

increases to 28 when a filter with 1 SD (Filter 100) was used, and to 34 when a filter with 2 SD (Filter 200) was used. The results show empirically that the implemented filter of 0.5 SD (Filter 50) is most effective of the four different sets of filters. The implication of this analysis is discussed below.

The difference between the two measures of sexual dimorphism was smallest with an equal number of males and females (100.3% approximation). As males outnumber females, ARM estimate underestimated the observed sexual dimorphism, and the difference increased: when there were twice as many males as females (sex ratio of 2:1), ARM underestimated the observed (99.5%). When the ratio was 3:1 or 4:1, ARM underestimated as well (97.8% or 96.6%, respectively). When females outnumbered males, ARM estimates underestimated the observed sexual dimorphism, and the difference increased as the ratio increased. However, the difference was smaller than when males outnumbered females (when the sex ratios were 2:1, 3:1, and 4:1, the approxima-
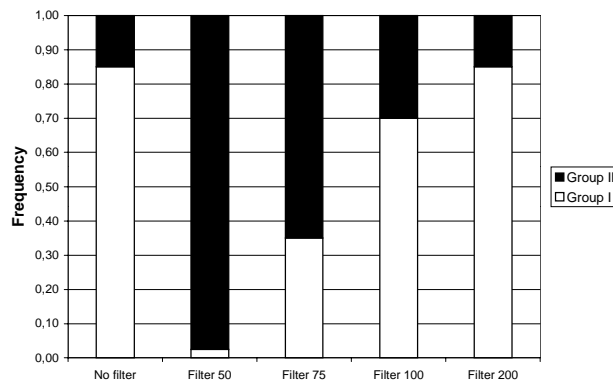

Fig. 2. Comparison of different filters on the gorilla sample. The x-axis represents five sets of results (see text for explanation). On the y-axis, the frequencies of variables are expressed, grouped into two groups of approximation: Group I that consists of variables that are underestimated more than 5%, and Group II that consists of variables that are within the acceptable range (no variable was overestimated more than 5% in the gorilla sample).

tion was 99.7%, 98.9%, and 97.9%, respectively). The amount of deviation did not exceed 5% in any case, but it is noted that bias in sex ratio has more effect on the performance of ARM (i.e., ARM underestimate more) when males outnumber females (see discussion below) than vice versa.

As the sample size changed from 80 to 6, the difference between the ARM estimates and the observed sexual dimorphism remained less than 2%, well within the acceptable threshold of deviation. When the sample size is 4, ARM estimate underestimated the most, 97.7%.

Discussion

The results of this study strongly suggest that ARM can be applied reliably to samples of humans, chimpanzees, and gorillas. It would be reasonable to assume that ARM is applicable to samples of extinct hominid species that could be reasonably assumed to have sexual dimorphism within the range exhibited among the three species.

The forty variables examined in this study include a wide range of metric variables. As ARM performed within the acceptable criteria for almost all of the forty variables, any metric variable seems to be eligible for ARM to be applied. The variables for which ARM performed less well do not appear to form any systematic pattern. It could be the case that variables with small values are less reliable: those variables that show more deviation between the ARM estimates and the observed sexual dimorphism have smaller values. Mandibular canine socket length and breadth in human samples have mixed-sex average of 5.9 mm and 5.8 mm, respectively,

the smallest averages of all variables. However, none of the three variables that fall outside the acceptable range for chimpanzees is the smallest in average for the chimpanzee sample: the mandibular canine socket breadth has the average of 11.2 mm, the midpoint shaft circumference of the tibia 63.5 mm, and the minimum shaft circumference of the tibia 58.1 mm. Similarly, the mixed-sex sample average for mandibular corpus height at symphysis in the gorilla sample is 41.7 mm, and is not of the smaller averages of the variables. In the absence of a convincing explanation at this point, random chance is as likely as any other explanation for the variables whose ARM estimates show more than 5% of deviation from the observed sexual dimorphism.

It is also possible that the observed reliability of the method is specific to the data sets used in this study and cannot be extended to other variables or other species. Future studies in empirical examinations of ARM regarding other species as well as theoretical examinations on the properties of ARM will provide answers to these questions.

One observed phenomenon merits further discussion. The exploration may in fact shed light on the theoretical properties of the ARM estimates. For example, the gorilla sample is characterized by a larger variance among the variables than the other two species (data not shown). Noting that the observed sexual dimorphism is a ratio of arithmetic means, it could be possible that the ARM estimates, mean of ratios, have the properties of a harmonic mean [LEE 2001]. Harmonic means, expressed as the reciprocal of the arithmetic mean of the reciprocal of the values of a set,

are less sensitive to variance than arithmetic means (P. Enflo, pers. comm.). Whether the ARM estimates are indeed harmonic means needs to be examined further, because this will have implications upon further application of the method to other variables with different properties of variance.

Conclusion

In understanding the evolution of hominid sexual dimorphism, fossils furnish the primary data source. Although several methods have been proposed to estimate the degree of sexual dimorphism in hominid fossils, none has fully addressed the problems imposed from using fossils as analytical samples. Most critical problems of unknown sex, fragmentary preservation, and small sample size have remained unresolved. This paper proposes an alternative, Assigned Resampling Method (ARM). The advantages of ARM lie in the heuristic aspect, in allowing direct comparisons of fragmentary materials without estimation of body size, and direct estimation of the degree of sexual dimorphism without depending on sex assessment of individual specimens. Using skeletal data sets in which individual specimens have independent records of sex, the ARM estimates are compared with the observed sexual dimorphism. Results of the comparison show that ARM provides reliable estimates of sexual dimorphism. Regardless of the assignment of sex in any particular drawing event, a sample of repeated drawing events results in an estimate that approximates the true sexual dimorphism. The effects of biased sex ratios and small sample size are explored using arbitrarily modified data sets. It is concluded that ARM provides a reliable alternative solution that addresses the problem of hominid fossil samples. With future research involving simulation studies, the theoretical basis of the method may be understood.

Acknowledgements

Appendix 1: ARM algorithm

The algorithm of ARM consists of the following steps:

(1) Two values, x_a and x_b, are randomly sampled with replacement from an empirical data set of size n [$x = (x_1, x_2, …, x_n)$].

(2) If *both* values of x_a *and* x_b are from the range of values more than mean + 0.5 SD, or if they are from the range of values less than mean - 0.5 SD, the pair is considered invalid.

(3) If not (2), x_a and x_b are compared. If x_b is larger than x_a, x_a and x_b form a ratio r by dividing x_b by x_a. Otherwise, x_a and x_b form a ratio r by dividing x_a by x_b.

(4) Steps (1), (2) and (3) are repeated numerous times until the size of r reaches $m = 1000$. This results in a distribution of ratios r_i ($i = 1, 2, …, 1000$).

(5) Sample mean of the above distribution is defined as the ARM estimate of sexual dimorphism.

(6) Steps (1) to (5) are repeated 100 times to yield the standard error of the ARM estimate.

Appendix 2: Measurement variables[*]

Description	Martin
Facial height: *nasion-prosthion*	M48
Facial breadth: distance between the most inferior points of zygomaxillary suture	M46(a)
Maximum cranial length: *glabella-opistocranion*	M1
Mandibular canine breadth	
Mandibular canine length	
Mandibular canine socket breadth	
Mandibular canine socket length	
Mandible corpus height at P4/M1	
Mandible corpus breadth at P4/M1	
Humerus length	M1
Maximum head diameter of humerus	
Midshaft circumference of humerus	M7(a)
Midpoint shaft diameter (A-P) of humerus	M5
Midpoint shaft diameter (M-L) of humerus	M6
Biepicondylar breadth (maximum) of humerus	M4
Minimum shaft circumference of humerus	M7
Breadth of the trochlea posterior face (ridge-ridge) at the base of fossa	
Breadth of the articular surface of the anterior face of humerus	
Maximum radius length	M1
Maximum head diameter of radius	
Midshaft circumference of radius	M5(5)
ML length, distal surface and styloid of radius	M5(6)
Maximum femur length	M1
Morphological length (standing on the condyles) of femur	
Perpendicular head diameter of femur	M18
Shaft circumference below lesser trochanter of femur	
AP diameter below lesser trochanter of femur	M10
ML diameter below lesser trochanter of femur	M9
Midshaft circumference of femur	M8
Midshaft AP diameter of femur	M6
Midshaft ML diameter of femur	M7(a)
Biepicondylar breadth of femur	M21
Distal articular surface breadth of femur	
Lateral condyle breadth at base midpoint of femur	M21(e)
Maximum tibial length	M1(a)
Maximum platform length of tibia	
Minimum platform breadth of tibia	
Midpoint shaft circumference of tibia	M10
Maximum ML distal end of tibia	M6
Minimum shaft circumference of tibia	M10(b)

* For complete description of the measurement variables, see LEE [1999].

[35]

References

ALBRECHT G.H., J.M.A. MILLER, 1997, *Variability profiles in fossil studies: the case of Homo habilis* (abstract), Am. J. Phys. Anthropol. Suppl., **24**, 65

ARSUAGA J.L., J.M. CARRETERO, C. LORENZO, *et al.*, 1997, *Size variation in Middle Pleistocene humans*, Science, **277**, 1086-1088

BENNETT K.A., 1981, *On the expression of sexual dimorphism*, Am. J. Phys. Anthropol., **56**, 59-62

BUIKSTRA J.E., D.H. UBELAKER (Eds.), 1994, *Standards for Data Collection from Human Skeletal Remains. Proceedings of a Seminar at the Field Museum of Natural History*, Arkansas Archeological Survey, Fayetteville, Arkansas

CHAKRABORTY R., P.P. MAJUMDER, 1982, *On Bennett's measure of sex dimorphism*, Am. J. Phys. Anthropol., **59**, 295-298

COHEN A.C., 1967, *Estimation in mixture of two normal distributions*, Technometrics, **9**, 15-28

COPE D.A., M.G. LACY, 1995, *Comparative application of the coefficient of variation and range-based statistics for assessing the taxonomic composition of fossil samples*, J. Hum. Evol., **29**, 549-576

DONG Z., 1997, *Mixture analysis and its preliminary application in archaeology*, J. Archaeol. Sci., **24**, 141-161

EFRON B., R.J. TIBSHIRANI, 1993, *An Introduction to the Bootstrap*, Chapman & Hall, New York

FLEAGLE J.G., R.F. KAY, E.L. SIMONS, 1980, *Sexual dimorphism in early anthropoids*, Nature, **287**, 328-330

GODFREY L.R., S.K. LYON, M.R. SUTHERLAND, 1993, *Sexual dimorphism in large-bodied primates: the case of the subfossil lemurs*, Am. J. Phys. Anthropol., **90**, 315-334

HAGER L.D., 1989, *The Evolution of Sex Differences in the Hominid Bony Pelvis*, University of California, Berkeley

HAGER L.D., 1991, *The evidence for sex differences in the hominid fossil record*, [in:] *The Archaeology of Gender*, D. Walde & N. Willows (eds.), Archaeological Association, University of Calgary, Calgary, 46-49

HAGER L.D., 1996, *Sex differences in the sciatic notch in great apes and modern humans*, Am. J. Phys. Anthropol., **99**, 287-300

HAMILTON M.E., 1982, *Sexual dimorphism in skeletal sample*, [in:] *Sexual Dimorphism in Homo sapiens*, R.L. Hall (ed.), Praeger, New York, 107-163

HÄUSLER M., P. SCHMID, 1995, *Comparison of the pelves of Sts 14 and AL 288-1: implications for birth and sexual dimorphism in australopithecines*, J. Hum. Evol., **29**, 363-383

JOHANSON D.C., T.D. WHITE, 1979, *A systematic assessment of early African hominids*, Science, **202**, 321-330

JOSEPHSON S.C., K.E. JUELL, A.R. ROGERS, 1996, *Estimating sexual dimorphism by method of moments*, Am. J. Phys. Anthropol., **100**, 191-206

JUNGERS W.L., 1988, *New estimates of body size in australopithecines*, [in:] *Evolutionary History of the "Robust" Australopithecines*, F.E. Grine (ed.), Aldine de Gruyter, New York, 115-125

KATZENBERG M.A., S.R. SAUNDERS (Eds.), 2000, *Biological Anthropology of the Human Skeleton*, Wiley, New York

KAY R.F., 1982, *Sexual dimorphism in Ramapithecinae*, Proc. Natl Acad. Sci. USA, **79**, 209-212

KIMBEL W.H., T.D. WHITE, 1988, *Variation, sexual dimorphism, and the taxonomy of Australopithecus*, [in:] *Evolutionary History of the "Robust" Australopithecines*, F.E. Grine (ed.), Aldine de Gruyter, New York, 175-192

KROGMAN W.M., M.Y. İŞCAN, 1986, *The Human Skeleton in Forensic Medicine*, C. C. Thomas, Springfield, IL

KURTÉN B., 1969, *Sexual dimorphism in fossil mammals*, [in:] *Sexual Dimorphism in Fossil Metazoa and Taxonomic Implications*, G.E.G. Westermann (ed.), Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 226-233

LEE S.-H., 1995, *Variability in Australopithecus afarensis: Does sexual dimorphism explain the humeral variability in Australopithecus afarensis?* (abstract), Am. J. Phys. Anthropol. Suppl., **20**, 132

LEE S.-H., 1999, *Evolution of Human Sexual Dimorphism: Using Assigned Resamplng Method to Estimate Sexual Dimorphism When Individual Sex is Unknown*, Ph.D. dissertation, University of Michigan, Ann Arbor

LEE S.-H., 2001, *Patterns of skeletal sexual dimorphism in human, chimpanzee, and gorilla* (abstract), J. Hum. Evol., **40**, A13

LEUTENEGGER W., B. SHELL, 1987, *Variability and sexual dimorphism in canine size of Australopithecus and extant hominoids*, J. Hum. Evol., **16**, 359-367

LIEBERMAN D.E., D.R. PILBEAM, B.A. WOOD, 1988, *A probabilistic approach to the problem of sexual dimorphism in Homo habilis: a comparison of KNM-ER 1470 and KNM-ER 1813*, J. Hum. Evol., **17**, 503-511

LOCKWOOD C.A., B.G. RICHMOND, W.L. JUNGERS, W.H. KIMBEL, 1996, *Randomization procedures and sexual dimorphism in Australopithecus afarensis*, J. Hum. Evol., **31**, 537-548

LOTH S.R., M. HENNEBERG, 1996, *The Taung child – it's a boy! Sexually dimorphic morphology in the immature human mandible and its application to fossil hominids* (abstract), Am. J. Phys. Anthropol. Suppl., **22**, 152

LOVEJOY C.O., K.F. KERN, S.W. SIMPSON, R.S. MEINDL, 1989, *A new method for estimation of skeletal dimorphism in fossil samples, with application to Australopithecus afarensis*, [in:] *Proceedings of the 2nd International Congress of Human Paleoanthropology*, Editoriale Jaca Books, Milan, 103-108

MARTIN L.B., 1983, *The Relationships of the Later Miocene Hominoidea*, University College, London

MARTIN R., K. SALLER, 1957, *Lehrbuch der Anthropologie*, Gustav Fischer Verlag, Stuttgart

MARTIN R.D., L.A. WILLNER, A. DETTLING, 1994, *The evolution of sexual size dimorphism in primates*, [in:] *The Differences Between the Sexes*, R. V. Short & E. Balaban (eds.), Cambridge University Press, Cambridge, 159-200

MCHENRY H.M., 1986, *Size variation in the postcranium of Australopithecus afarensis and extant species of hominoidea*, [in:] *Sexual Dimorphism in Living and Fossil Primates*, M. Pickford & B. Chiarelli (eds.), Il Sedicesimo, Firenze, 183-189

MCHENRY H.M., 1991, *Sexual dimorphism in Australopithecus afarensis*, J. Hum. Evol., **20**, 21-32

OXNARD C.E., D.S. LIEBERMAN, B.R. GELVIN, 1985, *Sexual dimorphisms in the dental dimensions of higher primates*, Am. J. Primat., **8**, 127-152

PEARSON K., 1894, *Contribution to the mathematical theory of evolution*, Philos. Trans. R. Soc. Lond. A, **185**, 71-110

PLAVCAN J.M., 1994, *Comparison of four simple methods for estimating sexual dimorphism in fossils*, Am. J. Phys. Anthropol., **94**, 465-476

PLAVCAN J.M., 2000, *Inferring social behavior from sexual dimorphism in the fossil record*, J. Hum. Evol., **39**, 327-344

RICHMOND B.G., W.L. JUNGERS, 1995, *Size variation and sexual dimorphism in Australopithecus afarensis and living hominoids*, J. Hum. Evol., **29**, 229-245

ROBERTSON C.A., J.G. FRYER, 1969, *Some descriptive properties of normal mixtures*, Scand. Actuarial J., **52**, 137-146

SCHWARTZ J.H., 1995, *Skeleton Keys: An Introduction to Human Skeletal Morphology, Development, and Analysis*, Oxford University Press, New York

SIEGEL M.I., 1976, *Comment on M.H. Wolpoff, Some aspects of the evolution of early hominid sexual dimorphism*, Curr. Anthropol., **17**, 600

SMITH R.J., 1980, *Rethinking allometry*, J. Theor. Biol., **87**, 97-111

SMITH R.J., 1996, *Biology and body size in human evolution*, Curr. Anthropol., **37**, 451-481

TAGUE R.G., C.O. LOVEJOY, 1998, *AL 288-1– Lucy or Lucifer: gender confusion in the Pliocene*, J. Hum. Evol., **35**, 75-94

TITTERINGTON D.M., A.F.M. SMITH, V.E. MAKOV, 1985, *Statistical Analysis of Finite Mixture Distributions*, J. Wiley & Sons, New York

TRINKAUS E., 1976, *Comment on M.H. Wolpoff, Some aspects of the evolution of early hominid sexual dimorphism*, Curr. Anthropol., **17**, 600-601

VITZHUM V.J., 1990, *Odontometric variation within and between taxonomic levels of Cercopithecidae: implications for interpretations of fossil samples*, Hum. Evol., **5**, 359-374

WELSH A.H., A.T. PETERSON, S.A. ALTMANN, 1988, *The fallacy of averages*, Am. Nat., **132**, 277-288

WOLPOFF M.H., 1975, *Sexual dimorphism in the australopithecines*, [in:] *Paleoanthropology: Morphology and Paleoecology*, R.H. Tuttle (ed.), Mouton, The Hague, 245-284

WOLPOFF M.H., 1976a, *Primate models for australopithecine sexual dimorphism*, Am. J. Phys. Anthropol., **45**, 497-510

WOLPOFF M.H., 1976b, *Some aspects of the evolution of early hominid sexual dimorphism*, Curr. Anthropol., **17**, 579-606

WOOD B.A., 1976, *The nature and basis of sexual dimorphism in the primate skeleton*, J. Zool. (Lond), **180**, 15-134

ZIHLMAN A.L., 1985, *Australopithecus afarensis: two sexes or two species?* [in:] *Hominid Evo-lution: Past, Present, and Future. Proceedings of the Taung Diamond Jubilee International Symposium*, P.V. Tobias (ed.), Alan R. Liss Inc., New York, 213-220

Streszczenie

U żyjących obecnie gatunków naczelnych znaczny dymorfizm płciowy skorelowany jest z czynnikami ekologicznymi, takimi jak proporcja płci i system kojarzeń. Chociaż nad kore-lacją taką u gatunków kopalnych badania jeszcze trwają [PLAVCAN 2000], uważa się, że dymorfizm płciowy jest jedną z głównych składowych zmienności w populacji, a w związku z tym określenie zmian jego natężenia w czasie jest zadaniem o fundamentalnym znaczeniu, które powinno być zrealizowane poprzez badanie materiałów kopalnych. Jednakże pewne cechy materiału kopalnego ograniczają możliwości wykorzystywania go do analizy dymorfi-zmu płciowego.

Przede wszystkim płeć okazu kopalnego prawie nigdy nie może być określona w sposób pewny, a często jest to po prostu niemożliwe. Fragmentaryczność znalezisk ogranicza moż-liwości porównywania elementów anatomicznych, ponieważ elementu obecnego na jednym fragmencie może brakować na innym. Przedstawiona praca zawiera przegląd metod stoso-wanych dotąd w celu oceny rozmiarów dymorfizmu płciowego w próbach kopalnych oraz nową propozycję omijającą ograniczenia dotychczasowych metod.

Problem oszacowania dymorfizmu płciowego w próbach kopalnych próbowano rozwiązać na kilka sposobów: od opartych na rozstępie między wartościami minimalnymi i maksymal-nymi [ARSUAGA *et al.* 1997; JOHANSON i WHITE 1979; JUNGERS 1988; LEE 1995; MCHENRY 1986; RICHMOND i JUNGERS 1995; ZIHLMAN 1985], bimodalności rozkładów [WOLPOFF 1975, 1976a,b], czy współczynniku zmienności [FLEAGLE *et al.* 1980; KAY 1982; KIMBEL i WHITE 1988; LEUTENEGGER i SHELL 1987; LOCKWOOD *et al.* 1996; MCHENRY 1991; PLAVCAN 1994; WOOD 1976], do bardziej skomplikowanych, jak „*technique dimorphism*" [LOVEJOY *et al.* 1989] i „*finite mixture analysis*" [DONG 1997; GODFREY *et al.* 1993; JOSEPHSON *et al.* 1996].

Wszystkie dotychczasowe metody oceny wielkości dymorfizmu płciowego mają ograni-czenia w stosowaniu do materiałów kopalnych. Jedno z tych ograniczeń polega na tym, że wybór metody w dużym stopniu zależy od uprzedniej wiedzy na temat dymorfizmu płciowe-go w próbie, która ma być zbadana.

Proponowana w tej pracy nowa metoda – metoda powtarzanych losowań par wartości zmiennej nazwana „*Assigned Resampling Method*" (ARM), ocenia rozmiary dymorfizmu płciowego w próbie o nieoznaczonej płci poprzez utworzenie rozkładu wzajemnych propor-cji w parach pomiarów wielokrotnie losowanych (ze zwrotem) z próby o wspólnym dla obu zmieszanych płci rozkładzie. Wspomnianą proporcję losowanej pary pomiarów oblicza się dzieląc wartość większą przez mniejszą. W celu skompensowania ewentualnego niedosza-cowania wartości dymorfizmu (proporcji wielkości pomiarów męskich do żeńskich) zasto-sowano „filtr" polegający na odrzucaniu wyników dla par, w których obie wylosowane wartości pochodziły z tego samego krańca zmienności mieszanego rozkładu (± 0,5 SD). Średnia rozkładu proporcji w parach jest w metodzie ARM oszacowaniem dymorfizmu płciowego.

Dokładność i czułość metody ARM przetestowano na próbach współczesnego człowieka, szympansa i goryla. Wykorzystano 40 pomiarów metrycznych wykonanych na każdym osobniku każdej z prób. Wyniki analizy (rys. 1) pokazały, że oszacowanie dymorfizmu metodą ARM dla zdecydowanej większości spośród 40 zbadanych zmiennych dało satysfakcjonujące wyniki we wszystkich trzech próbach. Porównanie wyników z zastosowaniem różnych filtrów wykazało (rys. 2), że proponowany filtr 0,5 SD jest najskuteczniejszy. Wykazano, że oszacowania dymorfizmu w próbach o różnych proporcjach płci są zadawalające, gdy proporcja ta nie jest bardziej zniekształcona niż 1:2 (liczebność jednej płci nie przeważa nad drugą więcej niż dwukrotnie). Najlepsze oszacowania metoda ARM zapewnia przy liczebnościach prób nie mniejszych od 20. Konkludując, metoda ARM może być z powodzeniem zastosowana do prób złożonych z kopalnych hominidów.