

Reproducibility of maximal exercise test data in the HERITAGE Family Study

JAMES S. SKINNER, KRISTINE M. WILMORE, ANNA JASKOLSKA, ARTUR JASKOLSKI, E. WARWICK DAW, TREVA RICE, JACQUES GAGNON, ARTHUR S. LEON, JACK H. WILMORE, D. C. RAO, and CLAUDE BOUCHARD

Indiana University, Bloomington, IN; Arizona State University, Tempe, AZ; Laval University, Quebec City, Quebec, CANADA; University of Minnesota, Minneapolis, MN; Texas A&M University, College Station, TX; and Washington University, St. Louis, MO

ABSTRACT

SKINNER, J. S., K. M. WILMORE, A. JASKOLSKA, A. JASKOLSKI, E. WARWICK DAW, T. RICE, J. GAGNON, A. S. LEON, J. H. WILMORE, D. C. RAO, and C. BOUCHARD. Reproducibility of maximal exercise test data in the HERITAGE Family Study. *Med. Sci. Sports Exerc.*, Vol. 31, No. 11, pp. 1623–1628, 1999. **Purpose:** The reproducibility of responses to maximal cycle ergometer testing was determined using data from the HERITAGE Family study at four Clinical Centers in the United States and Canada. **Methods:** Reproducibility was determined from maximal exercise test data obtained a) on 2 d in a sample of 390 subjects (198 men and 192 women), b) across 4 d in an Intracenter Quality Control (ICQC) substudy with 55 subjects who were not part of the main study, and c) across 2 wk in a Traveling Crew Quality Control (TCQC) substudy with the same eight subjects who were tested at each of the four centers. Reproducibility was evaluated using technical errors, coefficients of variation (CV) for repeated measures, and intraclass correlation coefficients (ICC) for selected variables obtained on the main cohort, as well as on the ICQC and TCQC substudies. **Results:** With the exception of systolic and diastolic blood pressures and respiratory exchange ratio, all the other variables (heart rate, ventilation, $\dot{V}O_2$, and $\dot{V}CO_2$) were highly reproducible, with CV below 10% and ICC over 0.86. These results were similar to those previously reported on the same subjects at a submaximal power output associated with 60% $\dot{V}O_{2max}$. Results were consistent for the main cohort, the ICQC sample, the TCQC sample, and across all four Clinical Centers. **Conclusions:** Day-to-day variations are small and reproducibility is high for maximal values of heart rate, ventilation, $\dot{V}O_2$ and $\dot{V}CO_2$ at each of the four Clinical Centers of the HERITAGE Family Study. **Key Words:** HEART RATE, BLOOD PRESSURE, VENTILATION, O_2 UPTAKE, CO_2 PRODUCTION, RESPIRATORY EXCHANGE RATIO

It is important to have accurate and reproducible measures of physiological variables during maximal exercise, especially maximal oxygen intake ($\dot{V}O_{2max}$), which is considered the best indicator of cardiorespiratory fitness (16). These maximal values are often used to prescribe exercise and to describe the intensity (% $\dot{V}O_{2max}$) of exercise. Reproducibility also is an important consideration when one wants to determine the significance of changes in $\dot{V}O_{2max}$ or other cardiorespiratory endurance phenotypes that might occur with endurance exercise training. That is, if trial-to-trial reliability is low and day-to-day biological variability is large, then it may be difficult to determine the true magnitude of the changes induced by training. These concerns are even greater when data are pooled from several centers using the same procedures, as there might be differences in reproducibility across these centers.

The HERITAGE Family study is a large, multicenter clinical trial studying the possible genetic bases for the

variation in response to endurance exercise training of various physiological measures and risk factors for cardiovascular disease and non-insulin-dependent diabetes mellitus. This study has been described in detail in a previous publication (2). The purpose of the present study was to determine the reproducibility of several cardiovascular, respiratory, and metabolic variables obtained during maximal exercise in 390 HERITAGE Family study subjects studied at the four participating Clinical Centers. In addition, reproducibility of maximal data was determined in 55 subjects tested four times within and in a group of eight subjects tested across the four Clinical Centers. The reproducibility of submaximal data was reported earlier (23).

METHODS

Subjects. The HERITAGE subjects came from families that included the natural mother and father (age 65 yr or less) and their natural children, 17–40 yr of age. The Caucasian families had at least three offspring, whereas the African-American families were smaller. This paper describes the results from the first 390 subjects (198 men and 192 women) studied at the four Clinical Centers (Arizona State University [Indiana University since January, 1996],

0195-9131/99/3111-1623/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 1999 by the American College of Sports Medicine

Submitted for publication May 1998.

Accepted for publication October 1998.

TABLE 1. Physical characteristics of the subjects in the HERITAGE Family Study, the Intracenter Quality Control (ICQC) Substudy, and the Traveling Crew Quality Control (TCQC) Substudy.

Subjects	Age (yr)	Height (cm)	Weight (kg)	Body Fat (%)
HERITAGE Family Study				
Women (N = 192)	34.8 ± 13.9	163.7 ± 6.5	68.1 ± 13.8	31.1 ± 9.7
Men (N = 198)	35.1 ± 14.7	177.6 ± 6.5	84.0 ± 16.0	22.9 ± 9.5
All subjects (N = 390)	34.9 ± 14.3	170.8 ± 9.5	76.2 ± 16.9	26.8 ± 10.4
ICQC Substudy				
Women (N = 25)	27.4 ± 8.4	165.6 ± 7.0	64.5 ± 11.5	24.4 ± 6.4
Men (N = 30)	28.5 ± 9.9	176.4 ± 6.3	76.1 ± 10.9	17.4 ± 6.3
All subjects (N = 55)	28.0 ± 9.2	171.4 ± 8.6	70.8 ± 12.5	20.6 ± 7.2
TCQC Substudy				
All subjects (N = 8)	32.6 ± 8.8	170.7 ± 6.0	68.0 ± 9.3	19.7 ± 7.6

* Data are expressed as mean ± SD.

Laval University, University of Minnesota, and The University of Texas at Austin). Subjects were healthy and sedentary and met a number of inclusion and exclusion criteria (2). They also had passed a medical examination with a physician that included a 12-lead electrocardiogram obtained at rest and during a maximal exercise test. The study protocol had been previously approved by the Human Subjects Committee at each of the four Clinical Centers. Informed consent was obtained from each subject.

Each Clinical Center recruited five additional subjects every 6 months over three 6-month periods to participate in an intracenter quality control (ICQC) substudy during the second and third year of data collection. Data were available from 55 subjects across all four centers. In addition, eight subjects (four men and four women) were tested in all four centers (four just before the onset of data collection in the first year and four during the second year) as part of an intercenter comparison. This traveling crew quality control (TCQC) substudy was done to compare results from the same subjects across centers. With the exception of family membership, the subjects in the two substudies met all criteria for admission to the HERITAGE Family Study. The physical characteristics of subjects in the three studies are presented in Table 1.

Exercise tests. Maximal, submaximal, and submaximal-maximal exercise tests were each performed on separate days before and after the 20-wk training program on a SensorMedics (Yorba Linda, CA) 800S cycle ergometer connected to a SensorMedics 2900 metabolic measurement cart. The electrocardiogram was used to monitor heart rate before exercise, during the last 15 s of each exercise stage, at the end of exercise, and for several minutes after the exercise tests. During each exercise stage, gas exchange variables ($\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}E$, and respiratory exchange ratio (RER)) were recorded as a rolling average of three 20-s intervals. The criteria for $\dot{V}O_{2max}$ were: RER > 1.1, plateau in $\dot{V}O_2$ (change of < 100 mL·min⁻¹ in the last three consecutive 20-s averages), and a heart rate (HR) within 10 beats·min⁻¹ of the maximal level predicted by age. All subjects achieved a $\dot{V}O_{2max}$ by at least one of these criteria. The majority of the exercise tests were conducted at the same time of day, with at least 48 h between tests.

For the first maximal test, subjects exercised at a power output (PO) of 50 W for 3 min, followed by increases of 25 W each 2 min until volitional exhaustion. Because a given

PO requires a similar $\dot{V}O_2$ (mL·min⁻¹) for all subjects, 50 W would require a higher % $\dot{V}O_{2max}$ for older, smaller, or less fit persons. For these subjects, the test started at 40 W, with increases of 10–20 W each 2 min thereafter; this was done to increase the number of stages they could do before reaching their maximum, as results of this test were used to select the PO for the subsequent tests.

The submaximal exercise test consisted of two stages. Subjects exercised 8–12 min at an absolute (50 W) and at a relative PO equivalent to 60% $\dot{V}O_{2max}$. This was done to get steady-state data for $\dot{V}O_2$, HR, and cardiac output before and after training.

The protocol for the submaximal-maximal exercise test was the same as that for the submaximal test, except that after exercising at 60% $\dot{V}O_{2max}$, subjects also exercised for 3 min at 80% $\dot{V}O_{2max}$. The resistance was then increased to the highest PO attained in the first maximal test. If subjects were able to pedal after 2 min, PO was increased each 2 min thereafter until they reached volitional fatigue.

Thus, each subject had duplicate submaximal and duplicate maximal exercise tests. The submaximal values for $\dot{V}O_2$, HR, and PO from the three tests were used to prescribe the various intensities for each subject's training program. Data from the submaximal tests were not used in this study but have been reported elsewhere (23).

For the ICQC substudy, subjects were tested once using the maximal exercise test protocol and then three more times over the next 2 wk using the submaximal-maximal exercise test protocol. For the TCQC substudy, subjects were tested with the maximal exercise protocol at the Quebec center. Shortly thereafter, they were tested once at each of the four Clinical Centers with the submaximal-maximal exercise protocol; these four tests were given over a 2-wk period, with at least 3 d between tests.

The HR was determined manually from electrocardiographic tracings, and values were recorded during the last 15 s of each stage. Blood pressure (BP) was obtained during the last minute of each stage using a Colin STBP-750 automated BP unit (San Antonio, TX). Two electrocondenser microphones embedded in the cuff synchronized the sound signal to the R-wave of the ECG. Earphones allowed the technician to confirm the BP values selected by the instrument's detection algorithm. At rest, the first Korotkoff sound was used to identify systolic BP, whereas the fifth Korotkoff sound (the point at which sound disappears) identified diastolic BP. During exercise, the fifth sound may not be heard, in which case the fourth sound (the point marked by the distinct, abrupt muffling of sound) was used.

Quality assurance, quality control, and statistical methodology. Several important quality assurance and quality control procedures were instituted at each of the four Clinical Centers. Staff from all Clinical Centers were trained centrally before data collection began and all staff had to be certified annually on each technique for which they were responsible. A detailed manual of procedures (MOP) was developed. Every 6 months, staff reviewed those sections of the MOP for which they were responsible. More informa-

TABLE 2. Reproducibility of cardiovascular, respiratory, and metabolic responses to maximal cycle ergometer exercise in 390 HERITAGE Family Study subjects across 2 days before training.

Variable	Max Test (mean ± SD)	Submax/Max Test (mean ± SD)	TE	CV	ICC
Heart rate (beats · min ⁻¹)	183.1 ± 15.4	182.6 ± 15.8	5.4	2.9	0.88
Systolic BP (mm Hg)	194.0 ± 26.4	190.3 ± 26.6	13.2	6.9	0.75
Diastolic BP (mm Hg)	84.1 ± 13.9	80.0* ± 14.2	9.6	11.6	0.54
VE (L · min ⁻¹)	96.1 ± 29.0	95.4 ± 27.3	9.1	9.5	0.89
VO ₂ (mL · min ⁻¹)	2358.8 ± 739.8	2341.7 ± 715.4	119.9	5.1	0.97
VCO ₂ (mL · min ⁻¹)	2715.9 ± 830.5	2712.6 ± 852.7	180.2	6.6	0.95
RER	1.16 ± 0.07	1.16 ± 0.08	0.05	4.2	0.52

TE, technical error; CV, coefficient of variation; ICC, intraclass correlation.

* Significantly lower than value obtained in the max test.

tion on quality control and quality assurance can be found in two recent publications (2,7).

All data were analyzed using the SAS statistical package. Data are expressed as mean ± SD. Technical errors (TE), coefficients of variation for repeated measures (CV), and intraclass correlation coefficients (ICC) were computed to evaluate the reproducibility in both the HERITAGE sample and the ICQC sample using the model of Shrout and Fleiss (18). With this model, the *i*th measurement on the *j*th 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 repeated measurements of the *j*th subject (or true score), and w_{ij} is a residual that encompasses the inseparable effects of the rater, the rater × subject interaction, and the error term. Both b_j and w_{ij} are assumed to be normally distributed and independent, with zero means and with standard deviations of σ_τ and σ_ω , respectively. To compute the ICC, PROC GLM in SAS was used to run a one-way repeated measures 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 (18):

$$ICC = (BMS - WMS) / [BMS + (k - 1) \times WMS]$$

where *k* is the number of replicate measurements on a subject. TE, or σ_ω , is the within-subject standard deviation, derived as the square root of the WMS from the ANOVA. The CV within subjects was computed as:

$$CVW = (100 \times \sigma_\omega) / \mu$$

The ICC, CV, and TE are all different measures of reproducibility. Whereas ICC and CV are unitless and are

therefore relative measures, the TE is scale-dependent and is therefore an absolute measure of reproducibility. The TE would be most informative for investigators who are familiar with the scale and units of measurements of a particular variable. The relative measures (ICC and CV) are generalizable to any measure.

A multiple-testing analysis of variance (ANOVA) was implemented using the general linear models (GLM) procedure to assess whether there were any differences across the two tests for the total sample and across the four tests for the ICQC sample and across the four Clinical Centers for the TCQC sample. Tukey's Studentized range (HSD) test was used to identify the source of significant differences between trials. The multiple-testing ANOVA controls for all potential sources of between-individual variation. Statistical significance was set at the 0.05 level.

RESULTS

The cardiovascular, respiratory, and metabolic responses to maximal exercise are presented in Table 2 for the total sample of 390 HERITAGE subjects across 2 d at baseline. With the exception of a drop in diastolic BP in the second test, there was little difference between the two trials in the mean values for all of the other variables. The CV were below 10% for all variables except diastolic BP, and the ICC were 0.88 to 0.95 for all variables except systolic BP, diastolic BP, and RER. Similar results were found across the four centers before and after training (data not shown).

Table 3 contains the results from the ICQC substudy on 55 subjects. There were similar results across the four trials, with the only significant difference being a lower but physiologically insignificant RER at the first maximal test. The

TABLE 3. Reproducibility of cardiovascular, respiratory, and metabolic responses to maximal cycle ergometer exercise in the 55 Intracenter Quality Control (ICQC) Substudy subjects across 4 days.

Variable	Max Test (mean ± SD)	Submax/Max Test			TE	CV	ICC
		Trial 1 (mean ± SD)	Trial 2 (mean ± SD)	Trial 3 (mean ± SD)			
Heart Rate (beats · min ⁻¹)	189.8 ± 11.5	189.3 ± 11.2	189.2 ± 10.6	188.8 ± 10.3	3.7	2.0	0.88
Systolic BP (mm Hg)	202.8 ± 23.4	195.8 ± 22.0	195.7 ± 22.1	195.7 ± 22.0	11.9	6.0	0.72
Diastolic BP (mm Hg)	82.0 ± 13.3	82.2 ± 14.1	80.7 ± 11.3	79.9 ± 10.9	8.6	10.6	0.52
VE (L · min ⁻¹)	124.8 ± 28.4	123.1 ± 26.6	124.1 ± 28.1	124.9 ± 29.3	9.1	7.3	0.90
VO ₂ (mL · min ⁻¹)	2916.1 ± 708.5	2902.2 ± 698.1	2890.0 ± 675.3	2860.7 ± 652.4	136.9	4.7	0.96
VCO ₂ (mL · min ⁻¹)	3446.5 ± 825.7	3514.7 ± 859.7	3519.8 ± 843.8	3478.2 ± 824.1	204.7	5.9	0.94
RER	1.18 ± 0.06*	1.21 ± 0.06	1.21 ± 0.07	1.21 ± 0.07	0.04	3.6	0.55

TE, technical error; CV, coefficient of variation; ICC, intraclass correlation.

* Value significantly lower than those of the other three trials ($P < 0.05$).

TABLE 4. Reproducibility of cardiovascular, respiratory, and metabolic responses to maximal cycle ergometer exercise in the eight Traveling Crew Quality Control (TCQC) Substudy subjects across four Centers over 2 weeks.

	Clinical Center				TE	CV	ICC
	Quebec (mean ± SD)	Minnesota (mean ± SD)	Texas (mean ± SD)	Arizona (mean ± SD)			
Heart rate (beats · min ⁻¹)	184.6 ± 12.1	183.3 ± 10.4	185.0 ± 9.7	181.8 ± 10.0	3.9	2.1	0.87
Systolic BP (mm Hg)	176.8 ± 13.0	176.1 ± 23.6	179.0 ± 18.4	179.3 ± 19.1	12.2	6.8	0.53
Diastolic BP (mm Hg)	71.5 ± 15.0	76.5 ± 11.8	72.5 ± 13.0	78.6 ± 7.0	12.0	16.0	-0.03
VE (L · min ⁻¹)	117.2 ± 33.4	109.7 ± 21.6	111.1 ± 23.0	114.3 ± 28.2	9.6	8.5	0.87
VO ₂ (mL · min ⁻¹)	2849.6 ± 646.3	2825.2 ± 569.3	2844.8 ± 615.1	2786.3 ± 603.8	114.6	4.1	0.97
VCO ₂ (mL · min ⁻¹)	3353.5 ± 866.4	3184.6 ± 670.0	3353.8 ± 800.2	3297.1 ± 781.9	181.5	5.5	0.95
RER	1.17 ± 0.08	1.13 ± 0.06	1.18 ± 0.08	1.18 ± 0.07	0.04	3.4	0.74

TE, technical error; CV, coefficient of variation; ICC, intraclass correlation.

TE, CV, and ICC values were similar to those found in the larger HERITAGE Family study sample shown in Table 2. Consistent with results from the HERITAGE sample, all ICC were 0.88 to 0.96 with the exception of those for systolic BP, diastolic BP, and RER.

Table 4 contains the data from the TCQC substudy on eight subjects. No significant differences were found across the tests done at the four centers. Again, with the exception of BP and RER, the ICC values for HR, ventilation, $\dot{V}O_2$, and $\dot{V}CO_2$ ranged from 0.87 to 0.97.

DISCUSSION

One of the purposes of the HERITAGE Family study is to determine the possible role of genetic factors in various phenotypes studied at baseline and in the different responses in these phenotypes to the same exercise training stimulus. The fact that we may need to detect fairly small changes in some variables and that data from four centers are being pooled to obtain an adequate sample size mandates that data be collected at each center in a highly reproducible manner. To be able to determine what role genetics might play, as well as to identify any relevant genes, it is also important to have accurate and reproducible measurements so that the sources of variation within and between families can be studied.

Because data obtained at maximal levels of exercise are so important for evaluating the response to training, the reproducibility of these variables was determined in three groups of subjects. A total of 390 HERITAGE subjects was tested twice at maximum before and after training in the four centers.

An additional 55 subjects (groups of 5 subjects in the ICQC substudy) were tested four times in each center over several 6-month periods to determine whether there was a drift in reproducibility over time. Change over time was considered important because there could have been greater error (less attention to detail, addition of new personnel who might have been trained differently, etc.) or less error (staff got better at taking measurements), both of which could influence the subsequent analyses and interpretation of results. More details about drift over time in the reproducibility of HERITAGE data can be found in a recent publication (4).

As well, eight subjects (two groups of 4 subjects) were tested in each of the four Clinical Centers. It could be argued

that the intracenter reliability is more critical because we want to detect changes resulting from training within each Clinical Center. If there is good reproducibility, then any systematic differences among the four centers should have little impact. Nevertheless, we felt that it was also important to determine whether the data across the centers were similar.

There is a fair amount of information in the literature on the reproducibility of $\dot{V}O_{2max}$ and maximal HR, but little information is available on such variables measured at maximal exercise as ventilation, RER, or BP. The reader should realize that the interpretation of the ICC in this study can be affected by the homogeneity or heterogeneity of the data. For example, maximal values for $\dot{V}O_2$ (and for ventilation due to differences in body size) are quite heterogeneous because of the large range in functional capacity within the three groups studied. Maximal values of HR are also heterogeneous due to its relationship with age. Because of this heterogeneity, one should expect higher ICC. On the other hand, maximal values of BP and RER can be quite similar in subjects with high and low functional capacities who are normotensive, as was the case in these studies. Therefore, the ICC of these data are not quite as reflective of reproducibility as would be the case for data on $\dot{V}O_2$, HR, and ventilation.

The reader also should be aware that most of the *r* values reported from the previous literature are not ICC but are Pearson product-moment correlations. Product-moment correlations are usually not appropriate for assessing reproducibility, except when there are only two measurements.

Even when studies are available, there is a problem because most of the information was obtained from small sample sizes. For example, of the 11 publications found reporting data on $\dot{V}O_{2max}$ and maximal HR on healthy subjects, five (5,6,11,17,20) had only 5–17 subjects and six (1,13,14,19,21,22) had 26–47 subjects. Five of these 11 studies also involved young students and athletes aged 18–26 yr (5,11,13,14,19), whereas six (1,6,17,20–22) had subjects aged 17–55 yr. There was one study of 16 patients aged 51–75 yr with chronic, stable congestive heart failure (CHF) who were tested to their “maximum” five or more times over a period of 3–22 months (10) and another study of peak $\dot{V}O_2$ in 51 patients aged 30–76 yr with mild to moderate CHF who were tested twice within 8 d (3). Thus, the present study has data on a larger and more heteroge-

neous sample of subjects; even the ICQC substudy had more subjects than was reported in these 13 studies.

$\dot{V}O_{2max}$. The reproducibility of $\dot{V}O_{2max}$ in data from HERITAGE (Table 2), from the ICQC substudy (Table 3) and from the TCQC substudy (Table 4) compare favorably with data available from other studies. The CV of 5.1% (HERITAGE), 4.7% (ICQC), and 4.1% (TCQC) were consistent and similar to values of 2 to 6% found in the literature (6,10,11,17). Similarly, the ICC of 0.97 (HERITAGE and TCQC) and 0.96 (ICQC) were consistent with most values presented in the literature that vary between 0.90 and 0.97 using various forms of exercise in healthy persons and CHF patients (1,3,5,13–15,17,20–22). Only the study by Taylor done in 1944 (19) reported a lower value of 0.70. Although not reported here, similar results ($r = 0.97$ – 0.99) were found before and after training across the four centers.

Maximal heart rate. The CV of 2.9% (HERITAGE), 2.0% (ICQC), and 2.1% (TCQC) were similar to values reported in the literature that range from 1.5 to 4.0% (6,11,17) in healthy subjects and 3.9% in 16 heart failure patients (10). The ICC values of 0.88 (HERITAGE and ICQC) and 0.87 (TCQC) were also similar to reported values of 0.90 (5) and 0.91 (1) but higher than values of 0.76 (13,14), 0.81 (19,22) and 0.82 (17,21). Similar results ($r = 0.82$ – 0.92) were found before and after training across the four centers.

Maximal ventilation. The CV of 9.5% (HERITAGE), 7.3% (ICQC), and 8.5% (TCQC) were similar to the only other value found in the literature of 7.0% (17). The ICC of 0.89 (HERITAGE), 0.90 (ICQC) and 0.87 (TCQC) were similar to values of 0.88 found in two studies (1,13) but higher than 0.78 (14,22), 0.73 (21), 0.58 (5), and 0.53 (19) found in five other studies. Values of 0.87 to 0.94 were found across the four centers before and after training.

Maximal RER. Although there is little or no difference in the mean values for RER (Tables 2–4), the ICC were 0.52 for HERITAGE, 0.55 for ICQC, and 0.74 for TCQC. Across the four centers before and after training, ICC ranged from 0.38 to 0.63. As these low coefficients were found in spite of the high reproducibility in the two variables used to calculate it ($\dot{V}O_2 = 0.96, 0.97, \text{ and } 0.97$ and $\dot{V}CO_2 = 0.94, 0.95, \text{ and } 0.95$ for HERITAGE, ICQC, and TCQC, respectively), it appears that RER should still be used as one criterion (e.g., $RER > 1.10$) to evaluate whether subjects have pushed themselves to maximum. Although the smaller ICC values for RER suggest that it is not highly reproducible, the ICC may be artificially small because the variability in maximal RER is also low as the subjects had to have an $RER > 1.10$ to be considered at maximum. Our values were similar to reported values of 0.52 (14), 0.48 (21), and 0.44 (22) but higher than 0.23 found in swimmers (13). No data were found in the literature on CV for maximal RER.

Blood pressure. We found little published data on the reproducibility of systolic BP at maximal exercise. The only CV found was obtained on 16 patients aged 51–75 yr with chronic stable congestive heart failure (10) who exercised to exhaustion during five or more incremental treadmill tests over a period of 3–22 months. The CV of 6.7% in that study

was similar to those found for the HERITAGE data (6.9%), the ICQC data (6.0%), and the TCQC data (6.8%). Relative to ICC, only one study was found. In this study (9), 156 CHD patients were tested, and the mean values were compared with those obtained 9 months later. The test-retest reliability coefficient was 0.67, a value similar to 0.75 from HERITAGE and 0.72 from ICQC but higher than 0.53 from TCQC. Data from these two studies may not be comparable because the clinical condition of the patients could have remained the same, improved, or worsened over time.

No data were found in the literature for diastolic BP during maximal exercise. The low ICC in HERITAGE (0.54), ICQC (0.52), TCQC (-0.03), and across the four centers (0.32 to 0.66) before and after training were not surprising, however, as it has been shown that the error in measuring diastolic BP becomes greater as exercise intensity increases (8). Although fourth-phase diastolic BP tends to remain steady during dynamic exercise in healthy subjects, attempting to find the point where the Korotkoff sounds disappear (fifth-phase diastolic BP) is often difficult because of the noise and movement associated with exercise (especially maximal exercise). In addition, even though the diastolic BP might not change, these sounds can often be heard all the way down to zero pressure (12), resulting in variable and inaccurate measurements. Therefore, only the fourth phase should be used to reflect diastolic BP during vigorous exercise, and the fifth-phase diastolic BP should be used at rest.

CONCLUSIONS

Similar results were found in the HERITAGE, ICQC, and TCQC data regarding the ICC values. For the most part, the results were consistent across the four Clinical Centers for each variable. There was no drift over time in these variables (4). There is also fair agreement from pre- to post-training and no trend for the ICC to become higher or lower after training. The data presented here also were similar to those obtained during submaximal exercise on the same subjects, as reported by Wilmore et al. (23), particularly at 60% $\dot{V}O_{2max}$.

With the exception of BP and RER, it is concluded that the day-to-day variations in responses to maximal cycle ergometer exercise were low in the HERITAGE subjects both before and after training, over several 6-month periods (ICQC) and in the same subjects tested in all four centers over a 2-wk period (TCQC). Thus, the reproducibility of responses to maximal exercise was high at each of the four centers in the HERITAGE Family Study. Reproducibility was excellent (CVs below 10% and ICC over 0.86) in those variables considered most important for evaluating the response to endurance exercise training (i.e., $\dot{V}O_{2max}$, $\dot{V}CO_{2max}$, maximal HR, and maximal ventilation). This was due in part to the careful and constant program of quality control and quality assurance implemented in the HERITAGE Family Study (7). This high level of reproducibility provides a robust foundation for investigating the role of

genetic and nongenetic factors in phenotypes associated with maximal exercise.

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. Jack H. Wilmore was partially supported by the Margie Gurley Seay Centennial Professorship, and

Arthur S. Leon is partially supported by the Henry L. Taylor endowed 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. Finally, the HERITAGE consortium is very thankful to those hard-working families whose participation has made these data possible.

Address for correspondence to: James S. Skinner, Department of Kinesiology, Indiana University, 11001 North 26th Street, Phoenix, AZ 85028; E-mail: js Skinner@iupui.edu.

REFERENCES

1. AUNOLA, S., and H. RUSKO. Reproducibility of aerobic and anaerobic thresholds in 20–50 year old men. *Eur. J. Appl. Physiol.* 53:260–266, 1984.
2. BOUCHARD, C., A. S. LEON, D. C. RAO, J. S. SKINNER, J. H. WILMORE, and J. GAGNON. The HERITAGE Family Study: aims, design, and measurement protocol. *Med. Sci. Sports Exerc.* 27: 721–729, 1995.
3. COHEN-SOLAL, A., F. ZANNAD, J. G. KAYANAKIS, P. GUERET, J. F. AUPETIT, and H. KOLSKY. Multicentre study of the determination of peak oxygen uptake and ventilatory threshold during bicycle exercise in chronic heart failure. *Eur. Heart J.* 12:1055–1063, 1991.
4. DAW, E. W., M. A. PROVINCE, J. GAGNON, et al. Reproducibility of the HERITAGE Family Study intervention protocol: drift over time. *Ann. Epidemiol.* 7:452–462, 1997.
5. FERGUSON, R. J., G. G. MARCOTTE, and R. R. MONTPETIT. A maximal oxygen uptake test during ice skating. *Med. Sci. Sports* 1:207–211, 1969.
6. FROELICHER, V. F., M. C. MAJ, H. BRAMMELL, et al. A comparison of the reproducibility and physiologic response to three maximal treadmill exercise protocols. *Chest* 65:512–517, 1974.
7. GAGNON, J., M. A. PROVINCE, C. BOUCHARD, et al. The HERITAGE Family Study: quality assurance and quality control. *Ann. Epidemiol.* 6:520–529, 1996.
8. GRIFFIN, S. E., R. A. ROBERGS, and V. H. HEYWARD. Blood pressure measurement during exercise: a review. *Med. Sci. Sports Exerc.* 29:149–159, 1997.
9. IRVING, J. B., R. A. BRUCE, and T. A. DEROUEN. Variations in and significance of systolic pressure during maximal exercise (treadmill) testing: relation to severity of coronary artery disease and cardiac mortality. *Am. J. Cardiol.* 39:841–843, 1977.
10. JANICKI, J. S., S. GUPTA, S. T. FERRIS, and P. A. McELROY. Long-term reproducibility of respiratory gas exchange measurements during exercise in patients with stable cardiac failure. *Chest* 97: 12–17, 1990.
11. KATCH, V. L., S. S. SADY, and P. FREEDSON. Biological variability in maximum aerobic power. *Med. Sci. Sports Exerc.* 14:21–25, 1982.
12. MADSEN, E. B., and V. FROELICHER. The use of exercise testing to evaluate patients after myocardial infarction In: *Heart Disease and Rehabilitation*, 2nd Ed. M. Pollock and D. Schmidt (Eds.). New York: Wiley, 1986, pp. 455–476.
13. MAGEL, J. R., and J. A. FAULKNER. Maximum oxygen uptake of college swimmers. *J. Appl. Physiol.* 22:929–933, 1967.
14. McARDLE, W. D., F. I. KATCH, G. PECHAR, L. JACOBSON, and S. RUCK. Reliability and interrelationships between maximal oxygen intake, physical work capacity, and step-test scores in college women. *Med. Sci. Sports Exerc.* 4:182–186, 1972.
15. McARDLE, W. D., F. I. KATCH, and G. PECHAR. Comparison of continuous and discontinuous treadmill and bicycle tests for max $\dot{V}O_2$. *Med. Sci. Sports Exerc.* 5:156–160, 1973.
16. MITCHELL, J. H., B. J. SPROULE, and C. B. CHAPMAN. The physiological meaning of the maximal oxygen intake test. *J. Clin. Invest.* 37:538–547, 1958.
17. NORDREHAUG, J. E., R. DANIELSEN, L. STANGELAND, G. A. ROSLAND, and H. VIK-MO. Respiratory gas exchange during treadmill exercise testing: reproducibility and comparison of different exercise protocols. *Scand. J. Clin. Lab. Invest.* 51:655–658, 1991.
18. SHROUT, P. E., and J. L. FLEISS. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86:420–428, 1979.
19. TAYLOR, C. Some properties of maximal and submaximal exercise with reference to physiological variation and the measurement of exercise tolerance. *Am. J. Physiol.* 142:200–212, 1944.
20. TAYLOR, H. L., E. BUSKIRK, and A. HENSCHEL. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J. Appl. Physiol.* 8:73–80, 1955.
21. WILMORE, J. H., J. A. DAVIS, R. S. O'BRIEN, P. A. VODAK, G. R. WALDER, and E. A. AMSTERDAM. Physiological alterations consequent to 20-week conditioning programs of bicycling, tennis, and jogging. *Med. Sci. Sports Exerc.* 12:1–8, 1980.
22. WILMORE, J. H., B. J. FREUND, M. J. JOYNER, et al. Acute response to submaximal and maximal exercise consequent to beta-adrenergic blockade: implications for the prescription of exercise. *Am. J. Cardiol.* 55:135D–141D, 1985.
23. WILMORE, J. H., P. R. STANFORTH, K. R. TURLEY, et al. Reproducibility of cardiovascular, respiratory and metabolic responses to submaximal exercise: the HERITAGE Family Study. *Med. Sci. Sports Exerc.* 30:259–265, 1998