International Journal of Obesity (1997) 21, 297–303

Reproducibility of anthropometric and body composition measurements: the HERITAGE Family Study

JH Wilmore[1], PR Stanforth[1], MA Domenick[1], J Gagnon[2], EW Daw[3], AS Leon[4], DC Rao[3], JS Skinner[5] and C Bouchard[2]

[1]*The University of Texas at Austin;* [2]*Laval University;* [3]*Washington University;* [4]*University of Minnesota; and* [5] *Indiana University*

OBJECTIVE: To determine the reproducibility of anthropometric and body composition measures using the HERITAGE Family Study protocol.
DESIGN: Anthropometric and body composition measures were obtained on three separate days within a 3-wk period at each of the four HERITAGE Clinical Centers.
SUBJECTS: Sixty men and women representative of the HERITAGE subject population, 15 from each of four Clinical Centers.
MEASUREMENTS: Anthropometric measures included eight skinfolds, three girths and one length; and body composition measures included stature, mass, hydrostatic weight, residual volume, and body density, from which relative fat, fat mass and fat-free mass were estimated.
RESULTS: Reproducibility as determined by technical error, coefficient of variation, and intraclass correlations was very high for the total sample. For example, intraclass correlations for the total sample generally ranged from 0.95–0.99 for the anthropometric measures, and from 0.97–1.00 for the body composition measures. The results across Clinical Centers were in close agreement with each other and with the pooled data.
CONCLUSIONS: The reproducibility of anthropometric and body composition measures using the HERITAGE Family Study protocol is sufficiently high that it should be possible to detect small changes in any of these measures and to determine the genetic basis of these changes consequent to a 20 wk endurance training program.

Keywords: reproducibility; body composition; anthropometric measurements

Introduction

The ability to accurately track changes in anthropometric and body composition variables following an exercise training program or dietary intervention is critically dependent on the reproducibility of the measurements, which reflects both trial-to-trial reliability as well as day-to-day biological variability. When either or both measurement error and day-to-day variability are close to the expected change in a variable following a given intervention, there is a strong likelihood that any true change will not be detected, namely it will not achieve statistical significance. With a significant dietary intervention, there are generally substantial changes in anthropometric measurements, including skinfold thicknesses and segment girths, and body composition measures, specifically mass, fat mass, and fat-free mass.[1,2] With endurance or resistance-type exercise training, in the absence of dietary intervention, the expected changes are considerably smaller,[3] in which case the reliability and reproducibility of measures are of much greater concern.

The HERITAGE Family Study is a large multicenter clinical trial investigating the possible genetic basis for the variability in the responses of physiological measures, and risk factors for cardiovascular disease and non-insulin-dependent diabetes mellitus, to endurance exercise training. This study includes four Clinical Centers (Indiana University, Laval University, the University of Minnesota, and The University of Texas at Austin) and a Data Coordinating Center (Washington University Medical School, St. Louis, MO). With a multicenter trial, it is critical that all clinical centers use exactly the same protocols for measurements. Further, to detect relatively small changes in critical variables, and to establish the possible genetic basis for these changes, it is crucial to have very high measurement reproducibility. Therefore, the purpose of this study was to establish the reproducibility of a series of anthropometric measures and body composition variables across three different days within a 3 wk period at each of the four Clinical Centers, as well as across the four centers combined. Details of the HERITAGE Family Study aims, experimental design, and measurement protocols have been presented in a previous publication.[4]

Correspondence: Dr JH Wilmore, Department of Kinesiology & Health Eduction, Bellmont Hall 222, The University of Texas at Austin, Austin, TX 78712, USA.
Received 4 June 1996; revised 11 December 1996; accepted 17 December 1996

Methods

Subjects

The HERITAGE Family Study subject population is composed of families, including the natural father and mother and at least two (African–American families) or three (Caucasian families) offspring 17 y of age or older. For this reproducibility study, the decision was made to use subjects who were not participants in the HERITAGE Family Study due to the nature of the study design. Each of the four Clinical Centers recruited subjects every six months over three consecutive six month periods to participate in this intracenter quality control substudy (ICQC substudy). Subjects were required to meet all criteria for admission to the HERITAGE Family Study[4] with the exception of family membership. Data were available from 60 subjects from all four centers. Their characteristics are presented in Table 1. The study protocol had been previously approved by each Clinical Center's Institutional Review Board, and informed consent was obtained from each subject.

Experimental design

Each subject completed the entire anthropometry and body composition assessment protocol on three different days within a period of 3 wk. On each of the three test days, the testing was conducted within ± 2 h of the time of testing on the other two days. For each test day, subjects reported to the laboratory at least 4 h post-prandial, having performed no formal exercise in the previous 4 h. The entire anthropometric and body composition test battery was administered on each of the three test days. This battery included the measurement of stature and body mass; one length, three girths, and eight skinfolds; and residual lung volume and hydrostatic weighing for the determination of total body density, from which relative body fat, fat mass, and fat-free mass were estimated.

When subjects arrived at the laboratory, they were questioned to ascertain if they had adhered to the 4 h post-prandial and 4 h post-exercise protocol. If they met these requirements, they were then instructed to change into a bathing suit, after which the measurement protocol was initiated.

Table 1 Physical characteristics of the 60 subjects from four clinical centers

Subjects	*Age, years*	*Height, cm*	*Weight, kg*
Women ($n=25$)	27.4 ± 8.4	165.6 ± 6.9	64.2 ± 11.5
Men ($n=35$)	28.5 ± 9.5	176.0 ± 6.1	76.4 ± 11.4
All subjects ($n=60$)	28.1 ± 9.0	171.6 ± 8.2	71.3 ± 12.9

Data are expressed as mean $\pm$ s.d.

Body composition and anthropometry methodology

Stature and body mass were measured to the nearest 0.1 kg and 0.1 cm respectively, using a balance beam scale and a stadiometer. A series of anthropometric measurements were then obtained. Measurements of upper-arm length and girth, to be used in the determination of the appropriate blood pressure cuff size for the subsequent measurement of both resting and exercise blood pressure, and abdominal and hip girth were made using a fiberglass anthropometric tape (Grafco Fiberglass Tape, Model 17-1340-2). Skinfold thickness was measured at the subscapular, biceps, triceps, midaxillary, suprailiac, abdominal, thigh, and calf skinfold sites using a Harpenden skinfold caliper (Quinton Instruments, Inc. #03496-001). The measurements were taken in accordance with the procedures recommended by Lohman, Roche, and Martorell.[5] All measurements were taken in duplicate. A third measurement was taken if the first two measurements differed by a predetermined amount: stature >0.5 cm; mass >200 g; length and girths >1.0 cm; and skinfolds >1.0 mm. When it was necessary to take a third measurement, the two closest measurements were averaged. When the third measurement fell equally between the first two, all three were averaged.

Hydrostatic weighing was used to assess body density according to the method of Behnke and Wilmore.[6] The subject was instructed to exhale completely to the point of residual lung volume, at which point a load cell interfaced with a computer was used to obtain the underwater measurement of body mass. Ten trials were obtained and the three highest values were averaged. Residual lung volume was assessed out of water in a seated position using the oxygen-dilution principle, as described by Wilmore[7] and modified by Wilmore *et al*,[8] at the Indiana, Minnesota and Texas Clinical Centers. A minimum of two trials were obtained, and a third trial was taken if the first two differed by more than 150 ml. An average of the first two trials, or the two closest trials, was used in the correction for the residual lung volume in the estimation of body density. At the Laval Clinical Center, residual volume was measured using the helium-dilution technique.[9,10] Relative body fat was estimated from body density using the equations of Siri[11] for Caucasian men, Lohman[12] for Caucasian women, Schutte *et al*[13] for African–American men, and Ortiz *et al*[14] for African–American women.

Quality assurance, quality control and statistical methodology

Important quality assurance and quality control procedures were instituted across all four Clinical Centers as described by Gagnon *et al*.[15] Staff from all Clinical

Centers were trained centrally on several occasions, and all staff from each Clinical Center had to attain certification on each technique for which they were responsible. Further, one or two staff members at each Clinical Center were responsible for all anthropometric, hydrostatic weighing, and residual volume measurements, and the same staff member was responsible for both pre- and post-training measurements on any given subject. For this quality control sample, one staff member at each Clinical Center was responsible for all measurements across the three trials. A detailed Manual of Procedures (MOP) was developed, and staff were required to review those sections of the MOP for which they were responsible every six months. Each year for the first two years of the study, a traveling crew of four subjects, two men and two women, went to each of the Clinical Centers over a 3–4 week period, and were tested according to the HERITAGE Family Study protocol at each Clinical Center, allowing comparisons to be made across the four Clinical Centers on these same four subjects.[15]

All data were analyzed using the SAS statistical package. Except where noted, data are expressed as mean ± s.d. Technical errors (TE), coefficients of variation (CV), and intraclass correlations (ICC) were computed to evaluate the reproducibility in each of the four Clinical Centers and the total combined sample using the model of Shrout and Fleiss.[16] The ICC provides an indication of the predictive value of the measurement by accounting for the population variance, while the TE is a measure of the magnitude of the error, and the CV is a measure of the magnitude of the error relative to the magnitude of the measurement. Under this model, the ith measurement on the jth subject, x_{ij}, is given by

$$x_{ij} = \mu + b_j + w_{ij}$$

where, μ is the population mean, b_j is the difference from μ of the mean of the measurements on the jth subject, and w_{ij} is the difference from $\mu + b_j$ of the ith measurement on the jth subject. Both b_j and w_{ij} are assumed to be normally distributed and independent with standard deviations of σ_τ and σ_ω respectively. σ_ω is the within-subjects standard deviation, also called technical error. The coefficient of variation within subjects was computed as:

$$\mathrm{CVW} = (100 \cdot \sigma_\omega)/\mu$$

To compute the intraclass correlation coefficient, PROC GLM in SAS was used to run an ANOVA, providing a between-subjects mean square (BMS) and a within-subjects mean square (WMS). These were used to estimate the ICC according to Shrout and Fleiss[16]:

$$\mathrm{ICC} = (\mathrm{BMS} - \mathrm{WMS})/[\mathrm{BMS} + (k-1) \cdot \mathrm{WMS}],$$

where k is the number of replicate measurements on a single subject.

A multiple testing analysis of variance (ANOVA) was implemented using the General Linear Models Procedure to assess if there were differences in measured values across the three test days for the ICQC sample. Turkey's Studentized Range (HSD) Test was used to determine between which trials there were significant differences. The multiple testing ANOVA used controls for all potential sources of variation. Statistical significance was set at the 0.05 level.

The three reliability measurements were tested for differences between centers. For the ICCs, a chi-square for homogeneity was computed, and if that chi-square was significant at the 5% level, approximate z-scores were computed between each pair of centers to determine which pairs were different, also at the 5% level. For the TE and CV, Bartlett's variance ratio test was used to determine which pairs of centers were significantly different at both the 5% and 1% levels.

Results

The mean ± standard deviations for the anthropometric variables on each of the three days for all four Clinical Centers combined are presented in Table 2, and for stature, mass, and the body composition variables in Table 3. Significant differences across trials by ANOVA were found only for the upper arm girth and the abdominal skinfold between Trial 1 and Trial 3.

The reproducibility of measurements for the four Clinical Centers combined, and for each Clinical Center individually, is presented in Tables 2 and 4 for the anthropometric variables, and in Tables 3 and 5 for stature, mass, and the body composition variables respectively. Reproducibility is presented as technical error, coefficient of variation, and intraclass correlation. Significant differences between Clinical Centers for TE, CV, and ICC (R) are indicated in Tables 4 and 5.

Discussion

At its conclusion, approximately 750 subjects will have completed the entire HERITAGE Family Study protocol, including anthropometric and body composition assessments both pre- and post-training. To determine pre- to post-training changes in these measurements, it is important to have accurate and highly reproducible measurements. The results of this study clearly demonstrate a high degree of reproducibility for the total sample of 60 subjects, and for each of the individual Clinical Centers, across all variables analyzed. While reproducibility of anthropometric and body composition measures has not been studied

Table 2 Reproducibility of anthropometric variables across three days for subjects in the intracenter quality control substudy ($n = 60$)

Variable	*Trial 1* mean ± s.d.	*Trial 2* mean ± s.d.	*Trial 3* mean ± s.d.	*TE*	*CV*	*R*
Length, cm						
Upper arm	36.9 ± 2.4	36.9 ± 2.3	36.8 ± 2.3	0.4	1.0	0.98
Girth, cm						
Upper arm	30.6 ± 3.6	30.5 ± 3.6	30.5 ± 3.6*	0.3	1.0	0.99
Abdominal	82.8 ± 11.6	82.8 ± 11.4	82.7 ± 11.4	0.9	1.0	0.99
Hip	98.4 ± 6.8	98.3 ± 6.9	98.3 ± 6.7	0.7	0.7	0.99
Skinfolds, mm						
Subscapula	13.2 ± 6.4	13.2 ± 6.4	13.2 ± 6.7	0.9	6.9	0.98
Biceps	6.1 ± 3.4	6.1 ± 3.4	6.2 ± 3.7	0.7	11.6	0.96
Triceps	12.3 ± 5.7	12.1 ± 5.6	12.1 ± 5.7	0.7	5.6	0.99
Midaxillary	10.5 ± 5.0	10.6 ± 5.4	10.5 ± 5.3	0.9	8.3	0.97
Suprailiac	18.3 ± 9.1	18.4 ± 9.0	18.3 ± 8.8	1.6	8.9	0.97
Abdominal	19.2 ± 9.0	18.8 ± 8.7	18.6 ± 8.7*	1.2	6.6	0.98
Thigh	17.6 ± 10.6	17.4 ± 10.3	17.4 ± 10.4	0.8	4.5	0.99
Calf	11.8 ± 6.5	11.6 ± 6.3	11.7 ± 6.5	1.0	8.2	0.98

*Significantly different from the first trial ($P < 0.05$).
TE = technical error; CV = coefficient of variation; R = intraclass correlation coefficient.

Table 3 Reproducibility of stature, mass, and body composition variables across three days for subjects in the intracenter quality control substudy ($n = 60$)

Variable	*Trial 1* mean ± s.d.	*Trial 2* mean ± s.d.	*Trial 3* mean ± s.d.	*TE*	*CV*	*R*
Stature, cm	171.7 ± 8.3	171.6 ± 8.2	171.7 ± 8.2	0.3	0.2	1.00
Mass, kg	71.5 ± 12.8	71.3 ± 12.9	71.3 ± 13.0	0.7	0.9	1.00
Mass in water, kg	2.56 ± 1.08	2.59 ± 1.08	2.60 ± 1.06	0.10	4.0	0.99
RLV, L	1.33 ± 0.43	1.34 ± 0.42	1.33 ± 0.41	0.07	5.6	0.97
Density, g/ml	1.052 ± 0.017	1.052 ± 0.017	1.053 ± 0.017	0.002	0.2	0.99
Relative fat, %	20.2 ± 7.5	20.0 ± 7.5	19.9 ± 7.4	0.8	4.1	0.99
Fat mass, kg	14.7 ± 7.2	14.5 ± 7.1	14.5 ± 7.0	0.6	4.3	0.99
Fat-free mass, kg	56.7 ± 9.5	56.8 ± 9.6	56.8 ± 9.8	0.6	1.1	1.00

RLV = residual lung volume
TE = technical error; CV = coefficient of variation; R = intraclass correlation coefficient.

Table 4 Reproducibility of anthropometric variables across three days at each of four clinical centers

Variable	*Clinical center A*			*Clinical center B*			*Clinical center C*			*Clinical center D*		
	TE	*CV*	*R*	*TE*	*CV*	*R*	*TE*	*CV*	*R*	*TE*	*CV*	*R*
Length, cm												
Upper arm	0.4 C**	1.2 C**	0.95	0.3	0.7 D*	0.99	0.2 A** D**	0.5 A** D**	0.99	0.5 C**	1.3 C** B*	0.96
Girth, cm												
Upper arm	0.3	1.0	1.00	0.3	0.9	0.99	0.3	1.1	0.99	0.4	1.2	0.99
Abdominal	1.0 B*	1.2 B**	0.99	0.5 A* C**	0.6 A** C**	1.00	1.1 B**	1.4 B** D*	0.99	0.7	0.8 C*	1.00
Hip	0.6	0.6	0.99	0.6	0.6	0.99	0.8	0.8	0.98	0.9	0.9	0.99
Skinfolds, mm												
Subscapula	0.7 B*	5.7	0.97	1.3 A* D*	8.5	0.98	0.9	7.0	0.99	0.7 B*	5.3	0.98
Biceps	0.6	10.9	0.93	0.7	11.8	0.97	0.4 D*	7.3 D*	0.99	1.0 C*	14.1 C*	0.84 [a]
Triceps	0.7	5.4	0.98	0.8	6.4	0.99	0.5	5.0	0.99	0.7	5.1	0.98
Midaxillary	0.5 B** C* D*	4.1 B** C** D**	0.99	1.1 A**	10.2 A**	0.97	0.8 A*	9.0 A**	0.97	0.9 A*	8.9 A**	0.94
Suprailiac	1.9	11.6	0.93	1.7	9.1	0.97	1.2	7.1	0.98	1.6	7.6	0.97
Abdominal	1.1	6.2	0.98	1.1	5.9	0.98	1.2	6.8	0.99	1.6	7.1	0.96
Thigh	0.8	4.6	0.99	0.8	4.4	1.00	0.9	5.0	0.99	0.7	4.1	0.99
Calf	1.1 B**	10.0 B**	0.96	0.5 A** D**	4.2 A** C* D**	0.99	0.9	8.7 B*	0.98	1.2 B**	9.3 B**	0.97

TE = technical error; CV = coefficient of variation; R = intraclass correlation coefficient.
[a] Fails homogeneity tests for ICCs (R), but no two centers were found different.
Note that the letters under numerical values for TE and CV indicate significant differences (* = 5% level; ** = 1% level).

Table 5 Reproducibility of stature, mass and body composition variables across three days for each of the four clinical centers

Variable	*Clinical center A*			*Clinical center B*			*Clinical center C*			*Clinical center D*		
	TE	*CV*	*R*	*TE*	*CV*	*R*	*TE*	*CV*	*R*	*TE*	*CV*	*R*
Stature, cm	0.4	0.2	1.00	0.2 D*	0.1 D*	1.00	0.3	0.2	1.00	0.4 A*	0.2 A*	1.00
Mass, kg	0.7	1.0	1.00	0.8	1.1	1.00	0.5	0.8	1.00	0.6	0.8	1.00
Mass in water, kg	0.07 C*	2.4 C*	1.00	0.08 C*	3.7	0.99	0.13 A* B*	4.9 A*	0.99	0.12	4.2	0.98
RLV, L	0.05 B*	4.1	0.99	0.1 A*	5.9	0.96	0.1	5.6	0.96	0.1	6.0	0.90
Density, g/ml	0.002	0.2	0.99	0.002	0.2	0.99	0.002	0.2	0.99	0.002	0.2	0.98
Relative fat, %	0.8	4.0	0.99	0.8	3.8	0.99	0.9	5.0	0.99	0.7	3.6	0.98
Fat mass, kg	0.6	4.3	0.99	0.7	4.1	0.99	0.7	5.0	0.99	0.6	3.6	0.99
Fat-free mass, kg	0.4	0.8	1.00	0.6	1.0	0.99	0.7	1.2	1.00	0.7	1.2	0.99

TE = technical error; CV = coefficient of variation; R = intraclass correlation coefficient; H_2O mass = mass while under water; RLV = residual lung volume.
Note that the letters under numerical values for TE and CV indicate significant differences (* = 5% level; ** = 1% level).

extensively, the results of this study are generally as good or better than what has been previously reported in the literature.

Wilmore and Behnke, in studies of both 133 young men[17] and 128 young women,[18] reported interclass correlations for repeat measurements on the same day for large battery of anthropometric measurements. The correlations ranged from 0.97–0.98 for skinfolds and from 0.98–0.99 for girths in men for the same sites used in the present study. For women, individual correlations were not reported, but out of 55 anthropometric measurement sites, 48 exhibited correlations >0.93. In a large sample of young and middle-aged men, Pollock *et al*[19] conducted repeat measurements on 18 subjects, and reported reliability estimates of 0.96–0.99 for skinfolds and 0.95–0.99 for girths. Bouchard[20] reported technical errors between 1.0 mm and 2.1 mm and intraclass correlations of from 0.94–0.98 for six skinfold measurements in 61 children and adults of each sex for replicate measures taken within a two week period. Finally, Mueller and Malina[21] reported intraclass correlations as an estimate of reliability for six girth and five skinfold measurement sites in 77 adolescents who had duplicate measurements taken within a period of three weeks. Correlations ranged from 0.88–0.98 for skinfolds and from 0.97–0.99 for girths.

Technical errors for the anthropometric measurements were generally ≤ 1.0 cm for lengths and girths and ≤ 1.0 mm for skinfolds (Table 2), which is considered very good. The suprailiac and abdominal skinfolds were exceptions. Lohman *et al*[5] have stated that the technical error for the suprailiac skinfold thickness is generally larger than those for other sites. The coefficients of variation were generally $< 10.0\%$ for skinfolds and $< 1.0\%$ for girths, well within accepted limits.

The reproducibility of stature, mass, residual volume and body composition measures was also high, with intraclass correlations for the total sample varying from 0.97–1.00. Mendez and Lukaski[22] evaluated the variability of body density in 17 young and middle-aged men measured across 4 d. Correlations between the initial and repeat measurements were 1.00 for mass, 0.97 for underwater weight, 0.87 for residual volume, and 0.99 for body density. Katch *et al*[23] reported a reliability correlation coefficient for body density measured on two separate days of 0.996 in 34 young men and women. Jackson *et al*[24] reported a correlation coefficient of 0.97 for body density measured on two different days within a 7 d period in 24 men and 44 women. Bouchard[20] reported technical errors and intraclass correlations for stature (0.3 cm and $R = 1.00$). mass (0.4 kg and $R = 1.00$), body density (0.0 gm/ml and $R = 0.97$), fat mass (1.2 kg and $R = 0.98$) and fat-free mass (1.2 kg and $R = 0.99$) in 61 children and adults of each sex for replicate measures taken within a 2 week period. Finally, Marks and Katch[25] determined the reliability of residual lung volume to be $R = 0.95$ when measured 18 times over a 10 d period in five men and five women.

The technical error for body density has ranged from 0.015[26]–0.003 gm/ml,[24] and is generally considered to be acceptable at 0.002 gm/ml,[27] which was the technical error computed for this study. The coefficient of variation for relative body fat has been reported to be 2.3% (namely percentage fat units) by McCrory *et al*,[28] considerably lower than the present study, but the two trials were obtained one immediately after the other on the same day.

Tables 4 and 5 reveal very close agreement across the four Clinical Centers for all three estimates of reproducibility; technical error, coefficient of variation and intraclass correlation. Only one variable (biceps) failed the homogeneity test for ICCs, and the post-hoc *z*-scores found no pair of centers which were significantly different, despite an ICC of $R = 0.99$ for center C and $R = 0.84$ for center D. While both of these values are within the range we generally consider good (namely $R > 0.80$), the failure to find significance between them may have more to do with the small sample size ($n = 15$ per center) than the magnitude of the difference.

Nine of 20 variables examined showed differences between centers in TE or CV at the 5% level, with four of these still showing significant differences at the 1% level. While a few differences found may be real, some are probably a result of the large number of tests conducted. Furthermore, those differences significant at the 1% level appear to be, at most, the difference between very good reliability and excellent reliability (for example upper arm length: a TE of 0.5 cm for Center D and a TE of 0.2 cm for center C are significantly different, but both are well below the 1.0 cm threshold for being considered very good).

Further, the individual Clinical Center reproducibility estimates were nearly identical to those for the total sample. Considerable effort was placed on standardizing the measurement protocol across all four centers, which included the training of all staff members from each Clinical Center at a common site on three occasions and a final certification examination. Also, detailed procedures were provided in the Manual of Procedures, both by photograph and verbal description, and staff were periodically monitored to assure compliance with these procedures.

Conclusions

In summary, the results from the analyses of the HERITAGE ICQC sample data on anthropometric and body composition variables suggest good measurement reproducibility. This is critical to any subsequent analyses of the full HERITAGE data set in which changes subsequent to endurance exercise training, and the genetic basis of these changes, are evaluated. Of particular importance was the close agreement across the four Clinical Centers, since the data from all Clinical Centers will be pooled for all subsequent analyses.

Acknowledgements

The HERITAGE Family Study is supported by the National Heart, Lung and Blood Institute through the following grants: HL45670 (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- RR000400. Further, JH Wilmore is supported by the Margie Gurley Seay Centennial Professorship and Art Leon is partially supported by the HL Taylor Professorship in Exercise Science and Health Enhancement. Thanks are expressed to all of the co-principal investigators, investigators, co-investigators, local project coordinators, research assistants, laboratory technicians, and secretaries who have contributed to this study [see Bouchard *et al*[4]]. Finally, the HERITAGE consortium is very thankful to those hard-working families whose participation has made these data possible.

References

1 Saris WHM, Koenders MC, Pannemans DLE, van Baak MA. Outcome of a multicenter outpatient weight-management program including very-low-calorie diet and exercise. *Am J Clin Nutr* 1992; **56**: 294S–296S.
2 Dengel DR, Hagberg JM, Coon PJ, Drinkwater DT, Goldberg AP. Effects of weight loss by diet alone or combined with aerobic exercise on body composition in older obese men. *Metabolism* 1994; **43**: 867–871.
3 Wilmore JH. Increasing physical activity: alterations in body mass and composition. *Am J Clin Nutr* 1963; **63** (Suppl.): 456S–460S.
4 Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J. The HERITAGE Family Study: Aims, design, and measurement protocol. *Med Sci Sports Exer* 1995; **27**: 721–729.
5 Lohman TG, Roche AF, Martorell F (eds). *Anthropometric Standardization Reference Manual*. Human Kinetics Books: Champaign, IL, 1988.
6 Behnke AR, Wilmore JH. *Evaluation and Regulation of Body Build and Composition*. Prentice-Hall: Englewood Cliffs, NJ, 1974, pp 20–24.
7 Wilmore JH. A simplified method for determination of residual lung volumes. *J Appl Physiol* 1969; **27**: 96–100.
8 Wilmore JH, Vodak PA, Parr RB, Girandola RN, Billing JE. Further simplification of a method for determination of residual lung volume. *Med Sci Sports Exer* 1980; **12**: 216–218.
9 Meneely GR, Kaltreider NL. The volume of the lung determined by helium dilution. Description of the method and comparison with other procedures. *J Clin Invest* 1949; **28**: 129–139.
10 Motley HL. Comparison of a simple helium closed with the oxygen open-circuit method for measuring residual air. *The Am Rev of Turberc Pulm Dis* 1957; **76**: 601–615.
11 Siri WE. Body composition from fluid spaces and density: analysis of methods. In: Brozek J, Henschel A (eds). *Techniques for Measuring Body Composition*. National Academy of Sciences National Research Council: Washington DC, 1961, pp 223–244.
12 Lohman TG. Applicability of body composition techniques and constants for children and youths. *Exer Sport Sci Rev* 1986; **14**: 325–357.
13 Schutte JE, Townsend EJ, Hugg J, Shoup RF, Malina RM, Blomqvist CG. Density of lean body mass is greater in Blacks than in Whites. *J Appl Physiol* 1984; **56**: 1647–1649.
14 Ortiz O, Russell M, Daley TL *et al*. Differences in skeletal muscle and bone mineral mass between black and white females and their relevance to estimates of body composition. *Am J Clin Nutr* 1992; **55**: 8–13.
15 Gagnon J, Province MA, Bouchard C *et al*. The HERITAGE Family Study: Quality assurance and quality control. *Ann Epidemiol* 1996, in review.
16 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420–428.
17 Wilmore JH, Behnke AR. An anthropometric estimation of body density and lean body weight in young men. *J Appl Physiol* 1969; **27**: 25–31.
18 Wilmore JH, Behnke AR. An anthropometric estimation of body density and lean body weight in young women. *Am J Clin Nutr* 1970; **23**: 267–274.
19 Pollock ML, Hickman T, Kendrick Z, Jackson A, Linnerud AC, Dawson G. Prediction of body density in young and middle-aged men. *J Appl Physiol* 1976; **40**: 300–304.
20 Bouchard C. 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. Ross Laboratories: Columbus, OH, 1985, pp 9–13.

21 Mueller WH, Malina RM. Relative reliability of circumferences and skinfolds as measures of body fat distribution. *Am J Phys Anthropol* 1987; **72**: 437–439.

22 Mendez J, Lukaski HC. Variability of body density in ambulatory subjects measured at different days. *Am J Clin Nutr* 1981; **34**: 78–81.

23 Katch FI, Hortobagyi T, Denahan T. Reliability and validity of a new method for the measurement of total body volume. *Res Q Exerc Sport* 1989; **60**: 286–291.

24 Jackson AS, Pollock ML, Graves JE, Mahar MT. Reliability and validity of bioelectrical impedance in determining body composition. *J Appl Physiol* 1988; **64**: 529–534.

25 Marks C, Katch V. Biological and technological variability of residual lung volume and the effect on body fat calculations. *Med Sci Sports Exerc* 1986; **18**: 485–488.

26 Akers R, Buskirk ER. An underwater weighing system utilizing 'force cube' transducers. *J Appl Physiol* 1969; **26**: 649–652.

27 Going SB. Densitometry. In: Roche AF, Heymsfield SB, Lohman TG (eds). *Human Body Composition*. Human Kinetics: Champaign, IL, 1996, pp 16–17.

28 McCrory MA, Gomez TD, Bernauer EM, Molé PA. Evaluation of a new air displacement plethysmograph for measuring human body composition. *Med Sci Sports Exerc* 1995; **27**: 1686–1691.