AMERICAN JOURNAL OF HUMAN BIOLOGY 9:415–424 (1997)

Race Differences in Reproducibilities: The HERITAGE Family Study

HABIB E. EL-MOALEM,[1] JACQUES GAGNON,[2] MIKE PROVINCE,[1] CLAUDE BOUCHARD,[2] ARTHUR S. LEON,[3] JAMES S. SKINNER,[4] JACK H. WILMORE,[5] AND D.C. RAO[1,6]*
[1]*Division of Biostatistics, Washington University Medical School, St. Louis, Missouri 63110*
[2]*Physical Activity Sciences Laboratory, Laval University, Quebec, Canada*
[3]*School of Kinesiology and Leisure Studies, University of Minnesota, Minneapolis, Minnesota 55455*
[4]*Department of Kinesiology, Indiana University, Bloomington, Indiana 47405*
[5]*Department of Kinesiology and Health Education, University of Texas, Austin, Texas 78712*
[6]*Departments of Psychiatry and Genetics, Washington University Medical School, St. Louis, Missouri 63110*

ABSTRACT The HERITAGE (**HE**alth, **RI**sk factors, exercise **T**raining **A**nd **GE**netics) Family Study is a multicenter clinical trial conducted by five institutions in the United States and Canada. The overall objective of the study is to investigate the role of the genotype in cardiovascular, metabolic, and hormonal responses to aerobic exercise training and the contribution of regular exercise to changes in cardiovascular disease and diabetes risk factors in white and black families. Since the accuracy of the assessment of the response to training depends on how repeatable or reproducible the measurements are, it is important to assess potential racial differences in reproducibilities, which may have implications for pooling data across races. The sample studied consisted of 96 blacks and 304 whites. The black sample had 46 males with mean age 33.6 ± 14.2 years and 40 females with mean age 33.9 ± 12.7 years. The white sample had 152 males with mean age 35.5 ± 14.9 years, and 152 females with mean age 34.9 ± 14.3 years. Reproducibilities, as measured by intraclass correlations among repeated measures, were comparable between whites and blacks for variables in the anthropometry, i.e, lipid, exercise test, and blood pressure domains. Reproducibilities in both races exceeded 0.85 for most of the variables. When the within-race reproducibilities are very high, statistical significance of any observed racial difference in the reproducibilities may not be very meaningful. There was a significant racial difference in the reproducibility for Apoprotein A1 (0.73 in blacks, 0.89 in whites, $P < 0.01$). However, this is not a cause for concern, since only one among 37 comparisons was significant. Am. J. Hum. Biol. 9:415–424, 1997.

It is well known that measurement error can seriously affect statistical analyses and interpretation. Hence, there is need for an index to assess the amount of such error. Measurements are made with error to a greater or lesser degree, and it is the goal of quality assurance/quality control procedures to quantify, estimate, explain, and ultimately reduce the magnitude of error to the smallest possible extent. Traditionally, two types of error are distinguished in quantitative measures: bias and precision. Bias is the degree to which there is a systematic deviation in the positive or negative direction from the "true" underlying value. Precision, in contrast, is the degree of repeat-

The HERITAGE Family Study is supported by the National Heart, Lung and Blood Institute through the following grants: HLA45670 (C. Bouchard, PI); HL47323 (A.S. Leon, PI); HL47317 (D.C. Rao, PI); HL47327 (J.S. Skinner, PI); and HL47321 (J.H. Wilmore, PI). Credit is also given to the University of Minnesota Clinical Research Center, NIH Grant MO1-RR00044400.

*Correspondence to: D.C. Rao, Washington University Medical School, Division of Biostatistics, Campus Box 8067, 660 S. Euclid Avenue, St. Louis, MO 63110-1093.

Received 13 December 1996; Accepted 14 March 1997

ability or reproducibility of the measurements, which might be determined, e.g., by the within-person standard deviation, also called technical error, or coefficient of variation of the repeated measures, or by the intraclass correlation coefficient.

The HERITAGE (**HE**alth, **RI**sk factors, exercise **T**raining **A**nd **GE**netics) Family Study is a multicenter clinical trial conducted by five institutions in the United States and Canada. The overall objective of the study is to investigate the role of the genotype in cardiovascular, metabolic, and hormonal responses to aerobic exercise training and the contribution of regular exercise to changes in cardiovascular disease and diabetes risk factors. The aims, design, and measurement protocol of the HERITAGE Family Study are described in detail by Bouchard et al. (1995). In short, the protocol called for recruitment of a total of 650 sedentary subjects comprising 90 two-generational nuclear families (father, mother, and at least three children) of Caucasian descent (white) and 40 families of African-American ancestry (black). Entire families, recruited at four clinical centers, go through an extensive test battery before and after a 20-week standardized exercise training program. Previous work by Gagnon et al. (1997) documented quality assurance and quality control measures that were implemented in the HERITAGE study. Fortunately, repeat measurements for select items in both the HERITAGE subjects, as well as a quality control study specifically designed to assess intracenter reproducibility are available. In a comprehensive study such as HERITAGE, it is important to rule out contamination of data caused by unreliable measurements so as to determine without ambiguity factors that contribute to variation in response to exercise training. Moreover, poolability of data across races is important to achieve the power to be able to detect fairly small changes in response to exercise training. Thus it is of interest to insure comparability of data quality and precision of measurements by checking to see if measurements in blacks are as reproducible as those in whites.

Finally, the terms "repeatability" and "reproducibility" are used interchangeably, namely, the extent of agreement among repeated measurements. When repeat measurements are known to differ only due to measurement error, such as when repeat measurements are taken in quick succession by the same technician, or if split blood samples are analyzed for lipids, the reproducibility or repeatability measure may be called "reliability." However, sometimes the variability in repeat measurements also includes other sources, such as day-to-day biological variability, e.g., when blood pressure is measured on separate days. Therefore, the more general term, the repeatability coefficient, or the reproducibility coefficient, is preferred. The primary goal of the present investigation was to evaluate the reproducibilities (precision) for a number of important variables (anthropometry, blood pressure, lipid, and power output at 50 W and 60%) measured in blacks and whites in the HERITAGE Family Study and to investigate potential race differences in reproducibilities.

MATERIALS AND METHODS

Anthropometry

In the anthropometric domain, height, weight, waist circumference, upper arm length, and four skinfolds (abdominal, biceps, calf, and subscapular) were taken during the same visit (prior to the training). The HERITAGE protocol required three measurements to be taken for these variables if certain criteria were not fulfilled. Otherwise, two readings sufficed. For example, a third stature measurement was obtained if there was >0.5 cm difference between the first two measurements. A third weight reading was obtained if there was >200 g difference between the first two measurements. For the circumference and skinfolds, a third measurement was taken if the first two measurements were not within 1.0 cm and 1.0 mm of each other, respectively. Details of all measurements, tests, and assays are given in the Manual of Procedures (HERITAGE, 1996).

Blood pressure

Resting blood pressures were measured on two different days during the pretraining phase using an automated colin STBP-780 unit. On each day, the first blood pressure reading was discarded and the next three valid ones were recorded (HERITAGE, 1996).

Lipids

Fasting plasma lipids were determined on two separate days with blood samples

drawn at least 24 hours apart. On each of the two days, the blood was drawn at each center and prepared according to a standard protocol before being sent to the central lab in Québec for analysis (HERITAGE, 1996). Ten blood variables were considered: plasma cholesterol, plasma triglycerides, LDL cholesterol, VLDL cholesterol, HDL cholesterol, HDL2 cholesterol, HDL3 cholesterol, apoprotein A1, total apoprotein B, and LDL apoprotein B.

Exercise tests

Three exercise tests were administered prior to training (HERITAGE 1996). The maximum rate of oxygen uptake (VO_{2max}) was determined during the first test (Max test). During the second test, participants exercised at an absolute power output of 50 W and at a relative power of 60% of their VO_{2max} determined in the initial test (Submax test). Subjects exercised for ~8 minutes at each power output to determine steady-state ventilation rate (VE), rate of oxygen uptake (VO_2), rate of carbon dioxide production (VCO_2) respiratory exchange ratio (RER = VCO_2/VO_2), systolic and diastolic blood pressure, heart rate, cardiac output (Q), and stroke volume (SV). During the third test, the participant repeated the 50 W and power output at 60% of VO_{2max} of the second test, after which the power was increased to 80% of VO_{2max} for 3 minutes and then continued until the subject reached exhaustion (Submax/Max test). The same variables were monitored during each test. Measurements were obtained in duplicate at 50 W and at 60% VO_{2max}, for both the Submax and Submax/Max tests.

Statistical methods

The reproducibility of quantitative measurements is traditionally estimated using a linear gaussian model, often an analysis of variance (ANOVA), which partitions the variance due to the different sources of error (Shrout and Fleiss, 1979). A good review of the statistical methodology may be found in Haas (1991).

Let X_{ij} be the i^{th} measurement on the j^{th} subject, $i = 1, \ldots k$, $j = 1, \ldots, n$, with k repeated measurements on n subjects. Assume the following random effects model for X_{ij}.

$$X_{ij} = \mu + \beta_j + e_{ij}$$

where μ is the overall population mean, β_j is the j^{th} subject effect assumed to be random and normally distributed with zero mean and a variance of σ^2_τ independent of the residual e_{ij} which is assumed to be normally distributed with zero mean and variance σ^2_w. The residual term includes the observer effect (if measurements are taken by multiple observers), the observer-by-subject interaction, and the error term.

The error can then be quantified on either an absolute or a relative scale, each of which has its advantages and disadvantages. The traditional estimate on the absolute scale is the within-subject standard deviation (SD), which is an estimate of σ_w, also called technical error (TE). Two relative scale measures of reproducibility are often seen in the literature: the intraclass correlation coefficient (ICC) and the coefficient of variation within subject (CV), which are defined as:

$$CV = (100^*\sigma_w)/\mu$$

$$\text{and } ICC = \sigma^2_\tau/(\sigma^2_\tau + \sigma^2_w).$$

To compute the ICC, PROC GLM in SAS was used to run an ANOVA, giving us a between-subjects mean square (*BMS*) and a within-subject mean square (*WMS*). *WMS* is an unbiased estimate of σ^2_w, and (*BMS-WMS*)/*k* is an unbiased estimate of σ^2_τ, where *k* is the number of repeat measurements. These estimates then would give a consistent estimate of the ICC (see Shrout and Fleiss, 1979):

$$ICC = (BMS—WMS)/[BMS + (k\text{-}1)\ WMS].$$

A large-sample, normal one-sided test was used to test for equality of the ICCs for blacks and whites, whereas the variance-ratio test was used to test for equality of the SDs in the two races. Although formal tests were carried out to see if the reproducibilities are comparable between the races, it should be noted that when the intraclass correlation coefficients are very high within each race (close to one), as is the case with many of the variables we studied, the large-sample test, as well as Fisher's Z-trans-

TABLE 1. Characteristics of the sample: Means and standard deviations

Data	Blacks		Whites	
	Males (n = 46) Mean SD	Females (n = 40) Mean SD	Males (n = 152) Mean SD	Females (n = 152) Mean SD
Age, yrs	33.6 ± 14.2	33.9 ± 12.7	35.5 ± 14.9	34.9 ± 14.3
Anthropometry				
Stature (height, cm)	175.8 ± 6.9	162.4 ± 8.1	178.1 ± 6.3	164.0 ± 6.0
Weight, kg	80.4 ± 15.6	70.7 ± 14.7	85.1 ± 16.0	67.4 ± 13.5
Waist, cm	87.3 ± 13.5	86.3 ± 13.8	95.3 ± 14.3	86.8 ± 14.8
Upper arm length, cm	38.4 ± 2.2	35.2 ± 1.8	38.4 ± 1.7	35.2 ± 1.9
Abdominal skinfold, mm	20.0 ± 11.7	25.4 ± 10.4	25.7 ± 11.3	26.4 ± 9.8
Biceps skinfold, mm	5.7 ± 3.5	11.8 ± 6.9	7.3 ± 4.6	11.8 ± 7.1
Calf skinfold, mm	9.4 ± 6.1	21.1 ± 9.3	12.3 ± 6.4	22.1 ± 8.2
Subscapular skinfold, mm	15.8 ± 9.1	21.4 ± 9.9	17.2 ± 8.1	17.5 ± 8.8
Blood pressure				
Systolic BP, mmHg	126.0 ± 10.0	120.9 ± 11.5	120.5 ± 10.4	113.3 ± 9.8
Diastolic BP, mmHg	71.8 ± 9.1	70.9 ± 8.5	68.4 ± 9.3	64.2 ± 7.4
Heart rate, beats/min	62.7 ± 7.9	70.3 ± 9.4	62.6 ± 8.6	66.8 ± 7.5
Lipids				
Plasma cholesterol, mmol/L	4.2 ± 0.9	4.0 ± 0.9	4.6 ± 1.0	4.5 ± 0.9
Plasma triglycerides, mmol/L	1.2 ± 0.8	0.8 ± 0.3	1.6 ± 1.0	1.2 ± 0.6
LDL cholesterol, mmol/L	2.8 ± 0.8	2.6 ± 0.8	3.1 ± 0.8	3.0 ± 0.8
VLDL cholesterol, mmol/L	0.4 ± 0.4	0.2 ± 0.1	0.6 ± 0.5	0.4 ± 0.2
HDL cholesterol, mmol/L	1.0 ± 0.2	1.2 ± 0.3	0.9 ± 0.2	1.1 ± 0.2
HDL2 cholesterol, mmol/L	0.3 ± 0.1	0.5 ± 0.3	0.3 ± 0.1	0.4 ± 0.2
HDL3 cholesterol, mmol/L	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1
Apoprotein A1, g/L	1.2 ± 0.2	1.2 ± 0.2	1.1 ± 0.2	1.2 ± 0.2
Total apoprotein B, g/L	0.8 ± 0.2	0.7 ± 0.2	0.9 ± 0.3	0.9 ± 0.2
LDL apoprotein B, g/L	0.7 ± 0.2	0.7 ± 0.2	0.8 ± 0.2	0.8 ± 0.2
Exercise test (50 W)				
Systolic blood pressure, mmHG	157.2 ± 17.3	159.4 ± 20.9	144.8 ± 17.4	142.9 ± 20.5
Diastolic blood pressure, mmHG	80.2 ± 11.1	81.5 ± 12.3	71.4 ± 11.1	70.9 ± 11.2
Heart rate, beats/min	109.9 ± 11.6	139.9 ± 18.4	107.8 ± 11.5	129.0 ± 15.2
Ventilation rate, L/min	31.8 ± 4.3	35.5 ± 6.2	31.9 ± 5.9	30.3 ± 5.1
Rate of carbon dioxide production, L/min	1018.4 ± 97.6	995.1 ± 108.2	1013.6 ± 121.5	912.8 ± 103.0
Rate of oxygen uptake, L/min	1108.4 ± 98.6	1011.4 ± 98.9	1115.4 ± 122.3	982.9 ± 96.2
Cardiac output, L/min	11.8 ± 1.5	11.2 ± 1.3	11.7 ± 1.8	11.1 ± 1.4
Stroke volume, L/min	108.7 ± 15.4	81.5 ± 13.8	109.4 ± 19.0	87.3 ± 14.2
Exercise test (60%)				
Systolic blood pressure, mmHG	182.1 ± 19.2	154.6 ± 17.1	172.6 ± 17.6	151.4 ± 16.1
Diastolic blood pressure, mmHG	81.9 ± 10.8	80.2 ± 10.9	73.1 ± 12.0	71.0 ± 11.5
Heart rate, beats/min	135.8 ± 15.1	135.2 ± 17.4	139.8 ± 16.6	142.7 ± 16.9
Ventilation rate, L/min	49.1 ± 8.6	32.9 ± 6.8	52.9 ± 8.4	36.0 ± 6.0
Rate of carbon dioxide production, L/min	1618.9 ± 330.2	929.0 ± 202.0	1784.6 ± 354.6	1103.8 ± 206.7
Rate of oxygen uptake, L/min	1647.0 ± 302.4	967.5 ± 213.3	1844.5 ± 361.1	1164.5 ± 212.3
Cardiac output, L/min	15.4 ± 2.8	11.3 ± 1.7	15.9 ± 3.1	12.2 ± 2.0
Stroke volume, L/min	114.8 ± 19.4	84.4 ± 11.9	114.4 ± 20.5	85.9 ± 13.6

formation, exaggerate differences and hence statistical significance may not be meaningful in such cases.

RESULTS

Selected clinically important variables were examined from each of the anthropometry, lipid, exercise test, and blood pressure domains. Demographic characteristics are presented in Table 1. It is clear that blacks and whites have comparable mean ages, as do males and females. Tables 2–4 show the absolute and relative measures of variability. The analyses were performed using pretraining data only. For each variable, two repeat measurements were used to examine reproducibility.

Anthropometry

The first and last valid measurements of each of the anthropometric variables were chosen for analysis of reproducibility. Since the measurements were made on the same occasion, this measures the reliability. For all anthropometric variables, the ICC was obtained separately for each race and is shown in Figure 1 and Table 2. The ICC was very high (>0.97) for all variables and was comparable in blacks and whites. The statistical significance of the race differences in

TABLE 2. Reproducibility coefficients (ICC), technical errors (TE), and coefficients of variation (CV) within subjects for antrhopometry and blood pressure in the pretraining data

Variable	ICC (blacks)	ICC (whites)	TE (blacks)	TE (whites)	CV (blacks)	CV (whites)
Anthropometry[1]						
Stature (height) (cm)	1.00	1.00	0.18	0.14*	0.11	0.08
Weight (kg)	1.00	1.00	0.09	0.17	0.12	0.22*
Waist circumference (cm)	1.00	1.00	0.74	1.01*	0.85	1.11
Upper arm length (cm)	0.97	0.99*	0.45	0.26*	1.21	0.70
Abdominal skinfold (mm)	0.99	0.99*	1.13	1.28	4.99	4.92
Biceps skinfold (mm)	0.99	0.99	0.54	0.73*	6.30	7.69
Calf skinfold (mm)	0.99	0.99	0.98	0.67*	6.61	3.89
Subscapular skinfold (mm)	0.99	0.99	0.97	0.97	5.29	5.60
Blood pressure[2]						
Systolic blood pressure (mmHG)	0.80	0.79	5.13	5.18	4.16	4.43
Diastolic blood pressure (mmHG)	0.75	0.77	4.62	4.37	6.50	6.58
Heart rate (BPM)	0.77	0.70	4.84	4.92	7.33	7.60

[1]N (blacks) = 84, N (whites) = 300.
[2]N (blacks) = 82, N (whites) = 299.
*Significantly different from the black value at the 0.01 significance level.

TABLE 3. Reproducibility coefficients (ICC), technical errors (TE), and coefficients of variation (CV) within subjects for lipids in the pretraining data[1]

Variable	ICC (blacks)	ICC (whites)	TE (blacks)	TE (whites)	CV (blacks)	CV (whites)
Plasma cholesterol (mmol/L)	0.95	0.95	0.20	0.21	4.74	4.71
Plasma triglycerides (mmol/L)	0.90	0.86	0.21	0.32	21.25	22.90
LDL cholesterol (mmol/L)	0.94	0.94	0.19	0.20	6.95	6.43
VLDL cholesterol (mmol/L)	0.88	0.87	0.10	0.14*	38.19	29.96
HDL cholesterol (mmol/L)	0.95	0.93	0.06	0.07	5.64	6.34
HDL2 cholesterol (mmol/L)	0.97	0.89*	0.04	0.06*	10.79	17.01
HDL3 cholesterol (mmol/L)	0.75	0.79	0.07	0.06	9.69	8.94
Apoprotein A1 (g/L)	0.87	0.90*	0.06	0.06*	4.90	4.88
Total apoprotein B (g/L)	0.96	0.95	0.05	0.06	5.91	6.20
LDL apoprotein B (g/L)	0.94	0.94	0.04	0.05	6.14	6.65

[1]N (blacks) = 82, N (whites) = 297.
*Race difference significant at the 0.01 level.

ICC for upper arm length and the abdominal skinfold is spurious.

Table 2 also shows the technical error (TE) and the coefficient of variation (CV) within subjects. A technical error within a subject chosen at random is the dispersion of that subject's measurements around the mean, which is then averaged over all subjects. The difference in TE between blacks and whites ranged from 0.01 for the subscapular skinfold to 0.3 for the calf skinfold. Except for weight and the abdominal and subscapular skinfolds, all other variables had a significant racial difference in the TE with whites having a higher TE than blacks for waist circumference and the biceps skinfold, and a lower TE for stature, upper arm length, and the calf skinfold.

The CV, in contrast, gives a relative measure of dispersion that is 100 times the TE scaled by the mean. Some prefer this measure of dispersion because it eliminates effects of mean differences from the dispersion comparisons. Differences in CV between blacks and whites ranged from 0.03% for stature to 2.64% for the calf skinfold. Blacks had a higher CV for stature, upper arm length, and the calf skinfold, and a lower CV for the other measurements.

Blood pressure

The average of the three valid readings on each of the two days was used to study the reproducibility of BP measurements in both races. Results of analysis using average measurements are shown in Table 2 and Figure 1. There was fairly good reproducibility, and there did not seem to be any significant differences between race-specific reproducibilities. Race differences in TE ranged from 0.14 for DBP to 0.19 for SBP, and CV differences ranged from 0.26% for DBP to 0.39% for HR. Thus it is seen that

TABLE 4. Reproducibility coefficients (ICC), technical errors (TE), and coefficients of variation (CV) within subjects for exercise test variables in the pretraining data[1]

Definition	ICC (blacks)	ICC (whites)	TE (blacks)	TE (whites)	CV (blacks)	CV (whites)
Absolute power output (50W)						
Systolic blood pressure (mmHG)	0.78	0.80	9.77	9.00	6.20	6.27
Diastolic blood pressure (mmHG)	0.73	0.74	6.72	6.17	8.35	8.70
Heart rate (BPM)	0.92	0.89	6.21	5.94	5.05	5.02
Ventilation rate (L/min)	0.81	0.85	2.57	2.12	7.71	6.86
Rate of carbon dioxide production (L/min)	0.76	0.85	54.88	50.13	5.46	5.21
Rate of oxygen uptake (L/min)	0.85	0.87	43.74	47.90	4.11	4.57
Cardiac output (L/min)	0.71	0.78	0.89	0.84	7.67	7.39
Stroke volume (L/min)	0.86	0.85	7.99	8.13	8.27	8.26
Relative power output (60% of VO_2MAX)						
Systolic blood pressure (mmHG)	0.86	0.80	9.05	9.66	5.35	5.96
Diastolic blood pressure (mmHG)	0.71	0.76	6.70	6.40	8.28	8.89
Heart rate (BPM)	0.85	0.87	6.67	6.35	4.92	4.49
Ventilation rate (L/min)	0.94	0.94	2.76	2.75	6.62	6.19
Rate of carbon dioxide production (L/min)	0.99	0.98	55.81	58.16	4.27	4.03
Rate of oxygen uptake (L/min)	0.99	0.98	46.27	57.52	3.46	3.82
Cardiac output (L/min)	0.90	0.94	1.02	0.84	7.47	5.94
Stroke volume (L/min)	0.89	0.90	7.92	7.26	7.79	7.27

[1]N (blacks) = 84, N (whites) = 298.

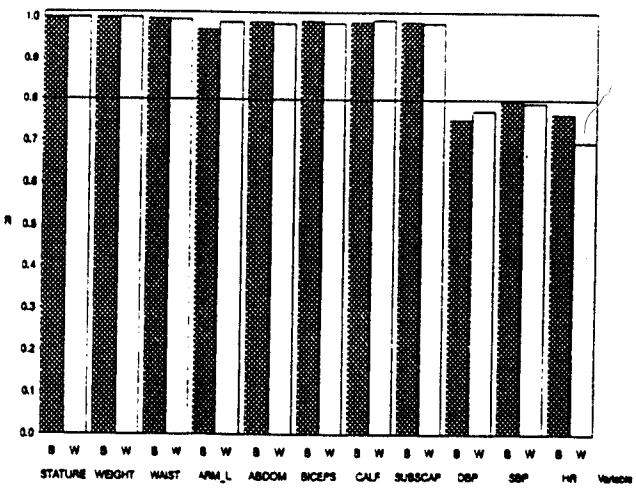

Fig. 1. Intraclass reproducibility coefficients (R) for anthropometry and blood pressure in the pretraining data. Stature height, weight = weight, waist = waist circumference, ARM L = upper arm length, ABDOM = abdominal skinfold, BICEPS = biceps skinfold, CALF = calf skinfold, SUBSCAP = subscapular skinfold, DBP = diastolic blood pressure, SBP = systolic blood pressure, HR = heart rate, B = black, W = white.

blacks and whites have comparable precision.

Lipids

The reproducibilities of 10 lipid variables were examined using fasting lipid values obtained on two separate days (Fig. 2). It is clear from the figure that both races had comparable and high ICC. Table 3 shows the ICC, TE, and CV. Apo A1 and HDL2 Cholesterol showed statistically significant differences between races in ICC and TE. The difference in TE between blacks and whites ranged from 0.001 for HDL cholesterol to 0.04 for plasma triglycerides, and the difference in CV ranged from 0.05% for plasma cholesterol to 12.9% for plasma triglycerides indicating little difference in reproducibility between the races.

Exercise tests

Reproducibilities were calculated for the exercise test variables using the four (re-

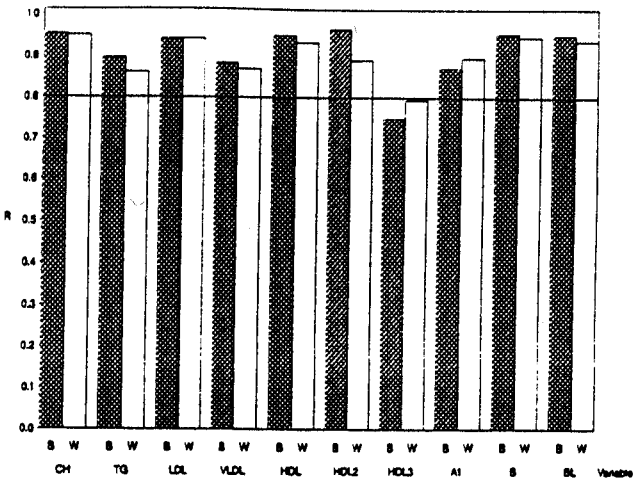

Fig. 2. Intraclass reproducibility coefficients (R) for LIPIDS in the pretraining data. CH = plasma cholesterol, TG = plasma triglycerides, LDL = LDL cholesterol, VLDL = VLDL cholesterol, HDL = HDL cholesterol (HDL-C), HDL2 = HDL2 cholesterol (HDL2-C), HDL3 = HDL3 cholesterol (HDL3—C), A1 = apoprotein A1 (apo A1), B = total apoprotein B (apo B), BL = LDL apoprotein B (LDL—apo B), B = black, W = white.

peat) measurements taken at each of 50 W and at 60% (two each under Submax and Submax/max tests). Figure 3 and Table 4 indicate comparability between blacks and whites at an absolute power output of 50 W. All variables had reasonably good ICC (>0.7) in both races. Moreover, no significant differences were noted. Figure 4 and Table 4 show similar results for the data obtained at a relative power output of 60% VO_{2max}. The ICC values in both races were fairly high (except for diastolic blood pressure) and nearly the same. None of the race differences were significant.

DISCUSSION

More than 650 participants are expected to complete the entire battery of measurements in the HERITAGE Family Study. The need to detect fairly small changes between pre- and posttraining measurements mandates the implementation of extensive quality control measures. Since the overall objective of the HERITAGE study is to investigate the role of the genotype in cardiovascular, metabolic, and hormonal responses to aerobic exercise training, it is imperative to have highly repeatable or reproducible readings. It is comforting to note that the HERITAGE measurements are highly reproducible. Comparability of the reproducibilities between whites and blacks is reassuring and justifies pooling the data for certain analyses when appropriate.

When interpreting the results, it should be remembered that the within-subject technical error gives direct information on the degree of variability on an absolute scale of measurement. However, interpretations of an absolute measure of error are difficult without knowing the population variability. Absolute measures are also difficult to compare across different scales. Marks et al. (1989) and Malina (1995) provide excellent discussions on the various measures of repeatability in general, with particular results specific to anthropometry.

Relative measures of variability automatically factor in estimates of the population level variability and are thus easier to compare across scales, since they are unitless. For instance, the ICC quantifies the amount of error relative to the total variation. High values of the ICC mean that the measurements are highly repeatable. The relative nature of the scale can be misleading, however, especially when the total variability of the underlying population is small.

Although it is certainly true that both the absolute and relative measures of reproducibility are scale dependent (i.e., in general, any monotonic transformation of the data, such as a logarithmic, will yield different CVs, TEs, and ICCs than the corresponding values on the original scale), what is most

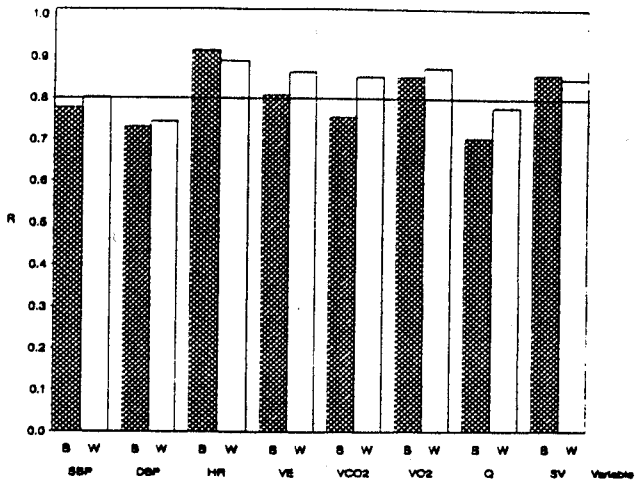

Fig. 3. Intraclass reproducibility coefficients (R) for the exercise test (50W) in the pretraining data. SBP = systolic blood pressure, DBP = diastolic blood pressure, HR = heart rate, VE = ventilation rate, VCO2 = rate of carbon dioxide production, VO2 = rate of oxygen uptake, Q = cardiac output, SV = stroke volume, B = black, W = white.

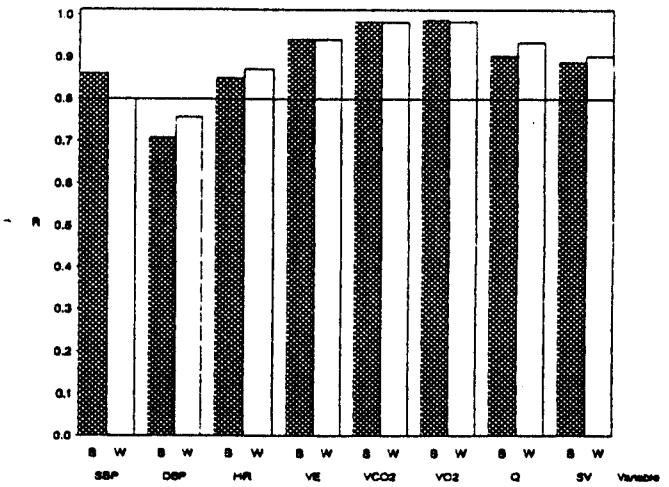

Fig. 4. Intraclass reproducibility coefficients (R) for the exercise test (60%) in the pretraining data. SBP = systolic blood pressure, DBP = diastolic blood pressure, HR = heart rate, VE = ventilation rate, VCO2 = rate of carbon dioxide production, VO2 = rate of oxygen uptake, Q = cardiac output, SV = stroke volume, B = black, W = white.

relevant here are the reproducibilities *for that scale on which actual data analyses will be done.* Whereas optimal scales for each variable for which the reproducibilities look "best" might be found, such findings would be of little interest and could even be misleading. It is much more important to quantify the degree of reproducibility on those scales that will actually be used (as reported throughout this work), since variations on those scales will ultimately affect power to estimate models and detect significant effects in data analysis.

There do not seem to be many studies in the literature similar to the current work. In assessing the etiologic role of maternal short stature, low prepregnancy weight index (BMI), and the low rate of gestational weight gain in idiopathic preterm labor, Kramer et al. (1995) demonstrated high in-

traclass correlations (>0.91) for these variables. Wilmore and Behnke (1969, 1970), in studies of both 133 young men and 128 young women, reported intraclass correlations for repeat measurements on the same day for a large battery of anthropometric dimension. The correlations ranged from 0.96 to 0.98 for skinfolds in men and were >0.93 in women. Pollock et al. (1976) conducted repeat measurements on 18 subjects and reported reliability estimates of 0.96–0.99 for skinfolds. Bouchard (1985) reported technical errors between 1.0 and 2.1 and intraclass correlations of 0.94–0.98 for six skinfold measurements in 61 children and adults of each sex, with replicate measurements taken within a 2-week period. Mueller and Malina (1987) reported intraclass correlations for five skinfold measurements in 77 adolescents, who had duplicate measurements taken within a period of 3 weeks, that ranged from 0.88 to 0.98. Wilmore et al. (submitted) found high reproducibility of anthropometric measures in the intracenter quality control substudy using the HERITAGE Family Study protocol; in the substudy, anthropometric measures were obtained on three separate days within a 3-week period at each of the four HERITAGE Clinical Centers. The intraclass correlation for the total sample, measuring reproducibility, varied from 0.97 to 1.00. Technical errors for the anthropometric dimensions were ≤1.0., and the coefficients of variation for skinfolds were <10%. The present work shows comparable results for anthropometric variables.

Zeidifard et al. (1972) determined the reproducibility of cardiac output, heart rate, and stroke volume in seven adults and three children during exercise at a VO_2 of 1200 ml/min (approximately midway between our mean values for 50 W and 60% of VO_2 max work rates). Coefficients of variation, across at least 4 days, of 5.7% for cardiac output, 6.8% for heart rate, and 5.6% for stroke volume were reported. Wolfe et al. (1978) used Pearson product-moment correlations to study the reproducibility of cardiovascular variables over two separate days at three power outputs in 20 men. The correlations ranged between 0.83–0.94 for cardiac output, 0.90–0.92 for heart rate, 0.81–0.94 for stroke volume, and 0.60 to 0.86 for systolic blood pressure. Paterson et al. (1982) conducted five repeat tests within 2–3 weeks in 12 boys at three submaximal power outputs. The coefficients of variation ranged from 6.6–8.5% for cardiac output, 4.3–6.0% for heart rate, and 7.2–10.8% for stroke volume. Kirby (1985) tested 15 subjects across three separate days and reported intraclass correlation coefficients for cardiac output of 0.69 at rest and at a lower workload and 0.87 at higher workloads. Becque et al. (1993) reported comparable intraclass correlations for steady-rate ventilation (VE), oxygen uptake (VO_2), and heart rate that ranged from 0.69 to 0.97 during submaximal cycle ergometry; reproducibilities for systolic and diastolic blood pressure (0.27 to 0.80) were lower.

Smith et al. (1993) evaluated results from 30 studies published from 1970 to 1992 to obtain estimates of the average intraindividual biological variability (CVb) in the concentrations of total cholesterol (CH), low-density lipoprotein cholesterol (LDLC), high-density lipoprotein cholesterol (HDLC), and triglyceride (TG). Composite estimates of the average CVb by different models of estimation ranged from 6.0–6.4% for CH, 6.2–7.5% for HDLC, 7.0–9.6% for LDLC, and 22.4–22.9% for TG.

In short, the pretraining HERITAGE data compare favorably with other studies and show little differences in reproducibilities between blacks and whites. The ICC as a relative measure used in this study indicates excellent reproducibility in both races. Further, the TE and CV as absolute and relative measures of precision also show that the data across races are comparable and justifies pooling of data across races when necessary.

ACKNOWLEDGMENTS

Thanks are expressed to all of the co-principal investigators, investigators, co-investigators, local project coordinators, research assistants, laboratory technicians, and secretaries who contributed to this study. The HERITAGE consortium is very thankful to those hard-working families whose participation made these data possible.

LITERATURE CITED

Becque MD, Katch V, Marks C, Dyer R (1993) Reliability and within subject variability of VE, VO2, heart rate and blood pressure during submaximum cycle ergometry. Int. J. Sports Med. 14:220–223.

Bouchard C (1985) Reproducibility of body-composition and adipose-tissue measurements in humans. In Roche AF (ed): Body-Composition Assessments in

Youth and Adults: Report of the Sixth Ross Conference on Medical Research. Columbus, OH: Ross Laboratories, pp. 9–13.
Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J (1995) The HERITAGE Family Study: Aims, design, and measurement protocol. Med. Sci. Sports Exerc. *27:*721–729.
Gagnon J, Province MA, Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH (1997) The HERITAGE Family Exercise Study: Quality assurance and quality control. Controlled Clinical Trials.
Haas M (1991) Statistical methodology for reliability studies. J. Manip. Physiol. Therap. *14:*119–132.
HERITAGE Consortium Steering Committee (1996) The HERITAGE Family Study: Manual of Procedures. Available from Physical Activity Sciences Laboratory, Laval University, Quebec, QC.
Kirby TE (1985) The CO_2 rebreathing technique for determination of cardiac output: Part I. J. Cardiac Rehab. *5:*97–101.
Kramer MS, Coates AL, Michoud MC, Dagenais S, Hamilton EF, Papageorgiou A (1995) Maternal anthropometry and idiopathic preterm labor. Obstet. Gynecol. *86:*744–748.
Malina RM (1995) Anthropometry. In PJ Maud and C Foster (eds.): Physiological Assessment of Human Fitness. Champaign, IL: Human Kinetics, pp. 205–219.
Marks GC, Habicht JP, Mueller WH (1989) Reliability, dependability, and precision of anthropometric measurements: The Second National Health and Nutrition Examination Survey 1976–1980. Am. J. Epidemiol. *130:*578–587.
Mueller WH, Malina RM (1987) Relative reliability of circumferences and skinfolds as measures of body fat distribution. Am. J. Phys. Anthropol. *72:*437–439.
Paterson DH, Cunningham DA, Plyley MJ, Blimkie CR, Donner AP (1982) The consistency of cardiac output measurement (CO_2 rebreathe) in children during exercise. Eur. J. Appl. Physiol. *49:*37–44.
Pollock ML, Hickman T, Kendrick Z, Jackson A, Linnerud AC, Dawson G (1976) Prediction of body density in young and middle-aged men. J. Appl. Physiol. *40:*300–304.
Shrout PE, Fleiss JL (1979) Intraclass correlations. Uses in assessing rater reliability. Psychol. Bull. *86:*420–428.
Smith SJ, Cooper GR, Myers GL, Sampson EJ (1993) Biological variability in concentrations of serum lipids: sources of variations among results from published studies and composite predicted values. Clin. Chem. *39:*1012–1022.
Wilmore JH, Behnke AR (1969) An anthropometric estimation of body density and lean body weight in young men. J. Appl. Physiol. *27:*25–31.
Wilmore JH, Behnke AR (1970) An anthropometric estimation of body density and lean body weight in young women. Am. J. Clin. Nutr. *23:*267–274.
Wolfe LA, Cunningham DA, Davis GM, Rechnitzer PA (1978) Reliability of noninvasive methods for measuring cardiac function in exercise. J. Appl. Physiol. *44:*55–58.
Zeidifard E, Silverman M, Godfrey S (1972) Reproducibility of indirect (CO_2) Fick method for circulation of cardiac output. J. Appl. Physiol. *33:*141–143.