AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 121:230–240 (2003)

Understanding Muscle Markers: Aggregation and Construct Validity

Elizabeth Weiss*

Environmental Dynamics Program, University of Arkansas, Fayetteville, Arkansas 72701

KEY WORDS musculoskeletal stress markers; osteophytes; composites; sex; robusticity; age

ABSTRACT Musculoskeletal markers are frequently used to reconstruct past lifestyles and activity patterns. Yet, the reliability of muscle marker measurements has been called into question because they allegedly fail to correlate with cross-sectional properties and exercise patterns, and are confounded by body size. In this study, the principle of aggregation was used to sum muscle markers over 7 insertion sites (4 humeral, 2 radial, and 1 ulnar) and examine the effects on them of body size, age, sex, and cross-sectional properties. Analyses were made of a sample of 91 (66 males, 25 females) Native British Columbians (3500–1500 years BP) and 18th century Quebec prisoners. Muscle markers were measured using three-point observer rating scales; size was measured by standard methods; age and sex were determined through pelvic, cranial, and dental morphology; and cross-sectional properties were calculated from radiographs. Whereas any single muscle marker component failed to correlate with age, size, sex, or cross-sections, aggregate muscle marker correlated with: age, r = 0.49; size, r = 0.38; sex, r = 0.40; and, cross-sections, r = 0.38; $P < 0.001$. Older individuals had greater muscle markers, as did larger individuals, males, and those with more robust cross-sections. Based on partial correlations and regression analyses, age was the best overall predictor of aggregate muscle marker. Am J Phys Anthropol 121:230–240, 2003.

Musculoskeletal stress markers and osteophytes (in this study referred to as "muscle markers") are distinct skeletal markings and bony projections that occur where a muscle, tendon, or ligament inserts into the blood-supplying periosteum and the underlying bony cortex. Large muscle markers are typically viewed as being the direct result of continued muscle use in daily and repetitive tasks (Hawkey and Merbs, 1995; Lai and Lovell, 1992; Nagy, 1998). Evidence from bone remodeling studies supports this perspective. When muscle insertion sites are subjected to stress, blood flow is increased, which stimulates bone-forming cells, resulting in bone hypertrophy and increased size of musculoskeletal stress markers (Chamay and Tschantz, 1972; Woo et al., 1981; Wolff, 1892).

Muscle markers have been used to reconstruct past lifestyles of populations, such as whether males and females differed in activity patterns, effects of shifts in subsistence patterns, and specific activities related to hunting and fishing (Chapman, 1997; Cook and Dougherty, 2001; Hawkey, 1998; Lai and Lovell, 1992; Nagy, 1999; Peterson, 1998). Probably the best established findings in the muscle marker literature relate to age and sex differences. Older individuals have greater muscle markers than do younger individuals, which many anthropologists relate to the stress of activity patterns that accumulates over time (Chapman, 1997; Nagy, 1998; Robb, 1998; Wilczak, 1998). Researchers using muscle markers to reconstruct past lifestyles frequently take age differences into account to enable more accurate reconstructions (e.g., Hawkey and Merbs, 1995; Nagy and Hawkey, 1995).

In most skeletal samples, males have greater muscle markers than do females (e.g., Cohen, 1989; Cook and Dougherty, 2001; Hawkey and Street, 1992; Nagy, 1999; Steen and Lane, 1998). However, some studies found certain muscle markers to be higher in females than in males (Chapman, 1997; Nagy and Hawkey, 1995). Sex differences are frequently attributed to differences in activity patterns (Chapman, 1997; Peterson, 1998; Wilczak, 1998). For example, Cook and Dougherty (2001) examined adults from Chirikof Island in Alaska in the 18th century and found that males had greater upper limb muscle markers than did females, which they attributed to the extensive rowing that males engaged in while hunting marmots. Yet it may be that some sex differences are due to differences in body

---

*Correspondence to: Elizabeth Weiss, 1265 Richmond St., Suite 1710, London, Ontario N6A 3M1, Canada.
E-mail: eweiss@anthrosciences.com

Received 28 February 2002; accepted 18 October 2002.

DOI 10.1002/ajpa.10226

size, especially in cases where body size is not controlled for.

Some anthropologists have begun to examine the effects of body size on muscle markers. For example, Zumwalt et al. (2000) examined lower and upper limb bones from nonhuman primates and found that muscle markers correlated with body weight and did not vary with locomotor type, raising the question of whether human research should also take body size into account. Human upper limbs, however, are unique because they are free of locomotor responsibilities and, as a result, are not weight-bearing. This lack of weight-bearing by human upper limbs may have decreased the influence of size on their muscle markers.

Another problem with using muscle markers as an index of activity is the lack of correlation with activity in experimental tests. In one study, Zumwalt et al. (2001) exercised sheep for 3 months and compared seven muscle sites with a nonexercised group. Only 1 of 7 muscle insertion sites was influenced by exercise. This lack of findings may be due to differences in animal muscles (e.g., variation in fast/slow twitch muscles) and bones (e.g., lack of secondary osteons) compared to humans. However, perhaps experiments with more strenuous exercise or with longer exercise periods could show stronger results.

Other difficulties with using muscle markers to infer activity patterns include the lack of consistent relations with cross-sectional geometries, which is another measure of activity patterns. For example, Berget and Churchill (1994) found that Aleut muscle marker data fit well with the cross-sectional geometry of arm bones. Their results supported conclusions made by Hawkey and Street (1992) about sex differences in Aleut activity patterns, in which males engaged in ocean rowing for hunting big game and females prepared hides, weaved, and rowed umiaks. Other anthropologists, however, did not find any relation between the muscle markers and cross-sectional properties (Bridges, 1997; Stirland, 1998). Bridges (1997) cautioned that muscle markers are not good indicators of overall activity, but may still be useful for reconstructing specific activities. Stirland (1998) and Jurmain (1990, 1999) concluded that muscle markers were typically too subjective a measure to be used in reconstructing lifestyle activities. Work by Hawkey and Merbs (1995) and Robb (1998) responded to some of these criticisms by formulating less subjective ways of collecting muscle marker data.

The present study attempts to solve some of these inconsistencies reported in the literature by examining the effects of size, age, sex, and cross-sectional robusticity in two populations using simple and aggregated muscle markers. The principle of aggregation is widely used in economics (e.g., Khamis and Hempstead, 1996), psychology (e.g., Dunn et al., 1993), and other social sciences (e.g., Van Rompaey et al., 1999) because it creates significant correlations where nonsignificant findings hampered predictability. Aggregation is similarly used in medical research to increase correlations and enhance predictability in order to identify high-risk individuals (e.g., Lord and Fitzpatrick, 2001). This is because by aggregating over several measures, error variance and specificity or idiosyncratic variance can be averaged out, leaving only "true score variance" (if there is any) to cumulate (Rushton et al., 1983; Spearman, 1904, 1910).

In order to explain aggregation, Spearman (1904) drew the analogy of repeatedly firing a gun while trying to hit a target. The bullets scatter randomly around the target, more of them hitting closer to the target than farther away from it, and the more shots fired, the greater the number of bullets that hit the target. The scatter of bullets around the target is analogous to error variance, which is a part of every measure. Spearman's formulization of aggregation states that every actual measurement, call it X, is composed of two parts: a "true score," t, and an error variance, e. Both t and e cannot be directly observed. Thus, $X = t + e$. Since e can have either a positive or negative sign and because e is random, e's value tends towards zero as more measurements of X are averaged in. In other words, by adding a large number of Xs, e is averaged out, leaving only t or the "true score" to accumulate. (More familiarly, the same principle is used when taking an average.) The Spearman-Brown formula may be used to demonstrate the increase in reliability (and thus validity) of the principle of aggregation. This shows that increasing the length of a measure by adding items increases reliability from r to $r' = (1 + n/N)/(r^{-1} + n/N)$, where N is the number of items in the original measure, n is the number of new items added to it, and r is reliability (which is always less than 1) (Spearman, 1927). Of course, aggregation is not useful in all cases, and whether to aggregate is, in the end, a choice the researcher must make (see Discussion).

The prediction is tested that the more muscle markers that are aggregated, the greater will be their correlation with body size, sex, age, and cross sections. In this study, humeral size is used as a proxy for body size. If this body size prediction is confirmed, it may call into question the accuracy of reconstructing past lifestyles through muscle markers without controlling for body size. In addition, it was tested whether other skeletal features, such as age and sex, correlate with muscle markers, as other anthropologists have found. Greater muscle markers among older individuals may signal that these markers have accumulated over life due to activity patterns. Differences between the sexes may be related to sexual division of labor or size. Finally, hypotheses will be evaluated concerning associations between muscle markers and cross-sectional robusticity, which has also been used to reconstruct lifestyles.

TABLE 1. Sample size, where left refers to number of individuals with only left arm bones, right refers to number of individuals with only right arm bones, both refers to number of individuals with both right and left arm bones available, and total refers to total number of individuals[1]

	Males			Females			
	Left	Right	Both	Left	Right	Both	Total
BC Amerinds	13	11	17	8	8	7	64
Euroamericans	7	10	8			2	27
Total	20	21	25	8	8	9	91

[1] BC, British Columbia.

TABLE 2. Sample demographics[1]

	Dates (in years BP)	Age range (in years)	Average age (in years)	Number of males	Number of females
BC Amerinds	3,500–1,500	18–69	30.6	41	23
Euroamericans	200	18–69	29.4	25	2

[1] BC, British Columbia.

MATERIALS AND METHODS

Sample

A skeletal sample of 91 adult individuals (66 males; 25 females) was examined, ranging from age 18–69 years from two populations (British Columbian Amerinds, Euroamericans), housed at the Canadian Museum of Civilization at Hull, Quebec (Tables 1, 2). Cybulski (1990, 1992, 1999) sexed the individuals using pelvic and cranial indicators. He also aged them by means of the pubic symphysis, ilium auricular surface, cranial suture closure, dental development, and epiphyseal union of the long bones, clavicles, vertebrae, and innominates (Cybulski, 1990, 1992, 1999). For the present study, individuals were excluded if they were not sexed or aged, if they lacked complete arm bones, if they were immature, and if they were severely pathological. Whenever possible, measures were taken for both the right and left arm bones. The humerus was x-rayed at 35% of bone length.

The British Columbian skeletal remains come from seven Prince Rupert Harbor sites located in the traditional territories of the Amerind tribes belonging to the Tsimshian language family and the Northwest Pacific Coast cultural area. The Tsimshin were fishers and gatherers during the short summers and whale hunters in winter (Cybulski, 1990; Weiss, 2001). These remains date from 3500–1300 years BP. The Euroamerican skeletal remains come from English prisoners of war who died after being captured by French Canadians about 200 years ago (Piedalue and Cybulski, 1997; Weiss, 2001).

Methods

This study used z-scores to create three composite variables: Aggregate Muscle Marker, Humeral Size, and Robusticity. For missing data on any of the elements that went into the composite, an average score based on all remaining data on that variable was substituted (8% of the data was replicated by means). This procedure tends to homogenize scores, thereby reducing differences and lessening the chance of getting a significant effect.

The Aggregate Muscle Marker composite was created by adding the z-scores for 42 component variables (a total of 7 muscle markers from 4 humeral insertion sites, 2 radial insertion sites, and 1 ulnar insertion site, with each marker scored in the three categories of ruggedness, stress lesions, and ossification, and with right and left upper limb bone scores added together). Deltoid, latissimus dorsi, pectoralis major, and teres major muscle markers were scored on the humerus (Fig. 1). Biceps brachii and supinator muscle markers were scored on the radius (Fig. 2). The triceps brachii muscle marker was scored on the ulna (Fig. 3). The tendinous muscle sites are the deltoid, latissimus dorsi, teres major, biceps brachii, supinator, and triceps major. The muscle to bone insertion site is the pectoralis major. These sites were chosen because: 1) they are easily distinguishable; 2) they have been associated with specific activities in the literature (e.g., Nagy, 1999); and 3) they have been used frequently in lifestyle reconstruction (e.g., Capasso et al., 1999; Kennedy, 1983, 1989; Merbs, 1983; Robb, 1998; Peterson, 1998).

The methods employed by Hawkey and Merbs (1995) for characterizing muscle markers were used on the seven muscle sites. These methods were chosen because: 1) the interobserver and intraobserver error rates are low; 2) the scoring establishes identifiable thresholds for each score; and 3) the guidelines for scoring muscle markers are straightforward, with photographs illustrating various scores. Each of the muscle insertion sites was scored on three dimensions: 1) robusticity (which is referred to as ruggedness in this study to avoid confusion with cross-sectional robusticity), 2) stress lesions, and 3) ossification. Within these three categories, there are four specific grades, with the absence of the expression being grade 0.

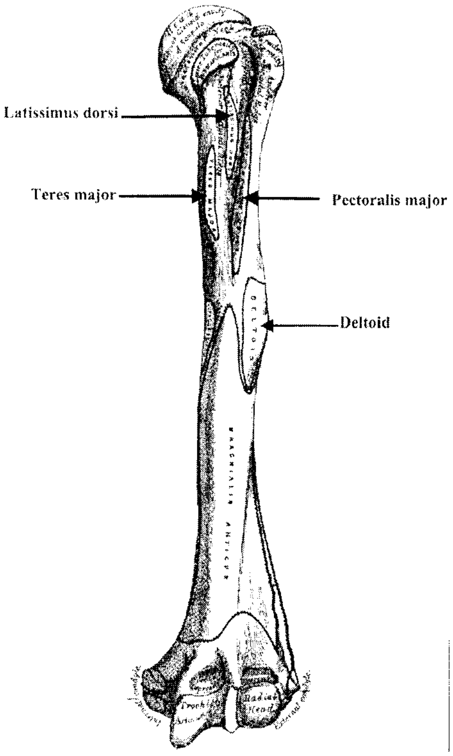

Fig. 1. Muscle markers on left humerus, anterior view (taken from Gray, 1977).

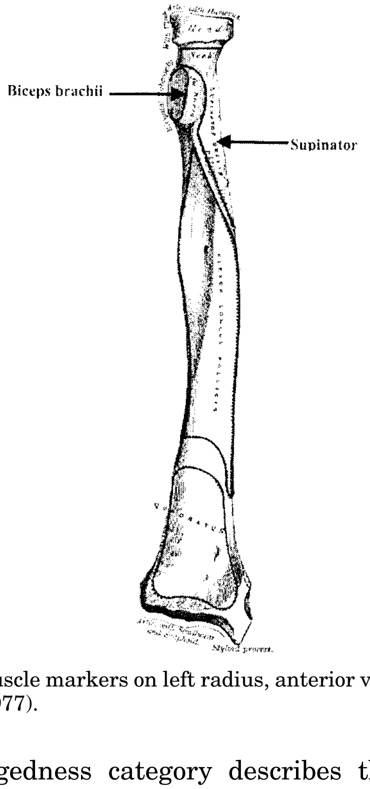

Fig. 2. Muscle markers on left radius, anterior view (modified from Gray, 1977).

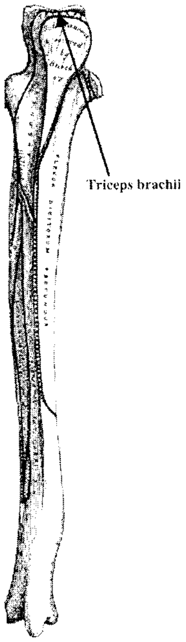

Fig. 3. Muscle marker on left ulna, posterior view (modified from Gray, 1977).

The ruggedness category describes the normal variation in areas where muscles attach. In ruggedness grade 1 (R1), the outer portion of the bone is only slightly rounded with elevation apparent when touched, although no distinct crests or ridges are present. In ruggedness grade 2 (R2), the outer portion of the bone is uneven, with a mound-shaped elevation clearly visible. In ruggedness grade 3 (R3), distinct sharp crests or ridges are present and there may be a small depression between crests, although this depression does not extend into the cortex or the outer portion of bone.

The stress lesion category is defined as pitting into the cortex. Stress lesion grade 1 (S1) is shallow pitting into the cortex, less than 1 mm in depth. In stress lesion grade 2 (S2), the pitting is between 1–3 mm in depth and covers a greater surface area, although not longer than 5 mm. In stress lesion grade 3 (S3), pitting is greater than 3 mm in depth and more than 5 mm in length.

The ossification category is used for scoring osteophytes. In ossification grade 1 (OS1), a slight exostosis (or bony spur) occurs, which is usually rounded and extends less than 2 mm from the cortex. In ossification grade 2 (OS2), a distinct bony spur is found, which is more than 2 mm but less than 5 mm in length. In ossification grade 3 (OS3), the bony spur extends more than 5 mm from the cortex and/or covers an extensive amount of surface.

A composite variable of Humeral Size was created by adding the three z-scores for humeral length, humeral vertical head diameter, and humeral epicondylar breadth (Fig. 4; Buikstra and Ubelaker, 1994). Humeral Size was calculated with data that were side-averaged following Ruff and Larsen (1990). These component traits are good proxies for body size, because they do not remodel (Ruff et al., 1991).

The humeral head was measured in millimeters, using a sliding caliper. The humeral head measure is the “direct distance between the superior and

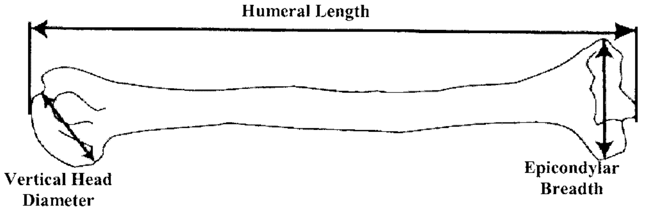

Fig. 4. Measurements of left humerus, anterior view (modified from Buikstra and Ubelaker, 1994).

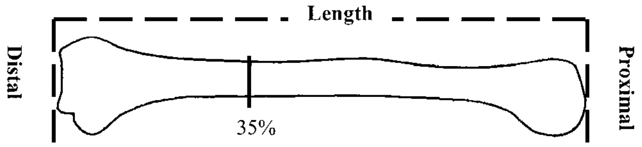

Fig. 5. Humeral location used to calculate cross-sections (adapted from Ruff and Larsen, 1990).

inferior points on the border of the articular surface" (Fig. 4; Buikstra and Ubelaker, 1994, p. 80). The humeral epicondylar breadth was measured in millimeters, using an osteometric board. This epicondylar measurement is the "distance of the most laterally protruding point of the lateral epicondyle from the corresponding projection of the medial epicondyle" (Fig. 4; Buikstra and Ubelaker, 1994, p. 80). Humeral length was measured in millimeters with an osteometric board. The humeral length measure is the "direct distance from the most superior point on the head of the humerus to the most inferior point on the trochlea" (Fig. 4; Buikstra and Ubelaker, 1994, p. 80).

A composite Robusticity variable was created by averaging the z-scores for five humeral cross-section variables: total area, cortical area, moment of inertia about the anteroposterior axis, moment of inertia about the mediolateral axis, and polar moment of inertia (all of which were taken at 35% of bone length, side-average, standardized by body size, and log-transformed).

In order to calculate cross-section variables, several steps were performed. Following Ruff and Larsen (1990), humeral lengths were used to mark the specific location on the humeri where cross-sectional geometry is calculated, i.e., 35%, of bone length, in which 100% is the proximal end (Fig. 5). Humeri were first x-rayed in the anteroposterior (AP) orientation (i.e., the anterior side of the bone faces upward), and then in the mediolateral (ML) orientation (i.e., the medial side of the bone faces upward). Before removing the bones from the x-ray machine, the magnification factor was calculated to remove any magnification errors. This was done using the formula

$$\text{Source-to-film distance}/(\text{Source-to-film distance} - \text{Object-center-to-film distance})$$

where source-to-film is the distance from the x-ray bulb to the film, and object-center-to-film is the distance from the middle of the bone to the film, which varied depending on the bone and position of the bone being x-rayed.

Next, each radiographed humerus was measured for inner and outer diameters at both orientations. These values allowed calculation of the derived values. The derived values for compressive strength are the total cross-sectional area (TA) and cortical cross-sectional area (CA). Total area is calculated with the formula:

$$\text{TA} = \Pi \times (\text{ML outer diameter} \times \text{AP outer diameter} / 4).$$

Medullary area (MA) is calculated with the formula:

$$\text{MA} = \Pi \times (\text{ML inner diameter} \times \text{AP inner diameter} / 4).$$

Then the cortical area is calculated by subtracting the medullary area from the total area (CA = TA − MA) (Biknevicius and Ruff, 1992).

For bending strength in particular planes, resultant values are anatomically oriented moments of inertia calculated about the mediolateral and anteroposterior axes (Iml and Iap). The formulas for Iml and Iap require the several steps sketched out in Biknevicius and Ruff (1992) and Fresia et al. (1990), and detailed in Timoshenko and Gere (1972). In short, the formulas for Iml and Iap are:

$$\text{Iml} = \Pi / 64 \times (\text{Tml} \times \text{Tap}^3 - \text{Mml} \times \text{Map}^3)$$

$$\text{Iap} = \Pi /64 \times (\text{Tap} \times \text{Tml}^3 - \text{Map} \times \text{Mml}^3)$$

where Tml = total mediolateral breadth; Tap = total anteroposterior breadth; Mml = medullary mediolateral breadth; and Map = medullary anteroposterior breadth.

Polar moment of inertia, which allows a determination of strength against overall bending and torsional strains and stresses (Runestad et al., 1993), is calculated using the formula:

$$\text{J} = \text{Iml} + \text{Iap}.$$

For all of these humeral geometric properties, percentage differences between the right and the left sides had to be calculated using the formula [right − left/right] on data not standardized by body size. Then the geometric properties were adjusted by this factor, depending on the direction of asymmetry, and whether the bone was from the left or the right side. For individuals with missing data for either the right or left humerus, mean bilateral asymmetry values for that sex and population were used. The averaged data were then standardized by body size, using the formulas described below.

Following Ruff et al. (1993), the resultant values are presented standardized for body size. Ruff et al. (1993) provided effective formulas for standardizing humeral areal (cortical area, CA and total area, TA) and inertial (moments of inertia, I and J) values. Areal measures (CA and TA) were standardized by dividing the result by humeral length cubed (HL^3) and then multiplying it by 10^8. Moments of inertia and polar moments of inertia (I and J) were stan-

TABLE 3. Means, SDs, and sample sizes for composite Aggregate Muscle Marker, Humeral Size, and Robusticity (z-scores), separately for males and females

Property	Sex	Mean	SD
Aggregate Muscle Marker	Males (n = 66)	4.730	18.9
	Females (n = 25)	−12.487	14.3
Humeral Size	Males (n = 66)	0.415	0.5
	Females (n = 25)	−1.094	0.4
Robusticity	Males (n = 66)	0.369	0.9
	Females (n = 25)	0.123	0.9

dardized by dividing the result by $HL^{5.33}$ and then multiplying it by 10^{12}. Finally, the data were log-transformed and converted to z-scores.

To allay concerns over averaging left and right arm bones (since bilateral asymmetry is common in arm bones), correlations for the right and left arm bones were carried out. Pearson correlations between the various size properties for the left and right arm bones ranged from 0.88–0.94 (mean r = 0.90, $P < 0.01$). Pearson correlations between the various cross-sectional properties for the left and the right arm bones ranged from 0.88–0.97 (mean r = 0.93, $P < 0.01$). The Pearson correlation between the left and right Aggregate Muscle Marker variable was r = 0.47, $P < 0.01$. With these high correlations, the author felt justified in combining left and right arm bones for the purpose of this study.

Statistical analysis

The data were analyzed using the statistical software program SPSS (version 10.0).

The data used here were tested for violations of assumptions of parametric tests. Independence of the variants will be handled through the use of adjusted error rates with critical alpha levels varying depending on the number of correlations per matrix; the formula used was (0.10/Number of tests) (Weiss and Hassett, 1982). Homogeneities of variances were tested through the Levene statistic test. A significance level greater than 0.05 signifies homogeneous variance. All of the variables were homogeneous (significances ranged from 0.10–0.72). To test whether the data are normally distributed, the Kolmogorov-Smirnov test was run, which is an appropriate test for skeletal samples because it "focuses on the greatest observed differences between two cumulative frequency distribution; loss of information from tied data is minimized" (Lovejoy, 1971, p. 105). Additionally, the Kolmogorov-Smirnov test combines high power efficiency (about 0.96) with a central expression of skewness, central tendency, and dispersion (Lovejoy, 1971). A significance value less than 0.05 indicates that the distribution of the data differs significantly from a normal distribution. All variables were normally distributed (significances ranged from 0.70–1.09). Thus, the aggregate measures met all the assumptions required to run parametric tests, and the relationships between variables were linear.

For each composite variable, means and standard deviations were calculated. Composite variable Aggregate Muscle Marker was correlated using two-tailed Pearson tests with both composite variables Humeral Size and Robusticity along with age (defined in six groups: 1 = 18–24 years old; 2 = 25–31 years old; 3 = 32–38 years old; 4 = 39–45 years old; 5 = 46–52 years old; and 6 = 53+ years old) and sex (Weiss and Hassett, 1982). Pearson tests were run separately for males and females on correlations between Aggregate Muscle Marker, Humeral Size, and Robusticity. Partial Pearson correlations controlling for age, sex, and Humeral Size were also run to determine causes of muscle markers. Critical alpha levels varied depending on the number of correlations per matrix; nonsignificant findings are marked "n.s." Additionally, a stepwise regression analysis (with an ANOVA to show significant variances) was run to test which factor best predicted Aggregate Muscle Marker scores (McCall, 1990).

RESULTS

Table 3 presents the means, standard deviations, and sample sizes for the composite variables used in this study, i.e., Aggregate Muscle Marker, Humeral Size, and Robusticity in z-scores. Aggregate Muscle Marker correlates significantly with age, r = 0.489; Humeral Size, r = 0.378; sex, r = 0.400; and Robusticity, r = 0.384; $P < 0.001$. (When disaggregating the side-averaged data for the right arm bones, Aggregate Muscle Marker correlates significantly with age, r = 0.400; Humeral Size, r = 0.361; sex, r = 0.291; and Robusticity, r = 0.253; $P < 0.01$; for the left arm bones, Aggregate Muscle Marker correlates significantly with age, r = 0.361; Humeral Size, r = 0.483; sex, r = 0.303, $P < 0.01$; and Robusticity, r = 0.160; n.s.).

When the sexes are separated, Aggregate Muscle Marker correlates significantly with age (r = 0.510, $P < 0.001$) and Robusticity (r = 0.381, $P < 0.001$) for males, but only marginally so in females (age, r = 0.377, $0.10 < P < 0.05$; Robusticity, r = 0.336, $0.10 < P < 0.05$). Aggregate Muscle Marker is also related to Humeral Size when the sexes are combined, although not to either sex when analyzed separately (males, r = 0.084, n.s.; females, r = 0.194, n.s.).

Figure 6 illustrates the increasing correlation between muscle markers and age, size, sex, and robusticity as a function of the number of muscle markers aggregated. For each variable, the correlation becomes higher with the number of muscle markers aggregated. However, the correlations level off after

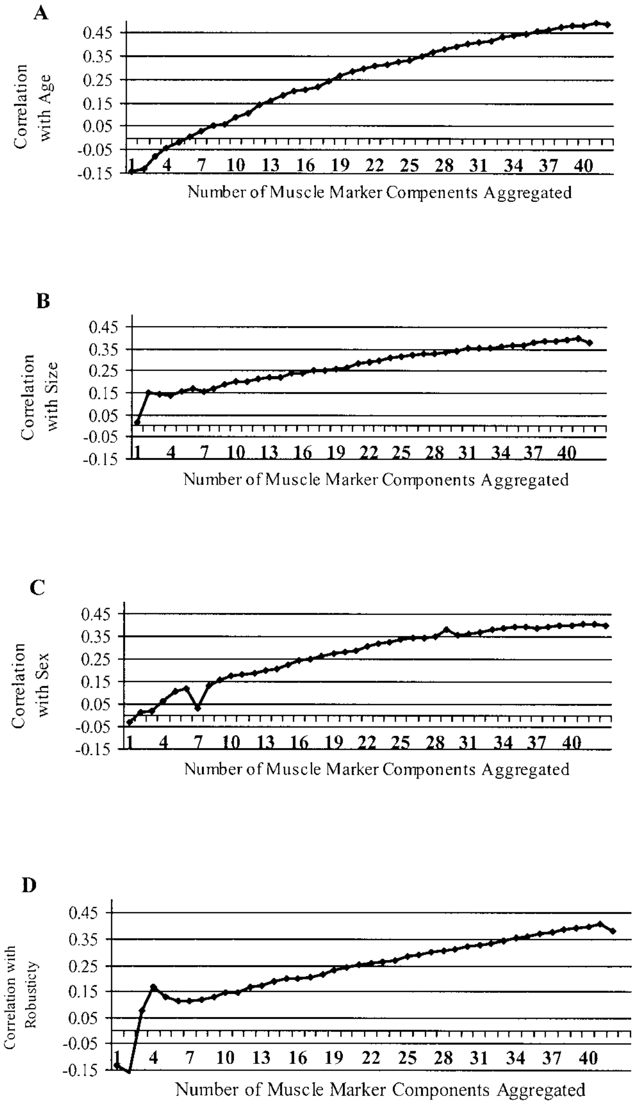

Fig. 6. Aggregation: as number of muscle marker components increase, so does relationship with age (**A**), size (**B**), sex (**C**), and robusticity (**D**).

30 or more variables are added, which implies that the correlations towards the end are cumulative "true scores." Age first correlates significantly with muscle markers when 16 muscle markers are aggregated (r = 0.210, $P < 0.05$). Size first correlates significantly with muscle markers when 12 muscle markers are aggregated (r = 0.215, $P < 0.05$). Sex first correlates significantly with muscle markers when 14 muscle markers are aggregated (r = 0.207, $P < 0.05$). Robusticity first correlates significantly with muscle markers when 18 muscle markers are aggregated (r = 0.217, $P < 0.05$).

Since Aggregate Muscle Marker is correlated with age and sex, partial correlations were carried out to reexamine correlations after controlling for age, sex, and both. When age is controlled for, Aggregate Muscle Marker continues to correlate significantly with Humeral Size (r = 0.389, $P < 0.001$), sex (r = 0.397, $P < 0.001$), and Robusticity (r = 0.351, $P < 0.001$). When sex is controlled for, Aggregate Muscle Marker still correlates significantly with age (r = 0.369, $P < 0.001$) and Robusticity (r = 0.369, $P < 0.001$), but no longer with Humeral Size (r = 0.104, n.s.). When age and sex are controlled for, Aggregate Muscle Marker continues to correlate significantly with Robusticity (r = 0.340, $P < 0.05$), but not with Humeral Size (r = 0.130, n.s.).

To determine whether the sex differences were related to size rather than activity patterns, a partial correlation controlling for size was run. When controlling for Humeral Size, sex no longer correlated significantly with Aggregate Muscle Marker (r = 0.174, n.s.). Additionally, Humeral Size and sex correlated significantly (r = 0.800, $P < 0.001$).

A regression analysis was carried out to determine what the best predictor of Aggregate Muscle Marker is from the variables in this study, i.e., Humeral Size, Robusticity, age, and sex. Table 4 presents the results and shows that age is the overall best predictor. The predictions become higher when sex, Robusticity, and Humeral Size are entered into the equation, in that order. Table 5 presents the ANOVA analysis of these models, which shows the F-ratios and significance levels for the various models.

Since previous studies examined relations within specific populations, the correlations here were also run on each population separately. The results remained virtually the same. In the British Columbian sample, the correlations became stronger (i.e., Aggregate Muscle Marker correlated with age, r = 0.52; sex, r = 0.42; size, r = 0.44; and cross-sections, r = 0.53, $P < 0.001$).

DISCUSSION

This study found that whereas single measures showed null findings, using multiple measures showed significant findings. Aggregate Muscle Marker correlated with: age, r = 0.49; size, r = 0.38; sex, r = 0.40; and cross-sections, r = 0.38; $P < 0.001$. Previously reported null findings (e.g., Bridges, 1997; Stirland, 1998; Zumwalt et al., 2001) may be in part the result of a failure to use aggregate measures, which are more reliable because they reduce error variance and increase reliability.

While the use of aggregation is not common in anthropology, the *principle of aggregation* rests on familiar procedures. For example, whereas using only one question from a multiple choice test is obviously an inaccurate way to measure knowledge, the sum of 60 multiple-choice questions typically provides a reliable and accurate indicator. Aggregation is common in sports (e.g., Hoop Stats in basketball, Loss Leaders in baseball), the stock market (e.g., the Dow Jones Index), education (e.g., Dow Jones Safety School Index), and demography (e.g., *Money Magazine* ratings of the top 10 metropolitan areas). Although not based on numbers, composite portraits used in forensics are based on aggregation; the whole face is more than the pieces (e.g., chin,

TABLE 4. Regression analysis of Aggregate Muscle Marker, Humeral Size, Robusticity, age, and sex (N = 91)

Model	*R*	*R*^{2}	Adjusted R^2
1. Age	0.489	0.239	0.230
2. Age and Sex	0.599	0.358	0.344
3. Age, Sex, and Robusticity	0.658	0.432	0.413
4. Age, Robusticity, and Humeral Size	0.682	0.465	0.447
5. Age, Sex, Robusticity, and Humeral Size	0.683	0.466	0.441

TABLE 5. ANOVA of predictor models from regression analysis (N = 91)

Model		Sum of squares	df	F
1. Age	Regression	8,017.435	1	27.90*
	Residual	25,574.188	89	
	Total	33,591.623	90	
2. Age and Sex	Regression	12,039.241	2	24.58*
	Residual	212,552.382	88	
	Total	33,591.623	90	
3. Age, Sex, and Robusticity	Regression	14,526.013	3	22.10*
	Residual	19,065.610	87	
	Total	33,591.623	90	
4. Age, Robusticity, and Humeral Size	Regression	15,624.217	3	25.22*
	Residual	17,967.406	87	
	Total	33,591.623	90	
5. Age, Sex, Robusticity, and Humeral Size	Regression	15,658.658	4	18.77*
	Residual	17,932.965	86	
	Total	33,591.623	90	

* $P < 0.001$.

nose, or eyes). Another example is dating fossil sites, in which the greater confidence stems from the more procedures used (e.g., paleoflora and paleofauna, radiocarbon or potassium-argon dating methods, artifacts, and the principle of sedimentation). The date of the Tsimshin site used in this study was based on over 30 radiocarbon dates, artifacts, and the principle of sedimentation (Cybulski, 1992). Examples that use statistics are also abundant. For example, Rushton and Erdle (1987) showed how correlations between a questionnaire on aggression and the predictability of aggressive acts committed increased from nonsignificant (0.10) to significant (0.54) when the number of items in the questionnaire increased from 1 to 11. Rushton et al. (1983) showed that aggregating many judges' ratings also increased correlations from nonsignificant to significant between ratings and actual behavior. Other examples come from the medical literature; for instance, Fischer et al. (1999) formed a composite measure for success of clinical treatment for multiple sclerosis, whereas earlier noncomposite measures were inadequate to determine outcomes of clinical treatment. Lord and Fitzpatrick (2001) used a composite measure to identify elderly individuals at risk of falling; the composite measure correlated with a risk of falling at 0.45, but any one indicator was not significantly correlated with falling risk. Finally, aggregation often makes sense biologically, as well as statistically. For instance, because muscles work in groups, the use of a single muscle marker to reconstruct activity patterns can result in a weak finding (Stirland, 1998).

It may be important to note, however, that there are times when aggregation is not helpful. For example, when examining a specific phenomenon, one may not want to aggregate in case the phenomenon is masked (such as when looking at certain rare behaviors performed by primates). At other times, specific bones may be of interest, and aggregating over too many bones (e.g., arm and leg bones) may hide specific findings. Another shortfall of aggregating is that one must decide which variables to aggregate; unreliable or poor items should not be aggregated or they will, in fact, reduce validity. Often, in science, it is ultimately a matter of gaining the greatest predictability. In the end, researchers must decide whether aggregation is appropriate for their study. In this study, aggregation was used for several reasons: 1) interobserver and intraobserver rates were low for the method of data collection used, which suggests that each item of data is reliable (Hawkey and Merbs, 1995); 2) aggregation made sense biologically because muscles work in groups; 3) general patterns were being examined rather than a specific phenomenon; and 4) aggregation streamlined the data analyses.

In review, the Aggregate Muscle Marker was constructed from 42 component variables (a total of 7 muscle markers from 4 humeral sites, 2 radial sites; and 1 ulnar site, with each marker scored in the three categories of ruggedness, stress lesions, and ossification, with right and left arm bones added together). Composite Robusticity was constructed from total area, cortical area, moment of inertia about the anteroposterior axis, moment of inertia about the mediolateral axis, and polar moment of inertia (all of which were taken at 35% of bone length, side-averaged, standardized by body size, and log-transformed). Finally, aggregate Humeral

Size was constructed from humeral head size, humeral length, and humeral epicondylar breadth.

This study found that age was the single best predictor of Aggregate Muscle Marker. Older individuals had greater muscle marker scores than did younger individuals. This finding held for both sexes separately, although only marginally so in females. The correlation with age and muscle markers corroborates many other studies (e.g., Chapman, 1997; Kennedy, 1983, 1989; Nagy, 1998; Robb, 1998; Wilczak, 1998). Anthropologists have hypothesized that older individuals have more muscle markers than younger individuals because they have experienced more stress over a lifetime of activities. Age differences also could be related to changes in bone structure due to the slowing down of bone remodeling, resulting in a thinner cortical bone with a greater diameter and a rougher external bone (Dewey et al., 1968; Mays; 2000).

Results from this sample also showed that using the composite measure Humeral Size as a predictor variable, individuals with larger humeri had greater muscle marker scores than did individuals with smaller humeri, a pattern that remained when controlling for age, but not when controlling for sex. The body size and muscle marker correlation supports the observation of Zumwalt et al. (2000) that, among nonhuman primates, muscle markers correlate with body size regardless of locomotor style. However, because the correlation in the present study only held when the sexes were combined, the relation may be a weak one. It is important to note that sex and size were highly correlated ($r = 0.80$, $P < 0.001$). It could also be that the present result was weak because it is based on human upper limbs, which are freed from weight-bearing. A stronger effect might be found when examining the lower limb or lower back, which do bear weight.

Males had greater muscle markers than did females (Fig. 7), a finding that is well-established in the literature (although some studies found females with greater muscle markers than males, e.g., Chapman, 1997; Nagy and Hawkey, 1995). Sex differences in muscle markers are often interpreted as due to sex differences in activity patterns (e.g., Chapman, 1997; Cook and Dougherty, 2001; Steen and Lane, 1998; Wilczak, 1998). Among the Tsimshin of British Columbia, the males rowed upon the ocean, whereas the females were more sedentary. One may conjecture that in this particular sample, males had greater muscle markers than females because of their age (12% of males and 0% of females were over 45 years old). Yet when age was controlled, males still had greater muscle markers than females. Most likely, males have greater muscle markers because of their size, males being, on average, larger, heavier, and with more muscle mass than females; in this study, Humeral Size and Sex are highly correlated. When a partial correlation was carried out controlling for Humeral Size, males and females no longer differed in muscle markers. In other words, in this study, the sex difference in muscle markers seems to be the result of sex differences in body size.

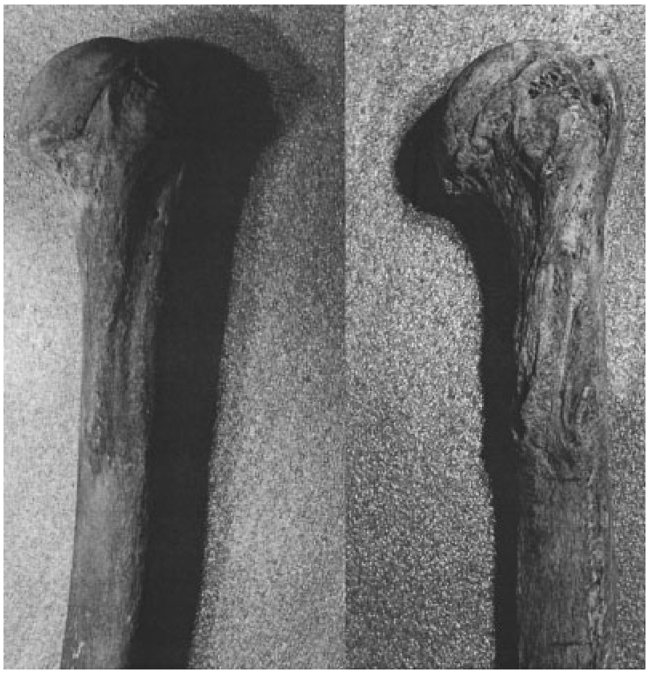

Fig. 7. **Left:** Female British Columbian Amerind humerus (catalog no. CP-2001 R006-17) with minor muscle markings. **Right:** Male British Columbian Amerind humerus (catalog no. CP-2001 R006-15) with major muscle markings. Photographs by Jerry Cybulski, courtesy of the Canadian Museum of Civilization.

Corroborating the finding of Berget and Churchill (1994) that cross-sectional properties correlated with muscle markers, the present study found that individuals with more robust humeri had greater muscle marker scores than did individuals with less robust humeri, a pattern that held for both males and females, and young and old. The robusticity and muscle marker correlation validates the use of muscle markers as indicators of activity levels. It is possible that some previous null findings of Bridges (1997), Stirland (1998), and Zumwalt et al. (2000, 2001) were due in part to their failure to use nonaggregate measures.

CONCLUSIONS

In conclusion, muscle markers in this sample from British Columbia and Quebec correlated with age, size, sex, and robusticity. Older individuals had greater muscle markers than did younger individuals; individuals with larger humeri had greater muscle markers than did individuals with smaller humeri; males had greater muscle markers than did females; and individuals with more robust humeri had greater muscle markers than did individuals with less robust humeri. Yet determining causality in a study such as this is difficult because of the colinearity among the variables and the small sample size. This study does confirm previous research suggesting that age should be taken into consideration when examining muscle markers. Finally, the effects of sex, body size, and population should also be controlled for when examining hypotheses about activity patterns on muscle markers in order to reconstruct past lifestyles. This study also suggests that aggregate measures may be useful (when appropriate and possible) to reduce error variance and enhance construct validity.

LITERATURE CITED

Berget KA, Churchill SE. 1994. Subsistence activity and humeral hypertrophy among Western Aleutian Islanders. Am J Phys Anthropol 18:55 [abstract].

Biknevicius AR, Ruff CB. 1992. Use of biplanar radiographs for estimating cross-sectional geometric properties of mandibles. Anat Rec 232:157–163.

Bridges PS. 1997. The relationship between muscle markings and diaphyseal strength in prehistoric remains from West-Central Illinois. Am J Phys Anthropol 24:82 [abstract].

Buikstra JE, Ubelaker DH. 1994. Standards for data collection from human skeletal remains. Fayetteville: Archaeology Survey.

Capasso L, Kennedy KAR, Wilczak CA. 1999. Atlas of occupations markers of human remains. Journal of Paleopathology, monograph publication 3. Teramo: Edigrafital SpA.

Chamay A, Tschantz P. 1972. Mechanical influences in bone remodeling. Experimental research on Wolff's law. J Biomech 5:173–180.

Chapman NE. 1997. Evidence for Spanish influence on activity induced musculoskeletal stress markers at Pecos Pueblo. Int J Osteoarchaeol 7:497–506.

Cohen MN. 1989. Health and the rise of civilization. New Haven: Yale University Press.

Cook DC, Dougherty SP. 2001. Row, row, row your boat: activity patterns and skeletal robusticity in a series from Chirikof Island, Alaska. Am J Phys Anthropol 32:53 [abstract].

Cybulski JS. 1990. Human biology. In: Suttles W, editor. Handbook of North American Indians, volume 7, northwest coast. Washington, DC: Smithsonian Institution. p 107–144.

Cybulski JS. 1992. A Greenville burial ground: human remains and mortuary elements in British Columbia coast prehistory. Archaeological Survey of Canada Mercury series paper 146. Hull: Canadian Museum of Civilization.

Cybulski JC. 1999. Trauma and warfare at Prince Rupert Harbour. Midden 31:5–7.

Dewey JR, Armelagos GJ, Bartley MH. 1968. Femoral cortical involution in Nubian archaeological populations. Hum Biol 41: 13–28.

Dunn G, Sham PC, Hand DJ. 1993. Statistics and the nature of depression. J R Stat Soc A 156:63–87.

Fischer JS, Rudick RA, Cutter GR, Reingold SC. 1999. The multiple sclerosis functional composite measure (MSFC): an integrated approach to MS clinical outcome assessment Mult Scler 5:244–250.

Fresia A, Ruff CB, Larsen CS. 1990. Temporal decline in bilateral asymmetry of the upper limb on the Georgia coast. In: Larsen CS, editor. The archaeology of Mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition, no. 68. New York: Anthropological Papers of the American Museum of Natural History. p 121–132.

Gray H. 1977. Anatomy, descriptive and surgical. New York: Bounty Books.

Hawkey DE. 1998. Disability, compassion and the skeletal record: using musculoskeletal stress markers (MSM) to construct an osteobiography from early New Mexico. Int J Osteoarchaeol 8:326–340.

Hawkey DE, Merbs CF. 1995. Activity-induces musculoskeletal stress markers (MSM) and subsistence strategy changes among ancient Hudson Bay Eskimos. Int J Osteoarchaeol 5:324–338.

Hawkey DE, Street S. 1992. Activity-induced stress markers in prehistoric human remains from the eastern Aleutian Islands. Am J Phys Anthropol 14:89 [abstract].

Jurmain R. 1990. Paleoepidemiology of a central California prehistoric population from CA-ALA-329: II. Degenerative disease. Am J Phys Anthropol 83:83–94.

Jurmain R. 1999. Stories from the skeleton: behavioral reconstruction in human osteology. London: Taylor and Francis, Ltd.

Kennedy KAR. 1983. Morphological variations in ulnar supinator crests and fossae as identifying markers of occupational stress. J Forensic Sci 28:871–876.

Kennedy KAR. 1989. Skeletal markers of occupations stress. In: Iscan M, Kennedy K, editors. Reconstruction of life from the skeleton. New York: Alan R. Liss. p 129–160.

Khamis S, Hempstead H. 1996. Consistency in aggregation principle for multilateral comparisons of purchasing power parities and real products. In: Prasada-Rao D, Salazar-Carrillo J, editors. International comparisons of prices, output and productivity. Contributions to economic analysis, volume 231. Amsterdam: Elsevier. p 187–193.

Lai P, Lovell NC. 1992. Skeletal markers of occupational stress in the fur trade: a case study from a Hudson's Bay Company fur trade post. Int J Osteoarchaeol 2:221–234.

Lord SR, Fitzpatrick RC. 2001. Choice stepping reaction time: a composite measure of falls risk in older people. J Gerontol [A] 56:627–632.

Lovejoy CO. 1971. Methods for the detection of census error in palaeodemography. Am Anthropol 73:101–109.

Mays S. 2000. Age-dependent cortical bone loss in women from 18th and early 19th century London. Am J Phys Anthropol 112:349–361.

McCall R. 1990. Fundamental statistics for behavioral sciences, 5th ed. New York: Brace and Jovanovich Publishers.

Merbs CF. 1983. Patterns of activity-induced pathology in a Canadian Inuit population. Archaeol Surv Can Pap 119:1–199.

Nagy BLB. 1998. Age, activity, and musculoskeletal stress markers. Am J Phys Anthropol 26:168–169 [abstract].

Nagy BLB. 1999. Bioarchaeological evidence for atl-atl use in prehistory. Am J Phys Anthropol 28:208 [abstract].

Nagy BLB, Hawkey DE. 1995. Musculoskeletal stress markers as indicators of sexual division of labor: multivariate analyses. Am J Phys Anthropol 20:158 [abstract].

Peterson J. 1998. The Natufian hunting conundrum: spears, atlatls, or bows? Musculoskeletal and armature evidence. Int J Osteoarchaeol 8:378–389.

Piedalue G, Cybulski JS. 1997. Buried prisoners in eighteenth-century Quebec. In: Poirier D, Bellatoni N, editors. In remembrance: archaeology and death. Westport: Bergin and Garvey. p 121–129.

Robb JE. 1998. The interpretation of skeletal muscle sites: a statistical approach. Int J Osteoarchaeol 8:363–377.

Ruff CB, Larsen CS. 1990. Postcranial biomechanical adaptations to subsistence changes on the Georgia coast. In: Larsen CS, editor. The archaeology of Mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition. Anthropological Papers of the American Museum of Natural History 68. New York: American Museum of Natural History. p 94–120.

Ruff CB, Scott W, Liu A. 1991. Articular and diaphyseal remodeling of the proximal femur with changes in body mass in adults. Am J Phys Anthropol 86:397–413.

Runestad JA, Ruff CB, Nieh JC, Thorington RW Jr, Teaford MF. 1993. Radiographic estimation of long bone cross-sectional geometric properties. Am J Phys Anthropol 90:207–213.

Rushton JP, Erdle S. 1987. Evidence for an aggressive (and delinquent) personality. Br J Soc Psychol 26:87–89.

Rushton JP, Brainerd CJ, Pressley M. 1983. Behavioral development and construct validity: the principle of aggregation. Psychol Bull 94:18–38.

Spearman C. 1904. "General intelligence": objectively determined and measured. Am J Psychol 15:201–292.

Spearman C. 1910. Correlation calculated from faulty data. Br J Psychol 3:271–295.

Spearman C. 1927. The abilities of man. New York: Macmillan.

Steen SL, Lane RW. 1998. Evaluation of habitual activities among two Alaskan Eskimo

populations based on musculoskeletal stress markers. Int J Osteoarchaeol 8:341–353.

Stirland AJ. 1998. Musculoskeletal evidence for activity: problems of evaluation. Int J Osteoarchaeol 8:354–362.

Timoshenko S, Gere J. 1972. Mechanics of materials. Monterey: Brooks/Cole Engineering Division.

Van Rompaey A, Govers G, Baudet M. 1999. A strategy for controlling error of distributed environmental models by aggregation. Int J Geogr Info Sci 13:577–590.

Weiss E. 2001. A cross-cultural study of humeri: environmental causes of morphology. Ph.D. dissertation, University of Arkansas, Fayetteville.

Weiss N, Hassett M. 1982. Introductory statistics. Reading: Addison-Wesley Publishing Co.

Wilczak CA. 1998. Consideration of sexual dimorphism, age, and asymmetry in quantitative measurements of muscle insertion sites. Int J Osteoarchaeol 8:311–325.

Wolff J. 1892. Das Gesetz der Transformation der Knochen. Berlin: A. Hirschwald.

Woo SL, Kuei SC, Amiel D, Gomez MA, Hayes WC, White FC, Akeson WH. 1981. The effect of prolonged physical training on the properties of long bone: a study of Wolff's law. J Bone Joint Surg [Am] 63:780–786.

Zumwalt AC, Ruff CB, Wilczak CA. 2000. Primate muscle insertions: what does size tell you? Am J Phys Anthropol 30:331 [abstract].

Zumwalt AC, Ruff CB, Lieberman DE. 2001. The influence of exercise on muscle insertion scars in sheep. Am J Phys Anthropol 32:170 [abstract].