AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 70:377-393 (1986)

Effects of Altitude, Ethnicity-Religion, Geographical Distance, and Occupation on Adult Anthropometric Characters of Eastern Himalayan Populations

PARTHA P. MAJUMDER, RANJAN GUPTA, BARUN MUKHOPADHYAY, PREMANANDA BHARATI, SUBRATA K. ROY, M. MASALI, A.W. SLOAN, AND AMITABHA BASU
Indian Statistical Institute, Calcutta 700035, India (P.P.M., R.G., B.M., P.B., S.K.R., A.B.), University of Turin, 10123 Turin, Italy (M.M.), and Department of Physiology, Medical School, Observatory 7925, Cape Town, South Africa (A.W.S.)

KEY WORDS Sociocultural factors, Multivariate analysis

ABSTRACT With a view to estimating the effect of altitude on body dimensions vis-à-vis ethnicity-religion, geographical distance, and occupation, a comprehensive multivariate statistical analysis was performed on data pertaining to 16 anthropometric characters collected from 1,103 individuals (643 males and 460 females) belonging to two ethnic groups—Sherpa and Lepcha. Samples were drawn from several locations in the eastern Himalayan region—Darjeeling and Kalimpong in West Bengal (India), and Nepal, situated at low (1,000–2,000 meters) and high (above 3,500 meters) altitudes. The individuals sampled practice different occupations and follow different religions. Significant age and sex effects were observed. The data were age-adjusted, and sexes were treated separately. A test of equality of mean vectors indicated heterogeneity among population groups. Almost all characters were found to contribute significantly to the ability to discriminate between the groups. The overall probability of correctly classifying an individual based on body dimensions into the group in which she or he actually belongs was high (between 0.64 and 0.77). Shape and size factors could be identified that explained about 50% of the total variance and yielded a reasonable separation of the groups. Results of four different types of multivariate statistical analyses were in agreement, and showed that altitude is most highly associated with body dimensions.

It is a common observation that striking differences exist between high and low altitude human populations with respect to various biomedical traits. Although no universally accepted cutoff point exists, in human biological studies, populations permanently living at altitudes above 3,000 meters are generally designated as "high-altitude populations" (Baker, 1978). Profound, though not always consistent, physiological, demographic, and anthropometric differences have been observed between high- and low-altitude populations (Baker and Little, 1976; Baker, 1978). These differences have sometimes been attributed to altitude related environmental stresses such as hypoxia, cold, etc. (Hurtado, 1964) or to sociocultural ones (Clegg et al., 1970; Weitz, 1984). There have been very few studies (Rothhammer and Spielman, 1972; Haas, 1976; Mueller et al., 1978) of the effects of microcultural factors on body dimensions at high altitudes, and even these studies did not reveal consistent patterns. Furthermore, some workers in this field (e.g., Cruz-Coke, 1968) believe that the genetic backgrounds of populations are not significant factors in their biological adaptation to high altitude.

A multidisciplinary study of the nature and extent of adaptation of human populations to various environmental stresses was initiated under the general title of Human Adaptabil-

---

Received September 17, 1984; revision accepted February 3, 1986.

ity Program in the Indian Statistical Institute in early 1976. The short-term objective of the program was to identify and measure the impacts of three sets of factors, e.g., physical environment, socioculture, and ethnicity, on health, as well as of human activity on environment, following recommendations of the IBP/HA panel (Weiner, 1969) and UNESCO/MAB program (UNESCO, 1973); the long-term objective was to determine the limits to human adaptation.

Since stresses, particularly those related to altitude, seem to have important, although sometimes inconsistent, effects on human biology and health and since there are no comprehensive data from the eastern Himalayan populations, studies were undertaken in this region. These studies were designed to circumvent some of the methodological weaknesses that were present in many earlier studies of similar nature. For instance, in many earlier studies, genetically different populations living in different altitudes were compared. Consequently, the possible effects of altitude could not be isolated from those owing to genetic differences between such populations. This problem of confounding of causal factors of variation was avoided in the present study in the following manner:

1. by comparing subgroups of the same ethnic group (a) living at different altitudes, (b) living at similar altitudes but practicing different occupations, (c) living at similar altitudes and practicing similar occupations but located geographically apart;
2. by comparing different ethnic groups living at similar altitudes and practicing similar occupations;
3. by comparing the same ethnic group living at similar altitudes, practicing similar occupations, located geographically close but having different microcultural characteristics, e.g., religion.

Univariate statistical analyses of the data collected by us thus far have revealed the following major differences between low- and high-altitude populations (Gupta, 1981):

1. There is a significant reduction in fertility among high-altitude populations compared to low-altitude populations; no such difference exists in respect of mortality.
2. The adult body dimensions are greater in high-altitude populations, which is in discordance with reports from other parts of the world (Frisancho, 1976, 1978) except Ethiopia (Clegg et al., 1970).
3. In concordance with other studies, the rate of physical growth among children at high altitudes is slower and more protracted than among children at low altitudes, but the magnitudes of differences are smaller than those observed in other populations. The "growth-spurt" is also absent among high-altitude children.
4. As expected for populations living under hypoxic stress, values of hematological parameters (e.g., packed cell volume, hemoglobin concentration, total red-cell count) are higher among high-altitude populations than among low-altitude ones.

Much of the data pertaining to adult body dimensions considered in previous papers have been analyzed by means of univariate statistical methods (Gupta, 1981; Gupta and Basu, 1981). We have, so far, not conducted any analysis considering all the variables jointly by means of multivariate statistical techniques. The purpose of the present paper is, therefore, to study the pattern of anthropometric variation as a whole (i.e., considering all the anthropometric variables together) by performing a comprehensive multivariate statistical analysis of the data, designed specifically to identify differences in body dimensions that are due to altitude in the context of differing occupational and ethnic-religious backgrounds as well as geographical distance.

MATERIALS AND METHODS

The geographical area covered in the present study is the eastern Himalayan mountainous terrain, specifically northeastern Nepal and northern West Bengal in India. The entire study area lies between 27°45′ N and 26°51′ N latitudes and between 86°40′ E and 88°53′ E longitudes. Figure 1 shows the regional location and the study area in Nepal and India. As mentioned in the opening section, the population groups and locations were chosen to enable us to study the effects of the major possible causal factors leading to biological variation in isolation, that is, to ensure by the sampling design that the effect of the causal factor of interest could be estimated in isolation from the effects of other causal factors. The causal factors considered here are altitude, ethnicity-religion, occupation, and geographical distance.

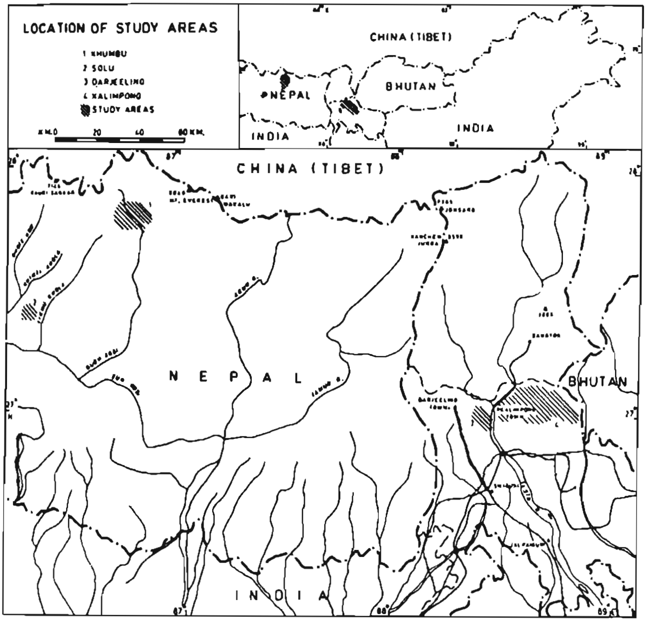

Fig. 1. Map locations showing study areas.

Two ethnic groups have been studied: Sherpas and Lepchas. At present, the Sherpas are distributed mainly in the Khumbu, Pharak, and Solu regions in Nepal, Western Sikkim, and the three hill subdivisions, namely, Darjeeling Sadar, Kurseong, and Kalimpong, in West Bengal, India. The Sherpas of Khumbu are an offshoot of the larger Tibetan population that migrated from eastern Tibet around 1533 A.D. (Haimendorf, 1964; Oppitz, 1974). Subsequently, through different waves of migration, the Khumbu Sherpas moved to the other locations that they presently occupy. The Sherpas of Khumbu live at altitudes between 3,500 and 4,050 meters, while the Sherpas of Solu (in Nepal) and Darjeeling Sadar and Kalimpong (both in West Bengal, India) live at altitudes between 1,000 and 2,000 meters. We have studied the Sherpas at all four of these locations. (Data on Sherpas of Solu were gathered by A.W.S.; the rest were gathered by the Indian Statistical Institute team.) The Sherpas hold uniform religious beliefs; and while in Khumbu and Solu the Sherpas generally practice agriculture, in Darjeeling some practice agriculture, while others are plantation laborers. In Kalimpong they are divided into three occupational clusters—agriculturists, plantation laborers, and those engaged in forest-based activities such as wood-cutting and timber-selling (Gupta, 1981).

The Lepchas are ethnically distinct from the Sherpas, and are widely believed to be the autochthones of the Sikkim-Darjeeling area of India. The religious beliefs of the Lepchas may be classified into three categories: Animism, Buddhism, and Christianity (Das, 1978). We have studied only the Buddhist and the Christian Lepchas living in the vicinity of Kalimpong, at low altitudes. Both the Lepcha subgroups studied are agriculturists. Religion was taken to be a broad indicator of socioeconomic status; for instance, Christian Lepchas were generally economically better off and had higher educational status.

Despite certain differences in specific dietary patterns, the Sherpas and Lepchas studied thrive on a cereal-dominated diet (Gupta, 1981). Similarly, although there are some minor differences in disease patterns, the common ailments in both the populations are influenza, gastrointestinal and upper respiratory tract diseases, tuberculosis, and leprosy (Gupta, 1981).

The choice of population groups and study locations, as mentioned above, obviously permits study of the effects of the various possible causal factors in isolation. We shall, however, not investigate the effects of the possible causal factors through pair-wise comparisons, but shall, by the use of multivariate statistical techniques, study the relative contributions of the causal factors in determining adult body dimensions.

Each individual included in this study was an adult (defined as being aged 20 years or older at the time of survey). Because of obvious difficulties in the field, the individuals studied could not be chosen by any rigorous sampling procedure, so that the sample comprises those adults who were cooperative. However, the adults included in the study comprised about 80% of the adults resident in each of the selected villages. Some difficulties were encountered in the assessment of age as there is no mandatory registration (or other written documentation) of births. Age was, therefore, estimated with the help of the traditional calendars (Tibetan animal element calendars), or by reference to important local events, or to the ages of individuals for whom reliable records existed, and cross-checked from a number of elderly individuals and on subsequent visits. The Tibetan animal element calendar, which is followed by all the populations under consideration, follows a 12-year cycle with 1-year periods (each named after a specific animal, natural phenomenon, etc.) within the cycles, so that the error in estimating an individual's true age is less than 1 year.

Measurements pertaining to 16 anthropometric characters were taken on each selected individual. The characters chosen are those included in the IBP basic list of recommended measurements and were measured following standard techniques (Weiner and Lourie, 1969). The names and codes of the characters chosen are listed in Table 1.

Brief descriptions, codes, and sample sizes of each of the population groups have been presented in Table 2. In all, 1,103 individuals comprising 643 males and 460 females from nine population groups were measured.

RESULTS

Sex and age effects

The basic descriptive statistics (minimum, maximum, mean, and standard deviation values) of each character were computed, separately for males and females, for each of the nine population groups. These statistics are not presented herein, but tables presenting these statistics may be obtained by writing to the authors.

The mean values for most anthropometric characters, as expected, were significantly different for the two sexes—females generally had lower mean values except for skinfold thicknesses. This indicates that the data for the two sexes within each population need

TABLE 1. List of codes for anthropometric variables

Serial No.	Variable name	Variable code
1	Height	HT
2	Sitting height	SH
3	Biacromial diameter	BAD
4	Biiliac diameter	BID
5	Weight	WT
6	Biceps girth	BG
7	Calf girth	CG
8	Head length	HL
9	Head breadth	HB
10	Bizygomatic breadth	BB
11	Bicondylar femur diameter	BFD
12	Morphological facial height	MFH
13	Nose height	NH
14	Nose breadth	NB
15	Triceps skinfold thickness (left)	TST
16	Subscapular skinfold thickness (left)	SST

TABLE 2 Descriptions, codes, and sample sizes of population groups

Serial No.	Description				Population code	Sample size		
	Population		Ethnic group	Occupation		Total	Male	Female
	Altitude[1]	Location						
1	Low	Kalimpong	Sherpa	Agriculture	LKASHA	72	26	46
2	Low	Kalimpong	Sherpa	Plantation	LKASHP	164	100	64
3	Low	Kalimpong	Sherpa	Forestry	LKASHF	153	94	59
4	Low	Kalimpong	Lepcha (Buddhist)	Agriculture	LKALBA	136	64	72
5	Low	Kalimpong	Lepcha (Christian)	Agriculture	LKALCA	257	133	124
6	Low	Nepal (Solu)	Sherpa	Agriculture	LNSSHA	76	76	0
7	Low	Darjeeling	Sherpa	Agriculture	LDASHA	81	48	33
8	Low	Darjeeling	Sherpa	Plantation	LDASHP	59	35	24
9	High	Nepal (Khumbu)	Sherpa	Agriculture	HNKSHA	105	67	38

[1] Low, 1,000-2,000m; High, $\geq 3{,}500$m.

to be considered separately and cannot be pooled.

Although all sampled individuals are adults (aged 20 years or older), it was considered necessary to check whether there are any significant age effects. For this, a linear regression line, of the form $X = \alpha + \beta A$ (where X denotes the value of a character for an individual of age A), was fitted for each character within each population separately for each sex. Tests of significance were performed to check whether $\beta = 0$. For many characters, especially CG, BAD, and NH (see Table 1), the null-hypothesis $\beta = 0$ was rejected at the 5% level in many populations, implying significant age effects. However, a careful examination of the estimated values of β showed that in almost all cases the numerical magnitudes of β are very small (<0.1), implying a small age effect. Furthermore, most β values are negative indicating a declining, albeit slight, trend of anthropometric measurements with advancing age. In order to determine whether a single regression line could be used to adjust for the effect of age on the values of a character, we performed tests to determine whether the regression coefficients are equal in all the population groups. The values of the F-statistic for testing this hypothesis were all significant at the 5% level, indicating that the effect of age on the characters are significantly different in the various population groups. The linear effect of age was, therefore, removed from the data by using the population group-specific and sex-specific regression line. The remaining analysis pertains to residual values of the characters after adjusting for the linear age effect, i.e., in the remaining analysis the value used of the i-th character for the j-th individual belonging to the k-th population group (treating the males and females separately) is $\tilde{X}_{ijk} = X_{ijk} - (\alpha_{ik} + \beta_{ik} A_{ijk})$, where X_{ijk} is the actual measured value, and A_{ijk} is the age of the individual under consideration. As expected, even after adjusting for the linear age effect, mean values for the two sexes were significantly different for many characters in most populations. (Note: Nonlinear regressions on age were not performed, because visual examination of the scatter diagrams did not indicate any nonlinear effect.)

The sexes within each population group, have, therefore, been treated separately, and will be denoted, for example, as LKASHA-M and LKASHA-F for males and females, respectively, of population group LKASHA. Of the nine population groups, there were no female samples from the population LNSSHA. Some results that we present pertain to differences among populations within the males (nine population groups) and within the females (eight population groups), while some pertain to differences between the sexes by considering all the 17 groups together.

Equality of mean vectors

To find out whether the population groups are distinguishable on the basis of anthropometric measurements, we performed a test of the null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_{17}$, where μ_i denotes the mean vector (of order 16 × 1) of the i-th population group (i 1 refers to LKASHA-M, i = 2 refers to LKASHA-F, etc.). The value of the Wilks' Λ-statistic for testing H_0 was 0.0045, and the value of the approximate F-statistic for testing the significance of this Λ-value (Rao, 1973, p. 472) was 28.995, which is significant at the 5% level with 256 and 12214.1 degrees of freedom. The null hypothesis was, therefore, rejected, which implies that the populations are distinguishable on the basis of the 16 anthropometric measurements considered. Within each sex, tests of the null hypotheses of equality of mean vectors (nine vectors for males and eight vectors for females) yielded the following results: (1) for males, the Wilks' Λ-value was 0.0204, and the corresponding approximate F-statistic value was 25.041 with 128 and 4,476.39 degrees of freedom; and (2) for females, the Wilks' Λ-value was 0.0289, and the F-statistic value was 18.481 with 112 and 2,834.98 degrees of freedom. Both the F-values are significant at the 5% level, showing thereby that even within each sex the populations are distinguishable.

Relationships among the population groups

Since the populations are distinguishable on the basis of the anthropometric characters, it was of interest to know the relationships among them. For this purpose, we considered all of the 17 groups together, without separating the populations within each sex, in order to examine relationships among population groups not only within sex, but also between the sexes. We computed the Mahalanobis' D^2 values between all pairs of populations, and tested the equality of all pairs of mean vectors by using a transformation of the D^2-statistic, which follows an F-distribution (Rao, 1973, p. 480). These values have been presented in Table 3. It is seen from this table that with the exception of the females of LKALCA and LKALBA, the hypothesis of equality of mean vectors was rejected for all pairs of population groups. The relationships among the population groups have been presented in the form of a dendrogram in Figure 2, which was constructed by performing a single linkage cluster analysis on the matrix of D^2-values. An examination of this figure reveals that the high-altitude groups are clearly different from the low-altitude groups. Also, irrespective of the ethnic-religious, location, and occupational factors, the two sexes are markedly different. The sociocultural and other factors, such as ethnicity-religion, location, and occupation, seem to play a small role in the differentiation of anthropometric characters because, for example, groups practicing the same occupation do not cluster together.

The principal components: Relationships among population groups and size and shape factors

Another method of studying relationships among the population groups with a view to identifying variables important in explain-

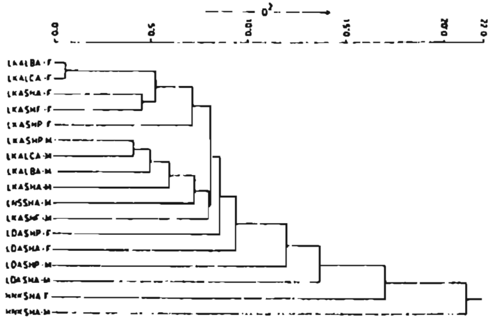

Fig. 2. Dendrogram depicting relationships among the population groups.

ing differences in body dimensions is principal component analysis. In this analysis, one examines whether a few linear combinations of the variables of the form

$$L_i = \sum_{j=1}^{16} l_{ij} V_j,$$

where V_j's are the anthropometric variables, and l_{ij}'s are the associated coefficients

$$(\sum_{j=1}^{16} l_{ij}^2 = 1 \; ; \; i = 1, 2, \ldots, 16)$$

can explain a large proportion of the total variance. If a few such linear combinations of the original variables can indeed be obtained, then one can use a smaller set of linear combinations of some of the original variables for purposes of examining relationships among the population groups. Moreover, by examining the magnitudes of the coefficients attached to the variables, these linear combinations can be described as size or shape factors, identifiable with morphological size and shape. For example, if for a particular principal component, L_i, l_{ij}'s are near zero for all variables except WT, TST, and SST, then this principal component may be called a size factor, since the component variables (weight and skinfold thicknesses) determine the body volume. We used the technique of principal component analysis to examine whether it is possible to obtain a reduction in dimensionality and identify size and shape factors.

The results of this analysis showed that three principal components explain about 60% of the total variance, and each of the remaining thirteen principal components explain only about 3–5% of the total variance. The coefficients attached to the variables in the first three principal components and the percentages of the total variance explained by each of these have been presented in Table 4. An examination of the magnitudes of the first two principal components reveal interesting features. In the first principal component, except for skinfold thickness variables the coefficients attached to all the variables are positive in sign and large in magnitude (mostly greater than 0.5). In the second principal component, the coefficients attached to the skinfold thickness variables are positive in sign and also large in magnitude compared to the coefficients attached to the remaining variables, a majority of which are small in magnitude, and many of which have a negative sign. We may, therefore, identify the first and second principal components with morphological shape and size, respectively. In order to discover how well the population groups can be separated in two dimensions, we plotted the values of the first two principal components computed at the mean values of the variables for each population group; this has been presented in Figure 3. This figure reveals relationships concordant with those depicted in Figure 2, and shows that the relationships of altitude and sex with body dimensions are strong, and the relationships of the other factors—ethnicity-religion, occupation, and location/geographical distance—are weaker.

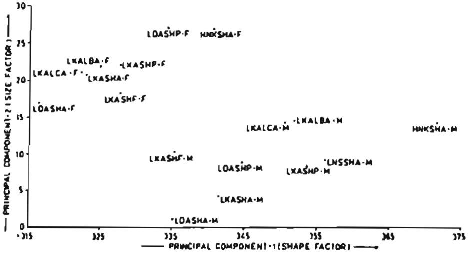

Fig. 3. Bivariate plot of the population groups on the basis of the first two principal components.

TABLE 3. Values of the F-statistic for testing equality of mean vectors between all pairs of population groups (below the diagonal) and the Mahalanobis-D^2 values between all pairs of population groups (above the diagonal)*

	LKASHA-M	LKASHA-F	LKASHP-M	LKASHP-F	LKASHF-M	LKASHF-F	LKALBA-M	LKALBA-F
LKASHA-M	—	19.03	6.09	22.64	8.00	16.27	10.03	19.23
LKASHA-F	19.53 (55)	—	21.99	10.20	13.48	4.68	22.34	6.20
LKASHP-M	7.82 (109)	42.64 (129)	—	19.99	9.62	16.58	8.55	20.57
LKASHP-F	25.92 (73)	16.84 (93)	48.17 (147)	—	16.85	11.42	21.20	8.45
LKASHF-M	10.09 (103)	25.60 (123)	28.77 (177)	39.85 (141)	—	8.03	13.11	13.99
LKASHF-F	18.16 (68)	7.27 (88)	37.96 (142)	21.83 (106)	17.92 (136)	—	21.67	6.48
LKALBA-M	11.48 (73)	36.80 (93)	20.57 (147)	41.77 (111)	30.86 (141)	41.01 (106)	—	21.20
LKALBA-F	22.62 (81)	10.52 (101)	52.96 (155)	17.60 (119)	35.03 (149)	12.85 (114)	44.18 (119)	—
LKALCA-M	13.16 (142)	44.49 (162)	14.90 (216)	50.76 (180)	29.95 (210)	40.17 (175)	13.53 (180)	49.75 (188)
LKALCA-F	23.45 (133)	10.74 (153)	68.63 (207)	19.00 (171)	46.49 (201)	15.83 (166)	54.56 (171)	1.53** (179)
LNSSHA-M	13.10 (85)	47.01 (105)	19.59 (159)	59.72 (123)	28.92 (153)	48.82 (118)	15.79 (123)	61.04 (131)
LDASHA-M	18.92 (57)	37.62 (77)	28.99 (131)	59.52 (95)	26.55 (125)	29.06 (90)	38.19 (95)	49.75 (103)
LDASHA-F	23.89 (42)	15.76 (62)	42.92 (116)	29.54 (80)	31.46 (110)	12.25 (75)	46.30 (80)	22.53 (88)
LDASHP-M	15.10 (44)	33.87 (64)	25.11 (118)	47.18 (82)	18.68 (112)	22.64 (77)	34.39 (82)	51.47 (90)
LDASHP-F	19.58 (33)	14.65 (53)	28.58 (107)	12.76 (71)	23.02 (101)	9.16 (66)	36.64 (71)	12.25 (79)
HNKSHA-M	39.12 (76)	74.92 (96)	69.84 (150)	103.48 (114)	60.19 (144)	68.12 (109)	84.68 (114)	116.04 (122)
HNKSHA-F	35.71 (47)	31.24 (67)	68.69 (121)	49.53 (85)	50.13 (115)	27.80 (80)	74.55 (85)	50.60 (93)

*The degrees of freedom of each F-value are 16 and the number is given in parentheses under the corresponding F-value.
**Not significant at the 5% level.

TABLE 3. Values of the F-statistic for testing equality of mean vectors between all pairs of population groups (below the diagonal) and the Mahalanobis-D^2 values between all pairs of population groups (above the diagonal) (continued)*

	LKALCA-M	LKALCA-F	LNSSHA-M	LDASHA-M	LDASHA-F	LDASHP-M	LDASHP-F	HNKSHA-M	HNKSHA-F
LKASHA-M	9.75	17.74	10.90	18.21	26.64	16.37	25.43	33.81	37.38
LKASHA-F	21.16	5.27	26.68	26.08	13.32	27.60	15.12	44.57	24.20
LKASHP-M	4.22	20.14	7.38	14.54	28.02	15.71	23.92	28.28	40.38
LKASHP-F	19.00	7.30	27.86	35.16	21.93	33.70	11.77	51.13	33.42
LKASHF-M	8.81	14.16	11.13	13.70	20.92	11.97	19.55	24.99	30.06
LKASHF-F	15.95	6.45	23.85	17.88	9.36	16.73	8.70	35.18	19.42
LKALBA-M	5.06	21.01	7.33	22.64	34.48	24.67	34.05	41.97	50.55
LKALBA-F	17.31	0.54	26.84	28.03	16.21	35.60	11.05	54.33	33.04
LKALCA-M	—	17.04	9.99	15.92	29.25	18.32	24.87	38.50	46.02
LKALCA-F	67.31 (240)	—	27.47	28.38	14.76	33.55	9.70	55.12	31.02
LNSSHA-M	29.79 (192)	79.66 (183)	—	19.53	38.55	23.08	36.08	21.87	43.78
LDASHA-M	34.53 (164)	60.46 (155)	35.26 (107)	—	21.84	13.94	29.82	35.68	42.23
LDASHA-F	47.72 (149)	23.58 (140)	54.70 (92)	26.34 (64)	—	21.06	12.75	45.62	17.01
LDASHP-M	31.22 (151)	56.19 (142)	34.00 (94)	17.16 (66)	22.09 (51)	—	25.16	22.79	28.95
LDASHP-F	31.19 (140)	11.97 (131)	40.60 (83)	29.42 (55)	10.90 (40)	22.04 (42)	—	44.01	18.21
HNKSHA-M	105.79 (183)	147.57 (174)	47.90 (126)	61.31 (98)	62.31 (83)	32.19 (85)	48.92 (74)	—	27.11
HNKSHA-F	84.03 (154)	55.58 (145)	68.42 (97)	55.16 (69)	18.56 (54)	32.59 (56)	16.51 (45)	40.63 (88)	—

*The degrees of freedom of each F-value are 16 and the number given in parentheses under the corresponding F-value.

**Not significant at the 5% level.

TABLE 4. The first three principal components

Variable	Estimates of coefficients for principal component 1	2	3
HT	0.827	−0.207	−0.043
SH	0.821	−0.194	0.061
BAD	0.748	−0.094	−0.241
BID	0.335	0.419	−0.331
WT	0.753	0.364	−0.293
BG	0.565	0.472	0.045
CG	0.454	0.319	−0.436
HL	0.633	−0.249	−0.098
HB	0.587	0.077	0.315
BB	0.438	0.199	0.714
BFD	0.526	0.108	0.585
MFH	0.712	−0.108	−0.223
NH	0.639	−0.193	−0.061
NB	0.522	−0.060	0.261
TST	−0.141	0.843	0.049
SST	−0.119	0.881	0.032
Percentage of total variance explained	34.72	14.88	9.60

Relative effects of causal/classificatory variables

The qualitative results of cluster and principal component anlayses indicate that the relative effects of altitude and sex are higher than the other causal/classificatory variables. To study the relative effects in a more quantitative manner, we have performed a canonical correlation analysis, by treating the five causal/classificatory variables—sex, altitude, ethnicity-religion, location/geographical distance, and occupation as one set, and the 16 anthropometric variables as the other set. Since there are five causal/classificatory variables, five canonical variates could be constructed. The canonical correlations turned out to be 0.875, 0.833, 0.552, 0.442, and 0.370. It is clear that in the present case, only the first two canonical variates are of practical significance, since the difference between the second and the third canonical correlations is rather large. Because we are interested in studying the relative effects of causal/classificatory variables, we shall only consider the coefficients attached to the canonical variates for the first set of variables. These coefficients are presented in columns 2 and 3 of Table 5. The first canonical variate represents the effect of sex, and the second canonical variate represents the effect of altitude. This is because the coefficients attached to sex in the first and to altitude in the second canonical variate are much higher in comparison with those attached to the other variables. Compared to these two variables, the remaining variables do not seem to have much effect on body dimensions. (It should, however, be pointed out that results of canonical correlation anlaysis are not in-

TABLE 5. Standardised coefficients attached to canonical variates for the set of causal/classificatory variables

	Canonical variate			
	Both sexes combined		Male	Female
Variable	Variate-1	Variate-2	Variate-1	Variate-1
Sex	0.889	0.458	—	—
Altitude	−0.295	0.816	0.840	0.824
Ethnicity-religion	0.129	−0.101	0.373	0.388
Location/geographic distance	−0.107	0.[illegible]43	−0.198	−0.139
Occupation	0.010	0.267	0.223	0.195

dependent of the numerical codes used for discrete variables. In the results presented here, the codes used for the classificatory variables were 1, 2, 3, etc. For example, male = 1, female = 2 for sex; Sherpa = 1, Lepcha-Buddhist = 2, Lepcha-Christian = 3 for ethnicity-religion, etc. We have also analyzed the data with two other sets of codes; the results were qualitatively similar.)

Since, in the pooled data, sex differences in body dimensions were found to be strong, it was considered logical to analyze the data separately for each sex in order to check whether the ranking of the relative effects of the other four classificatory variables remained unaltered. This was done by considering separately the nine population groups for males and the eight population groups for females, with respect to the remaining four classificatory variables—altitude, location/geographical distance, ethnicity-religion, and occupation. The four canonical correlations for males were 0.860, 0.633, 0.590, and 0.476, and for females these were 0.857, 0.613, 0.570, and 0.396. It appears that for both sexes, only one canonical variate is of practical significance. The standardised coefficients attached to the first canonical variates have been presented in columns 4 and 5 of Table 5 for males and females, respectively. These coefficients indicate that even within each sex, altitude has the strongest effect on body dimensions.

Are all the anthropometric characters significant discriminators?

The fact that when all the anthropometric characters were considered jointly the population groups showed significant differences among themselves is not unexpected. It was of interest to find out whether and how well the population groups could be distinguished on the basis of a fewer number of variables. Such an attempt is useful, especially in view of the fact that significant correlations were noted between many of the characters. It may well be that once a subset of the 16 variables is considered, the remaining variables do not contribute significantly to the discriminating ability because these variables are highly correlated with the variables already considered. A reduction of dimensionality and the identification of important discriminating variables is useful both for purposes of data collection and analysis, since one will then need to collect and handle data on a smaller number of variables to reach the same conclusions regarding the population groups. Furthermore, such an identification will also yield insights into the processes of biological adaptation at high altitudes. Since the consideration of all possible subsets of variables is enormously time consuming, we resorted to a stepwise procedure (see, e.g., Roy and Majumder, 1984). This analysis was also done by combining the two sexes (that is, resulting in 17 groups) and separately for each sex. The results of the stepwise discriminant analysis for the combined data have been presented in columns 2 and 3 of Table 6. These results showed that the values of the Wilks' Λ-statistic were significant (as judged by the approximate F-statistic) at the 5% level at all steps of the procedure until the final step when all the 16 variables were introduced in the discriminant functions. This indicates that all the variables considered provide significant discrimination among the population groups when sex is also taken into consideration. Of the 16 variables, the most important discriminator variable is sitting height (SH), and the least important is head breadth (HB). One may also consider the problem of discriminating among the population groups within each sex. In this case, data were analyzed separately for males and females. Among males, as the figures in columns 4 and 5 of Table 6 show, only 14 variables were significant for purposes of discrimination. The same number of variables discriminated among females (see columns 6 and 7 of Table 6). While among males head length (HL) and height (HT) were not found to be useful for discrimination, among females the variables excluded as significant discriminators were the two head measurements HL and HB. It may be pointed out that even when both sexes are considered together, HL and HB turned out to be the least important discriminating variables. Another striking difference between the sexes is that except for the most important discriminating variable (bizygomatic breadth [BB]), the order of entry of the other variables in the discriminant functions was generally reversed in females compared to the males.

The ability to discriminate among the population groups

Having found that all the variables are significant discriminators when sex is taken into consideration, the question that arises is

TABLE 6. Results of step-wise discriminant analysis[1]

	Both sexes combined		Male		Female	
Step No.	Variable entered	Wilks' Λ	Variable entered	Wilks' Λ	Variable entered	Wilks' Λ
1	SH	0.4272	BB	0.4447	BB	0.6211
2	BB	0.2038	BAD	0.2451	BID	0.3403
3	SST	0.1134	SST	0.1736	SH	0.2507
4	BAD	0.0662	WT	0.1228	MFH	0.1901
5	BFD	0.0477	BFD	0.0912	NB	0.1493
6	WT	0.0343	BG	0.0703	NH	0.1198
7	NB	0.0248	NB	0.0555	CG	0.0981
8	BG	0.0182	CG	0.0463	BFD	0.0812
9	MFH	0.0145	TST	0.0400	HT	0.0698
10	TST	0.0117	SH	0.0364	TST	0.0596
11	NH	0.0096	NH	0.0309	WT	0.0501
12	BID	0.0080	MFH	0.0277	BG	0.0422
13	CG	0.0067	BID	0.0248	SST	0.0368
14	HT	0.0058	HB	0.0224	BAD	0.0317
15	HL	0.0051	—	—	—	—
16	HB	0.0045	—	—	—	—

[1] All values of Wilks' Λ are significant at the 5% level.

this: How well can we discriminate among the populations using these variables? Judgment on this question is made by the probability of correct classification, which is computed on the basis of the estimated linear classification functions. The classification rule is to classify an individual into that population group for which the value of the classification function (obtained by entering the measurements of the individual in question) is the maximum. To examine how well the classification functions perform, the standard practice is to classify the sampled individuals using the estimated classification functions. Since it is known that the inclusion of the data on the individual to be classified, while estimating the parameters, results in an overestimation of the probability of correct classification, we resorted to a jackknife approach in which estimation of parameters is done after eliminating the data on the individual to be classified. The results have been presented in Table 7, which show that the overall probability of correct classification is 0.71. This table also shows that only a negligible proportion of individuals actually belonging to the low-altitude population groups is misclassified as belonging to the high-altitude groups, and vice versa. Therefore, the classification functions perform very well in discriminating between high- and low-altitude groups. Within the low-altitude groups, however, the performances of the classification functions were in some cases very good (e.g., for LDASHA) while in some cases fairly poor (e.g., for LKALBA-F, LKALCA-F). This again is indicative of the relatively strong association with altitude in comparison with the other factors (e.g., occupation, geographical distance, etc.). It may be noted that in the above exercise of classification, an individual was classified only on the basis of the anthropometric measurements without any knowledge of the sex of the individual. It may be pertinent to repeat this exercise and find out how well an individual with known sex status can be classified into one of the population groups (nine for males and eight for females) on the basis of his/her anthropometric measurements. The results have been presented in Tables 8 and 9, respectively, for males and females. It is seen that the probabilities of correct classification are 0.77 and 0.64, respectively, for males and females. In corroboration of the results of previous analyses, it is seen from Tables 8 and 9 that even within each sex the probability of misclassification of a low-altitude resident to a high-altitude group, or the vice versa, is very small, while within the low-altitude groups this probability is much higher. It may also be pointed out that separate sex-specific classification functions do not increase the ability of correct classification.

DISCUSSION

We begin with a few general remarks. An inspection of the mean values of the anthropometric characters revealed that, by and

TABLE 8. Results of jackknifed classification for males

Actual population group	Percentage correctly classified	No. of individuals classified into population group								
		LKASHA-M	LKASHP-M	LKASHF-M	LKALBA-M	LKALCA-M	LNSSHA-M	LDASHA-M	LDASHP-M	HNKSHA-M
LKASHA-M	61.5	16	7	2	1	0	0	0	0	0
LKASHP-M	67.0	13	67	3	2	6	5	1	2	1
LKASHF-M	74.5	2	4	70	1	5	6	0	6	0
LKALBA-M	73.4	4	2	0	47	5	5	0	1	0
LKALCA-M	68.4	7	18	4	7	91	5	1	0	0
LNSSHA-M	85.5	0	6	2	1	0	65	0	1	1
LDASHA-M	93.8	1	0	0	0	0	0	45	2	0
LDASHP-M	85.7	0	2	1	0	0	0	1	30	1
HNKSHA-M	92.5	0	1	1	0	0	2	0	1	62
Total	76.7	43	107	83	59	107	88	48	43	65

TABLE 9. Results of jackknifed classification for females

Actual population group	Percentage correctly classified	No. of individuals classified into population group							
		LKASHA-F	LKASHP-F	LKASHF-F	LKALBA-F	LKALCA-F	LDASHA-F	LDASHP-F	HNKSHA-F
LKASHA-F	73.9	34	2	4	0	6	0	0	0
LKASHP-F	85.9	3	55	4	0	1	0	1	0
LKASHF-F	71.2	5	1	42	4	2	1	3	1
LKALBA-F	34.7	3	5	7	25	30	1	1	0
LKALCA-F	43.5	9	11	3	43	54	0	3	1
LDASHA-F	87.9	0	0	2	1	0	29	1	0
LDASHP-F	83.3	0	0	2	1	0	1	20	0
HNKSHA-F	89.5	0	0	0	0	1	1	2	34
Total	63.7	54	74	64	74	94	33	31	36

TABLE 7. *Results of jackknifed classification—both sexes combined*

Actual population group	Percent correctly classified	No. of individuals classified into population group							
		LKASHA-M	LKASHA-F	LKASHP-M	LKASHP-F	LKASHF-M	LKASHF-F	IKALBA-M	LKALBA-F
LKASHA-M	69.2	18	0	3	0	4	0	0	0
LKASHA-F	73.9	0	34	1	3	0	3	0	3
LKASHP-M	68.0	10	0	68	0	2	1	1	0
LKASHP-F	79.7	0	4	0	51	1	5	0	0
LKASHF-M	70.2	3	3	5	2	66	2	1	0
LKASHF-F	59.3	0	6	1	1	3	35	0	2
LKALBA-M	79.7	2	0	1	1	0	1	51	0
LKALBA-F	36.1	0	3	0	4	1	4	0	26
LKALCA-M	67.7	6	0	20	3	3	1	5	0
LKALCA-F	47.6	1	11	0	5	2	3	0	37
LNSSHA-M	88.2	0	0	4	0	1	0	1	0
LDASHA-M	93.8	0	0	0	0	1	0	0	0
LDASHA-F	87.9	0	1	0	0	0	1	0	0
LDASHP-M	82.9	1	1	1	0	0	1	0	0
LDASHP-F	75.0	0	0	0	2	1	0	0	0
HNKSHA-M	94.0	0	0	1	0	0	0	0	0
HNKSHA-F	89.5	0	1	0	0	0	0	0	0
Total	71.0	41	64	105	72	85	57	59	68

TABLE 7. *Results of jackknifed classification—both sexes combined (continued)*

Actual population group	No. of individuals classified into population group								
	LKALCA-M	LKALCA-F	LNSSHA-M	LDASHA-M	LDASHA-F	LDASHP-M	LDASHP-F	HNKSHA-M	HNKSHA-F
LKASHA-M	1	0	0	0	0	0	0	0	0
LKASHA-F	0	2	0	0	0	0	0	0	0
LKASHP-M	8	0	6	0	0	2	1	0	1
LKASHP-F	0	2	0	0	0	0	1	0	0
LKASHF-M	4	1	3	0	0	4	0	0	0
LKASHF-F	0	5	0	1	0	2	2	0	1
LKALBA-M	3	0	4	0	0	0	1	0	0
LKALBA-F	0	33	0	0	0	0	1	0	0
LKALCA-M	90	0	4	1	0	0	0	0	0
LKALCA-F	0	59	0	0	0	0	6	0	0
LNSSHA-M	2	0	67	0	0	1	0	0	0
LDASHA-M	0	0	0	45	0	2	0	0	0
LDASHA-F	1	0	0	0	29	0	1	0	0
LDASHP-M	0	0	0	1	0	29	0	0	1
LDASHP-F	0	1	0	0	1	1	18	0	0
HNKSHA-M	0	0	1	0	0	1	0	63	1
HNKSHA-F	0	1	0	0	1	0	1	0	34
Total	109	104	85	48	31	42	32	63	38

large, the Sherpas and Lepchas are shorter and heavier than most other Indian populations (Basu et al., 1980). The fact that there were significant sex differences and that the males had significantly larger body dimensions than females is in concordance with many earlier studies. Furthermore, the finding that females had a larger accumulation of subcutaneous fat—as measured by skinfold thicknesses—than males has been observed in many other populations (Eveleth and Tanner, 1976). The small but significant decrease in adult body dimensions with advancing age has also been reported earlier for the Andean Quechua (Frisancho, 1976).

The population groups studied belong to different ethnic categories and different physical environmental and sociocultural niches. It is, therefore, not at all unexpected that these groups revealed significant differences in respect of body dimensions. What was interesting, however, was that despite the characters being significantly correlated, none could be discarded for purposes of discrimination when both males and females were considered together.

The overall probability of misclassification was low—about 0.3. In fact, considering altitude alone, the probability of misclassifying an individual actually belonging to a high-altitude population to a low-altitude population was only 0.083, and, the probability of misclassification of the converse kind was also very small, 0.003 (see Table 7). This indicates that altitude is strongly correlated with adult body dimensions. An examination of the pattern of misclassification among the low-altitude populations did not, however, reveal any of the other factors—for instance, ethnicity-religion and occupation—to be significant.

When allocation of individuals was done on the basis of both anthropometric measurements and knowledge of the sex status, for both males and females only 14 of the 16 variables turned out to be significant discriminators. Within each sex also, the pattern of misclassification revealed that the association of altitude with body dimensions is the strongest.

It may be worth pointing out that based on several anthropometric measurements and genetic markers, Spielman and Smouse (1976) and Smouse and Spielman (1977) obtained much lower probabilities of correct classification (between 0.4 and 0.6) when they tried to classify Yanomama Indians by villages or village clusters. That we have obtained higher probabilities of correct classification is not, however, surprising since we are dealing with a more disparate—genetically, socioculturally, and environmentally—group of individuals.

These results on the relative effects of altitude, ethnicity-religion, location/geographical distance and occupation on body dimensions were also corroborated by other types of statistical analyses—cluster analysis, canonical correlation analysis, and principal component analysis.

The single linkage clustering pattern (Fig. 2) showed that the high and low-altitude population groups are clearly separated. Among the low-altitude populations, those of Darjeeling and those sampled from Kalimpong are also well separated. This is indicative of a significant role played by geographical distance. However, the male Sherpas sampled from Solu do not stand out distinctly, although the geographical distance between Solu and Darjeeling or between Solu and Kalimpong is much greater than that between Darjeeling and Kalimpong. The role played by geographical distance is, therefore, not very clear. Another interesting feature that is revealed is that the two sexes are very clearly separated in the Kalimpong samples, and also rather well in the Darjeeling samples. The pattern of clustering did not reveal any significant roles of ethnicity (because Sherpas and Lepchas did not fall in separate clusters), religion (because Christian and Buddhist Lepchas were not well separated), or occupation (because groups practicing agriculture, plantation, and forestry did not form distinct clusters).

The fact that size and shape factors could be identified is interesting, and as seen from Figure 3, these two factors alone led to a considerable separation of the population groups. Despite significant differences with the pattern of clustering depicted in Figure 2, there are considerable similarities. For example, a comparison of Figures 2 and 3 shows that the groups LKALBA-F, LKALCA-F, LKASHA-F, LKASHF-F, and LKASHP-F are in the same cluster. Differences were, of course, expected since the first two principal components only explain about 50% of the total variance. The most significant difference was that the high-altitude females seem to cluster with low-altitude Darjeeling Sherpa female plantation workers. By and large, however, the conclusions drawn from the pattern of clustering depicted in Figure 3 were similar to those derived earlier. The results of the canonical correlation analysis also showed exactly the same features.

To conclude, the findings of this study on the effect of altitude vis-à-vis geographical distance, ethnicity-religion, and occupation on adult body dimensions are: (1) altitude has a significant effect; (2) ethnicity-religion and occupation have no discernible effect; (3) the effect of geographical distance is inconsistent. Our finding of the profound effect of altitude on anthropometric variability is consistent with that obtained among the Aymaras of Bolivia (Andes) by Rothhammer and Spielman (1972). We are unaware of similar studies in other high-altitude populations. Our observation on the effect of geographical distance is, of course, intuitively understandable, as geographical distance per se is not likely to affect physical growth and development; it can work only through physical environmental, sociocultural, and/or genetic differences occurring between the study areas/populations. It is plausible that the environmental and/or genetic differences occurring between the study areas/populations in Kalimpong and Darjeeling were such that they affected the physiological processes determining adult body dimensions, while those occurring between Kalimpong and Solu study areas/populations were such that they did not. The major question that remains unanswered at this stage concerns the manner in which the composite physical environmental stresses of high altitude (i.e., hypoxia, cold, terrain, etc.), affects anthropometric characteristics. More specifically, one needs to enquire into the relative effects of genetic composition and environmental plasticity, as well as the relative effects of the various sets of environmental stresses (i.e., physical environmental and sociocultural), in the determination of a given population's response to the "altitude" stress-cluster. One may also reckon that "environment" may include many presently undefined components, and complex interactions among the various defined and undefined components.

ACKNOWLEDGMENTS

Comments of Professor J.K. Ghosh and Dr. T. Krishnan of the Indian Statistical Institute (Calcutta), Professor W.J. Schull and Mr. S. Naruya of the Center for Demographic and Population Genetics (Houston), and the anonymous reviewers, on earlier drafts of the paper, were helpful in improving the pa substantially. Thanks are due to Mr. P.K. and Mr. C. Pal for help in data transcript to Mr. P.C. Kumar for drawing the figu to Mr. K.K. Halder and Mr. R.M. Sarkar secretarial assistance, and to the Repro phy Unit, ISI, for photographic and xero services. We gratefully acknowledge th nancial and logistic support rendered by Indian Statistical Institute, Calcutta.

LITERATURE CITED

Baker, PT (ed) (1978) The Biology of High Altitud ples. Cambridge: Cambridge University Press.

Baker, PT, and Little, MA (eds) (1976) Man in the A A Multidisciplinary Study of High Altitude Que Stroudsburg, PA: Dowden, Hutchinson and Ross.

Basu, A, Majumder, PP, Ghosh, AK, and Biswa (1980) Human biological variations in Asia, wit cial reference to India. In J Hiernaux (ed): H Biological Diversity. Paris: Masson, pp. 197–274.

Clegg, EJ, Harrison, GA, and Baker, PT (1970) Th pact of high altitude on human populations. Biol. 42:486–518.

Cruz-Coke, R (1968) Genetic Characteristics of Hig titude Populations in Chile. Washington, D.C.: American Health Organization.

Das, AK (1978) The Lepchas of West Bengal. Cal Editions Indian.

Eveleth, PB, and Tanner, JM (eds) (1976) World Variation in Human Growth. Cambridge: Camb University Press.

Frisancho, AR (1976) Growth and functional de ment at high altitude. In PT Baker and MA (eds): Man in the Andes: A Multidisciplinary Stu High Altitude Quechua. Stroudsburg, PA: Do Hutchinson and Ross, pp. 180–207.

Frisancho, AR (1978) Human growth and develop among high-altitude populations. In PT Baker The Biology of High Altitude Peoples. Cambr Cambridge University Press, pp. 117–171.

Gupta, R (1981) The Impact of Altitude on Huma ulations. Calcutta: Calcutta University (Ph.D. in anthropology).

Gupta, R, and Basu, A (1981) Variations in body d sions in relation to altitude among the Sherp eastern Himalayas. Ann. Hum. Biol. 8:145–151.

Haas, JD (1976) Prenatal and infant growth and opment. In PT Baker and MA Little (eds): Man Andes: A Multidisciplinary Study of High Alt Quechua. Stroudsburg, PA: Dowden, Hutchinso Ross, pp. 161–179.

Haimendorf, C von F (1964) The Sherpas of Nepal cutta: Oxford Book Company.

Hurtado, A (1964) Acclimatization to high altitud WH Weihe (ed): The Physiological Effects of Hig titude. Oxford: Pergamon Press, pp. 1–17.

Mueller, WH, Schull, VN, Schull, WJ, Soto, P, and hammer, F (1978) A multinational Andean geneti health programme: Growth and development hypoxic environment. Ann. Hum. Biol. 5:329–35

Oppitz, M (1974) Myths and facts, reconsidering data concerning the clan history of the Sherpas. von F Haimendorf (ed): Contribution to the Anth ogy of Nepal. Warminster: Aris and Phillips Ltd 232–243.

Rao, CR (1973) Linear Statistical Inference and its Applications. New York: Wiley.

Rothhammer, F, and Spielman, R (1972) Anthropometric variation in the Aymara: Genetic, geographic and topographic contributions. Am. J. Hum. Genet. *24*:371–380.

Roy, J, and Majumder, PP (1984) Choosing a subset of variables for discrimination. Technical Report No. ASC/1984/3, Indian Statistical Institute, Calcutta.

Smouse, PE, and Spielman, RS (1977) How allocation of individuals depends on genetic differences among populations. Excerpta Medica. International Congress Series No. 411:255–260.

Spielman, RS, and Smouse, PE (1976) Multivariate classification of human populations. I. Allocation of Yanomama Indians to villages. Am. J. Hum. Genet. *28*:317–331.

UNESCO (1973) Programme on Man and the Biosphere (MAB), Working Group on Project 6: Impact of Human Activities on Mountain and Tundra Ecosystems, Final Report. Paris: UNESCO.

Weiner, JS (1969) A Guide to the Human Adaptability Proposals, IBP Handbook No. 1. Oxford: Blackwell Scientific Publications.

Weiner, JS, and Lourie, JA (1969) Human Biology: A Guide to Field Methods. IBP Handbook No. 9. Oxford: Blackwell Scientific Publications.

Weitz, CA (1984) Biocultural adaptations of the high altitude Sherpas of Nepal. In JR Lukacs (ed): The People of South Asia. New York: Plenum Press, pp. 387–420.