Validity and Reproducibility of Crutch Force and Heart Rate Measurements to Assess Energy Expenditure of Paraplegic Gait

Maarten J. IJzerman, PhD, PT, Gert Baardman, PhD, Martin A. van 't Hof, PhD, Herman B. K. Boom, PhD, Hermie J. Hermens, PhD, Peter H. Veltink, PhD

ABSTRACT. IJzerman MJ, Baardman G, van 't Hof MA, Boom HBK, Hermens HJ, Veltink PH. Validity and reproducibility of crutch force and heart rate measurements to assess energy expenditure of paraplegic gait. Arch Phys Med Rehabil 1999;80: 1017-23.

Objective: To determine the validity and reproducibility of heart rate (HR) and crutch force measurements to estimate energy expenditure during paraplegic walking. Usefulness of these outcome measures in comparative trials was assessed in terms of responsiveness.

Design: Cross-sectional validity was determined using one single (first) measurement. Longitudinal validity as well as reproducibility were calculated using repeated measurements.

Setting: Oxygen uptake and HR during steady state as well as axial crutch load were measured at subjects' self-selected walking speeds.

Patients: Ten subjects with thoracic-level spinal cord injury were included in the study. All subjects had considerable experience with ambulation in the advanced reciprocating gait orthosis (ARGO).

Main Outcome Measures: Oxygen uptake ($\dot{V}O_2$, mL/min) and oxygen cost (EO_2, mL/m) were used as criterion standards. Crutch peak force (CPF), crutch force time integral (CFTI), HR, and physiological cost index (PCI) were used to estimate energy expenditure.

Results: The PCI was found to be sensitive to detect differences between sessions in criterion standard ($r = .86$). Smallest detectable difference (ie, point where difference exceeds measurement error) ranged from approximately 15% for CPF to 33.7% and 41.8% for EO_2 and PCI, respectively.

Conclusions: Although PCI is expected to be a valid measure for within-patient differences in $\dot{V}O_2$, responsiveness was lower compared to EO_2 and CPF. The limited number of patients who can be included in studies on paraplegic locomotion requires reproducible outcome measures. Therefore, CPF and EO_2 are advocated in favor of PCI.

© *1999 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation*

ASSESSMENT OF THE energy expenditure of walking is frequently performed to evaluate the effectiveness of walking systems for patients with paraplegia.[1-5] For this purpose, measurement of oxygen uptake ($\dot{V}O_2$) and oxygen cost (EO_2) are the most uniformly accepted standards.[1,3-5] Usually, oxygen uptake is expressed either per unit time (oxygen uptake, $\dot{V}O_2$) or per unit distance (oxygen cost, EO_2).[4] EO_2 is thus considered an efficiency measure, because it relates the $\dot{V}O_2$ to the performance. During steady state exercise, $\dot{V}O_2$ as well as EO_2 can be transferred into energy uptake (J/min/kg) and energy cost (J/m/kg), respectively.[6,7] In that case, the respiratory exchange ratio (RER) can be used to determine the energy production of 1L of oxygen.[7] Some objections can be made, however, against transformation of EO_2 to energy cost in paraplegic walking because of the disturbed metabolic pathways and partially anaerobic energy supply.[8]

Although $\dot{V}O_2$ measurements are a primary choice for assessing energy expenditure, they are cumbersome to conduct, the instrumentation is expensive for a routine laboratory, and the measurements require trained personnel.[5] Thus, for estimating energy expenditure, other parameters have been used instead, including the physiological cost index (PCI)[2,5,9-12] and crutch forces.[13-15] PCI is usually calculated by dividing the difference in heart rate (HR) between rest and steady state by walking speed and has been proposed to replace $\dot{V}O_2$ measurements because of the close association between HR and $\dot{V}O_2$ in submaximal exercise and during steady state conditions.[9] An implicit assumption made while using PCI is that the sympathetic control of HR is unaffected in patients with high-level spinal cord injury (SCI) (T6 and above). Although it is difficult to justify this assumption, Bar-on and Nene[16] conducted an arm ergometry study in SCI patients and found a linear relation between HR and $\dot{V}O_2$ in a group of patients with high-level SCI. This study does indicate that both HR and $\dot{V}O_2$ increase at higher workloads and that there should be cardiac control during exercise, either via a sympathetic drive or via humoral mechanisms. More importantly, this finding indicates that, even in patients with high-level SCI, HR may be used to estimate differences in $\dot{V}O_2$ within patients.

Winchester and colleagues[5] found a statistically significant difference in the PCI between reciprocating gait orthosis (RGO) and isocentric RGO, whereas EO_2 did not change significantly. They consequently claimed sensitivity of the PCI to detect changes in energy expenditure. This conclusion, however, can only be justified after studying the behavior of the PCI with respect to EO_2.

A high test-retest reliability of PCI has been claimed by Nene and Jennings,[9] who presented data from two consecutive tests in healthy adults and adolescents. Although they presented a nonparametric test for systematic errors, they did not calculate reproducibility statistics.

Typically, two parameters can be derived from crutch forces, ie, crutch force time integral (CFTI)[13,17] and crutch peak force (CPF) or peak axial load.[17,18] CFTI has been advocated in preference over PCI especially in subjects with lesions between T1 and T6, because of their loss of control of HR.[13] A second

advantage of CFTI could be that it is directly related to shoulder muscle effort and not affected by factors that may influence physiologic measurements.[13] Although CFTI and CPF are frequently used as outcome measures, neither data about their relation with $\dot{V}O_2$ during walking nor data about their reproducibility are available.

This study investigated whether HR and crutch force measurements can be used to detect differences in energy expenditure between groups (cross-sectional validity) as well as within individuals (longitudinal validity or sensitivity to change).[19] Reproducibility and responsiveness of these variables were studied to draw conclusions on their usefulness in comparative trials on energy expenditure during paraplegic walking.

ESTIMATION OF OXYGEN UPTAKE ($\dot{V}O_2$) AND OXYGEN COST (Eo_2)

Engsberg and colleagues[10] studied the relation of PCI and energy expenditure and compared PCI with $\dot{V}O_2$.[10] Although PCI and $\dot{V}O_2$ may have a conceptual relation, we prefered to divide estimators of energy expenditure into those that estimate "uptake" ($\dot{V}O_2$:mL/min) and those that estimate "cost" (Eo_2:mL/m). HR (beats/min) has been described as an estimator of $\dot{V}O_2$; PCI (beats/m) should thus be related to Eo_2. Katz and coworkers[20] reported a high correlation of PCI and Eo_2 ($r_{x,y} = .88$) in a comparative trial of two hip-knee-ankle-foot orthoses for children with myelomeningocele. This correlation between PCI and Eo_2 is considered spurious, however, since both PCI and Eo_2 are obtained by normalizing for walking speed (ie, $PCI = (HR_{ss} - HR_{rest})/v$, and $Eo_2 = \dot{V}o_2/v$, where "ss" is steady state walking, "rest" is at rest, and "v" is walking speed). Dividing two sets of independent observations (x, y) by the same denominator (z) may yield those spurious correlations.[21]

Because of the statistical errors in the relation between PCI and Eo_2, we will study the validity of the PCI by means of the relation between ($HR_{ss} - HR_{rest}$) and $\dot{V}O_2$. It is assumed that these variables represent the same conceptual relation but are not subject to statistical errors.

Although axial crutch forces probably do not provide sufficient information on the actual energy expenditure of the upper body, many authors have interpreted axial crutch forces as being an estimate of $\dot{V}O_2$.[13-15] Assessment of the validity of axial crutch forces in estimating $\dot{V}O_2$ is required. At least two parameters can be derived from crutch force assessments: CFTI over the stride and CPF, which is the required force on both crutches to obtain foot clearance during midswing phase (fig 1). Because CPF and CFTI are not normalized for either walking speed or stride length, we will relate these variables to $\dot{V}O_2$ instead of Eo_2.

METHODS

Subjects

Ten subjects with paraplegia were included in the study (table 1). All subjects had a complete thoracic lesion and were previously provided with an advanced reciprocating gait orthosis (ARGO) by a qualified orthotist. Level of SCI was between T4 (three subjects) and T12 (two subjects). Five subjects had T8 or T9 lesions. One subject had severe extension spasticity, and seven subjects had only marginal spasticity. Prescription of the orthosis and an extensive training by a physical therapist in use of the ARGO was conducted prior to inclusion in the study. Exclusion criteria for brace prescription were, among others, cardiac and pulmonary dysfunction as well as severe shoulder
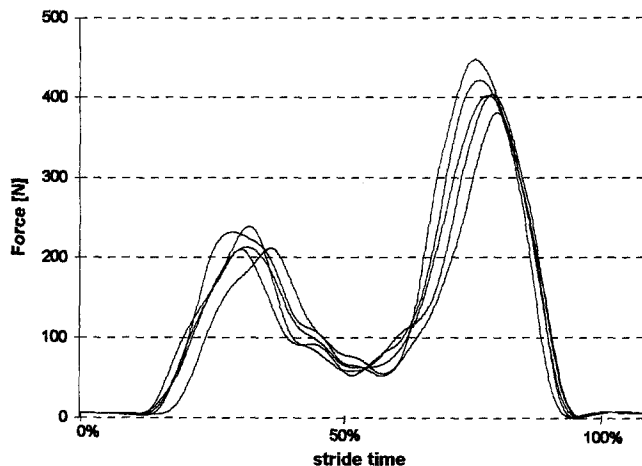

Fig 1. Typical crutch force profiles of one subject during one trial comprising five strides. Usually five trials were performed to obtain approximately 20 to 30 strides. Left crutch force is presented from left heel strike to left heel strike. The first peak corresponds with the crutch peak force on stance side (left crutch is stance side), the second peak with crutch peak force on the swing side.

and arm complaints. Important criteria for inclusion to the present study were sufficient experience with walking in the ARGO and the ability to walk independently for at least 15 minutes. Informed consent was obtained from all subjects before inclusion in the study. The study was approved by the local medical ethics committee.

Study Design

All subjects were asked to come to the gait laboratory three times with an interval of 2 weeks between the measurements. The first measurement was conducted to familiarize subjects with the testing procedures. The second measurement was used to determine cross-sectional validity of HR_{ss}, CPF, CFTI, and ($HR_{ss} - HR_{rest}$). The second (test) and third (retest) measurements were used to determine longitudinal validity of these variables. Reproducibility and responsiveness was determined using the second and third measurements as well.

Measurements

A standardized measurement protocol was used including physiologic, spatiotemporal, and kinetic assessments.[17] All measurements were done in the same sequence. Repeated measurements for each subject were done at the same time of the day. All measurements were performed at a self-selected, comfortable walking speed. Subjects were asked to refrain from tobacco, coffee, and food for at least 2 hours before arriving at the gait laboratory.

Table 1: Relevant Data of Study Sample (n = 10)

	Mean	σ	Range
Age (yrs)	36.2	9.8	28.0, 59.0
Weight (kg)	73.0	14.4	53.0, 100.0
Lean body mass (kg)	55.0	10.9	39.0, 75.0
Walking speed (m/sec)	.21	.10	.09, 0.41
Stride length (m)	.87	.17	.49, 1.05
Cadence (strides/min)	15.3	3.6	11.1, 23.3

Physiologic Assessment

Breath-by-breath measurement of energy expenditure was conducted by means of a metabolic cart.[a] Subjects were provided with a facemask containing a flexible gas tube, which did not interfere with the subjects' walking pattern. Rest metabolism was assessed for 5 minutes, after which subjects were asked to stand up. When HR approached a stable level, patients were asked to walk at a comfortable, self-selected speed for 10 minutes along a 125-meter pathway. HR (beats/min), $\dot{V}O_2$ (mL/min/kg), $\dot{V}CO_2$ (mL/min/kg), RER, tidal volume (L), breathing frequency (breaths/min), and expiratory volume (L/min) were measured.

HR was determined during rest (HR_{rest}) as well as during steady state (HR_{ss}). HR_{ss} is defined as the average heart rate during steady state walking, where steady state walking is visually judged using plateau values of expiratory volume (V_e) and $\dot{V}O_2$. In previous experiments steady state during walking was found to be delayed (within approximately 6 minutes) in a group of spinal cord injured subjects.[17,22] Consequently, steady state EO_2 and PCI could only be calculated during the last few minutes of the 10 minute walk according to:

$$EO_2 = \frac{\dot{V}O_2}{v} \ (mL/m/kg) \tag{1}$$

$$PCI = \frac{HR_{ss} - HR_{rest}}{v} \ (beats/m) \tag{2}$$

where v = walking speed during steady state.

No transformation of oxygen cost to energy cost was made because of the expected anaerobic contribution to energy supply and because of the disturbed basal metabolic pathways.[8]

Kinetic and Spatiotemporal Assessments

Biomechanical assessments were made in the gait laboratory on a fixed walking lane of 7.5m. Crutch force data was sampled at 200Hz by means of strain gauges.[b] Force data were filtered off-line using a linear phase second order Butterworth filter (F_{3dB} = 5Hz). All data were split into strides representing a right-left step sequence. CFTI over the stride, CPF, stride length, and walking speed (m/sec) were calculated. CFTI and CPF of left and right crutch were averaged and normalized for body weight.[23] Each assessment comprised five trials to be ascertained from approximately 20 to 30 strides for averaging (fig 1).

Assessment of Validity, Reproducibility, and Responsiveness

Validity. Assessment of validity is generally conducted by calculating the association between the new variable and a criterion standard. Validity is strongly dependent on the type of outcome measure: should it be used either discriminatively, predictively, or evaluatively.[19,24,25] Cross-sectional validity will be determined to assess the ability of the new variable to discriminate between subjects. Longitudinal validity is the ability of the new variable to detect (clinically important) changes in criterion standard across patients or within patients (evaluative purpose).[19] However, instead of applying clinically important differences it will be determined by using the (small) test-retest differences in criterion standard ($\delta_{1-2}\dot{V}O_2$). It is assumed that, if such small differences can be detected, the outcome measure will also be able to detect larger differences that may be of more relevance.

Reproducibility and responsiveness. Lack of reproducibility may be either a problem of systematic errors or random errors.[26,27] The intraclass correlation coefficient (ICC) is often preferred over the Pearson's correlation as a measure of reproducibility, because it combines systematic and random errors into one statistic.[26,28-31] Interpretation of Pearson's correlation, as well as of ICC, however, can sometimes be misleading, because the (cross-sectional) between-subject variance is the variance of interest, whereas the within-subject variance is more relevant to detect longitudinal changes.[26,32-34] The approach to detect longitudinal changes in time is to calculate the standard error of measurement (SE_m) or mean square error (MS_{error}).[28,32,33,35] The SE_m can be used to calculate the smallest detectable difference (SDD), which is the point where the difference between two consecutive assessments exceeds the measurement error or noise.[33,34]

An additional criterion that is required to assess the usefulness of health status instruments is responsiveness.[29,32] To be able to detect a clinically relevant difference, a high signal-to-noise ratio is required. This signal-to-noise ratio for an evaluative instrument is called the responsiveness index.[25] The estimate of "signal" is, for instance, the difference that is considered clinically relevant.[32] The estimate for "noise" may be the within-subject variability in stable subjects.[29,32,33,35]

Statistical Analysis

Before statistical testing all data were checked for statistical assumptions.

Cross-sectional validity. Scatter plots were made for HR_{ss}, CFTI, CPF, and ($HR_{ss} - HR_{rest}$) versus $\dot{V}O_2$. Pearson product moment correlation coefficients ($r_{x,y}$) were calculated to estimate the association between criterion standard and either CFTI, CPF, HR_{ss}, or ($HR_{ss} - HR_{rest}$). Fisher z transformation was used to calculate a 95% confidence interval (CI) for the correlation coefficient.[36] Correlations above .80 were considered good.

Longitudinal validity. Longitudinal validity was determined by correlating differences (δ_{1-2}) between test and retest measurements for HR_{ss}, CFTI, CPF, and ($HR_{ss} - HR_{rest}$) with test-retest differences in the criterion standard ($\delta_{1-2}\dot{V}O_2$). Pearson's product moment correlation coefficient ($r_{\delta x, \delta y}$) was calculated and Fisher z transformation was used to calculate a 95% CI for the correlation.[36]

Reproducibility and responsiveness. Analysis of variance (ANOVA) was conducted to estimate the between-subject, within-subject, and residual variance components.[30,31,33,34] $ICC_{2,1}$ were calculated using a random effects model.[30,31] SE_m, SDD, and responsiveness (Guyatt's effect size) were calculated according to:

$$SE_m = \sqrt{MS_{error}} \ ^{.33-35}$$

$$SDD = 2.23 * \sqrt{2} * SE_m \ ^{.33,34}$$

$$\text{Effect size} = \frac{\Delta}{\sqrt{2 \cdot MS_{error}}} \ ^{.25,32}$$

where: Δ = clinically relevant difference and $\sqrt{2}.MS_{error}$ = within-subject variability ($\sigma_{\delta i}$).[25]

Guyatt's effect size was calculated using a clinically relevant difference (Δ) of 20%, which was arbitrarily chosen in a study on the influence of the reciprocal cable linkage in an ARGO.[17]

The SE_m and SDD are expressed in the unit of measurement of the variable. SDD is also given as relative difference with respect to the baseline ARGO assessment. The SDD is inter-
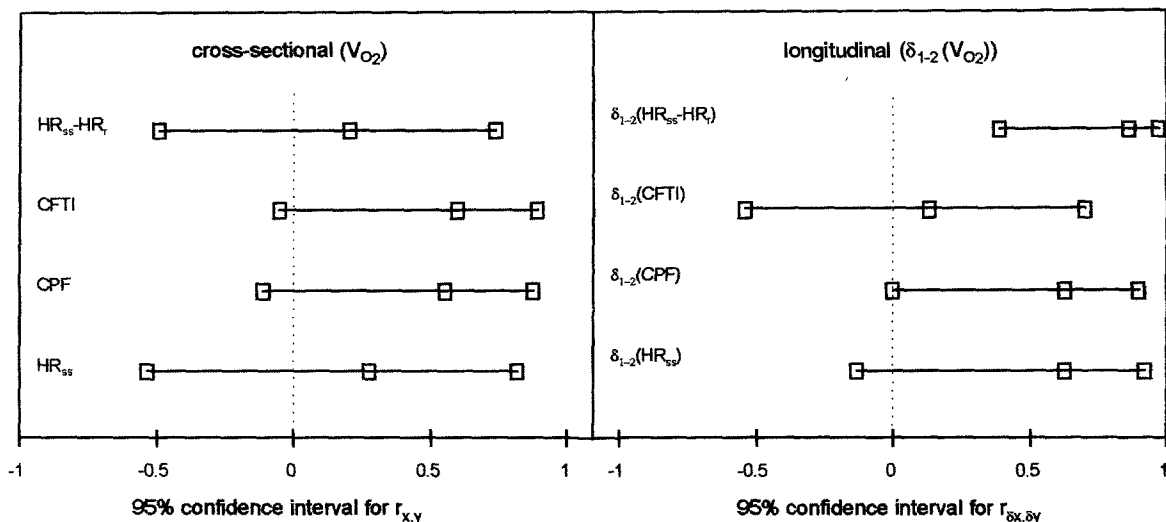
Fig 2. The 95% confidence intervals for the Pearson's correlation coefficients of each of the outcome measures. Confidence intervals are calculated using a Fisher z transformation. None of the outcome measures appeared to be cross-sectionally valid (left part of figure). Longitudinal validity (right part of figure) is studied using test-retest differences in criterion standard, V_{O_2}. ($HR_{SS} - HR_{rest}$) is valid to detect changes in V_{O_2}. It is thus assumed that PCI will be valid to detect changes in E_{O_2}.

preted as a confidence interval, ie, the difference should be at least the SDD to be 95% sure of a true difference.[33] The value 2.23 is chosen, rather than 1.96, to achieve 95% confidence in this sample of 10 subjects. All analyses were done using SPSS.[c] A p level of .05 was considered significant if statistical testing was performed.

RESULTS

During the trials it appeared that one subject was an outlier because he walked considerably slower than the other subjects: he was a 59-year-old man with a T9 lesion. Although slow, he walked safely and could maintain the experiments without any problems.

Repeated test of HR on the second measurement day failed in two patients because of technical reasons; thus, a total of eight subjects were available for the analysis of longitudinal validity and reproducibility of HR_{SS}. Also, eight subjects were available to study reproducibility of the PCI.

Cross-Sectional Validity

None of the variables appeared to be cross-sectionally valid in predicting V_{O_2}. Pearson's correlation coefficients for each of the outcome measures with V_{O_2} are presented in figure 2. CFTI and CPF had the highest correlation of .60 and .56, respectively. HR during steady state has no cross-sectional relation with V_{O_2} ($r_{x,y} = .28$). The relation between PCI and E_{O_2} was studied by means of the correlation between ($HR_{SS} - HR_{rest}$) and V_{O_2} ($r_{x,y} = .21$) (fig 3). As the correlation was only .21, we concluded that PCI will not be able to compare E_{O_2} between subjects.

Longitudinal Validity

Pearson correlation coefficients for each of the outcome measures with δ_{1-2} (V_{O_2}) are presented in figure 2. The differences in V_{O_2} (δ_y) between the consecutive measurements in the same orthosis were very small and not statistically significant (table 2). Nevertheless, they could be detected quite accurately by means of ($HR_{SS} - HR_{rest}$) ($r_{\delta x,\delta y} = .86$, fig 4). Detection of within-patient differences in V_{O_2} (test-retest) was

moderate for HR_{SS} as well as CPF ($r_{\delta x,\delta y} = .63$ and $r_{\delta x,\delta y} = .63$ respectively, fig 2). CFTI cannot be used to detect small differences in V_{O_2} between two consecutive measurements ($r_{\delta x,\delta y} = .13$).

Reproducibility and Responsiveness

Table 2 summarizes mean and standard deviation (range for skewed variables) of first and second assessments in ARGO ($ARGO_1$ and $ARGO_2$). No systematic measurement errors were found between two consecutive assessments in ARGO. E_{O_2} was not significantly reduced during the second assessment. Table 3 presents the relevant reproducibility indexes of each of the outcome measures. SDD is presented as absolute as well as relative compared with baseline.

ICCs ranged from .99 to .60 (table 3); ICC of both E_{O_2} and PCI were considerably high (.94 and .92, respectively). However, the SDDs of PCI and E_{O_2} were high and differences over 41.8% and 33.7%, respectively, should be found, before one can conclude that there is a detectable change beyond measurement
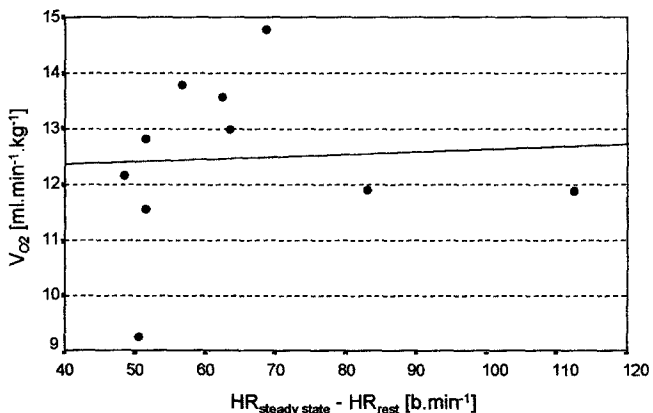

Fig 3. Plot of cross-sectional relation between V_{O_2} and ($HR_{SS} - HR_{rest}$). Pearson's correlation coefficient ($r_{x,y}$): .21 (95% CI: −.48, .74).

Table 2: Crude Data of Test and Retest With 95% Confidence Intervals for the Difference

	ARGO₁ $\bar{x} \pm \sigma$ (or range)	ARGO₂ $\bar{x} \pm \sigma$ (or range)	95% CI for $\delta_{1\text{-}2}$
$\dot{V}O_2$	17.6 (2.0)	17.6 (3.5)	−1.85, 1.77
EO_2	1.69 (.86, 3.48)	1.56 (.81, 3.30)	−.06, .31*
v	.21 (.09, 0.41)	.21 (.10, 0.43)	−.01, .01*
HR_{SS}	141.3 (20.2)	145.6 (26.2)	−16.13, 7.46
PCI	5.9 (2.7, 12.3)	6.0 (2.4, 10.8)	−1.10, .96*
CFTI	6.54 (1.44)	6.56 (1.59)	−.53, .48
CPF	4.44 (.38)	4.42 (0.41)	−.19, .23

Abbreviations: ARGO₁ or ₂, advanced reciprocating gait orthosis (1, first assessment; 2, second assessment); CI, confidence interval; $\dot{V}O_2$, oxygen uptake; EO_2, oxygen cost; v, walking speed; HR_{SS}, heart rate at steady state; PCI, physiological cost index; CFTI, crutch force time interval; CPF, crutch peak force.
* Calculated after transformation to "normality."

Table 3: Reproducibility Statistics

	ICC₂,₁	SE_m	SDD	SDD (%)	Responsiveness*
$\dot{V}O_2$	.60	1.77	5.58	31.7	1.41
EO_2	.94	.18	.57	33.7	1.33†
v	.99	.01	.03	14.7	2.97†
HR_{SS}	.81	9.98	31.47	22.3	2.0
PCI	.92	.87	2.47	41.8	.96†
CFTI	.89	.50	1.57	24.0	1.85
CPF	.71	.21	.66	14.8	2.98

Abbreviations: ICC₂,₁, intraclass correlation; SE_m, standard error of measurement; SDD, smallest detectable difference; $\dot{V}O_2$, oxygen uptake; EO_2, oxygen cost; v, walking speed; HR_{SS}, heart rate at steady state; PCI, physiological cost index; CFTI, crutch force time integral; CPF, crutch peak force.
* Using Guyatt's effect size.
† Calculated after transformation to "normality."

error. Although the ICC of CPF was moderate (.71), the SDD was only 14.8%. Walking speed appeared a reproducible outcome measure as well (SDD = 14.7%). Guyatt's effect size (responsiveness) ranged from .96 to 2.98 for PCI and CPF, respectively (table 3).

DISCUSSION

In this study we found that none of the outcome measures appeared cross-sectionally valid in predicting $\dot{V}O_2$. CFTI and CPF had the highest correlation of approximately .60 (95% CI: −.10, .98). Correlations of either HR_{ss} or (HR_{ss} − HR_{rest}) with $\dot{V}O_2$ were very low.

It is important to conclude that heart rate cannot be used to compare subjects' $\dot{V}O_2$ during steady-state walking. The slope of the relation between HR and $\dot{V}O_2$ depends, for instance, on the subjects' maximal aerobic power and a comparison of HR_{ss} would mainly reflect differences in maximal aerobic power.[6] PCI is calculated by subtracting HR_{rest} from HR_{ss} and this might be considered as a correction for differences in maximal aerobic power, ie, subjects with a high HR_{rest} will also have a high HR during walking at the same workload.[12] However, it is doubtful whether objective measurements made on a resting individual reveal the capacity for physical exercise on maximal aerobic power.[6] A low HR_{rest} may indicate a high aerobic power, but not necessarily. This may be an explanation for the lack of correlation between (HR_{ss} − HR_{rest}) and $\dot{V}O_2$ found in this study.
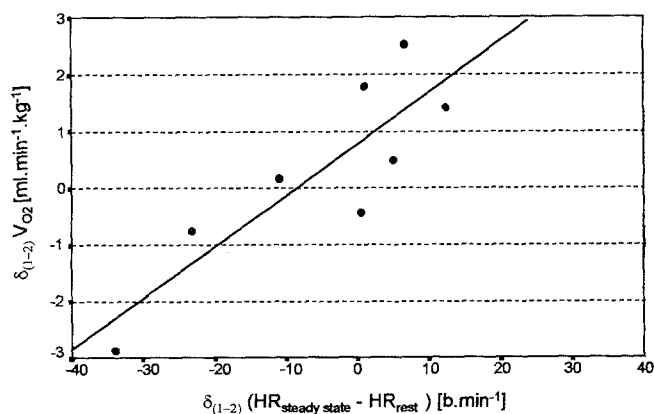

Fig 4. (Longitudinal) validity of (HR_{ss} − HR_{rest}) to detect changes in $\dot{V}O_2$. Correlation ($r_{\delta x, \delta y}$) approached .86, which was considered a good relation (95% CI: .39, .97).

Because (HR_{ss} − HR_{rest}) cannot be used to compare subjects' $\dot{V}O_2$ during walking, it is concluded that PCI cannot be used either. A practical consequence is that it is not useful to calculate subjects' PCIs while walking in a specific orthosis to compare the data with other studies involving different subjects.

Though PCI cannot be used in a between-subject comparison, it should be noted that most authors have used the PCI in a within-subject comparison.[2,5,15] Although the physiologic explanation appears adequate (Astrand and Rodahl[6]: linear relation between HR and $\dot{V}O_2$), it is surprising that so many authors have used an outcome measure for which validity has not yet been assessed. Since (HR_{ss} − HR_{rest}) is valid to predict within-subject differences in $\dot{V}O_2$ ($r_{\delta x, \delta y}$ = .86), it is expected that PCI will be able to detect changes in EO_2. Although promising, a few remarks should be made if the use of PCI is being considered in a comparative trial. Although a linear relation has been shown between HR and $\dot{V}O_2$ in high-level SCIs, one must be aware that the sympathetic drive can be either absent or affected. Use of PCI in such patients should thus be considered carefully.

HR has a close linear relation with $\dot{V}O_2$, but only up to submaximal loads.[6,7] It is described that this relation can bend in healthy subjects at loads above submaximal, ie, $\dot{V}O_2$ still increases but HR has reached a maximum.[6] It can be doubted whether walking in patients with paraplegia is a submaximal workload and as a consequence whether differences in HR accurately predict differences in $\dot{V}O_2$.

The reproducibility part of this study promotes studying the within-subject variability (ie, SE_m and SDD) rather than the ICCs alone. The SDD is not dependent on between-subject variability and thus gives a better view on the usefulness of the variables for evaluative purposes. SDD of walking speed and CPF are approximately 15% and are preferable as outcome measure in evaluative studies. A difference in $\dot{V}O_2$ and EO_2 of approximately 35% should be measured before it can be considered as a true difference. PCI is even worse; a difference of more than 40% is considered a true difference.

Responsiveness in this context is an additional criterion in which a signal (important difference) is combined with the noise in stable subjects (within-subject variability). The effect size that is calculated combines the clinically relevant difference with the within-subject variability in stable subjects.[32] However, it is quite often a problem to specify a clinically relevant difference without having sufficient clinical experience with the outcome measure and its use in clinical trials.[32,33] In a study on the influence of the reciprocal cable linkage in the ARGO, the clinically relevant difference for the main outcome measures was arbitrarily set at 20%.[17] Using this difference,

responsiveness of the main outcome measures $\dot{V}o_2$ and Eo_2 is 1.41 and 1.33, respectively. Responsiveness can best be interpreted with regard to the sample size that is required to detect this difference. Guyatt and colleagues provided a simple formula for sample size calculations using the effect size.[32] To detect a difference of 20% in Eo_2 and $\dot{V}o_2$ at least 7 and 6 subjects (paired observations) should be included ($n_{subjects}$: $[(Z_\alpha + Z_\beta)\sigma/\Delta]^2$, where $\sigma = \sqrt{2} * MS_{error}$ $(EO2) = 26$, $\Delta = .20 * 1.69 = .338$, and $Z_{\alpha = .05} + Z_{\beta = .10} = 3.24$). Comparably, 13 subjects should be included to detect a difference of 20% in PCI. It is concluded from the previous calculations that, although PCI appeared to be valid to detect differences in Eo_2, the probability of a type II error will increase when using PCI as main outcome measure in a small study. It is assumed that loss of sympathetic control of HR in patient with lesions above T6 contributes to more variability.[16] In addition, HR_{rest} is subject to variation because of parasympathetic nerve activity. Subtracting HR_{rest} from HR_{ss} may cause additional variability.

Because only a limited number of subjects can be included in the majority of studies on paraplegic locomotion, outcome variables like walking speed and CPF should be preferred because of their reproducibility. Although this study has not shown that CPF is valid for detecting (small) test-retest differences in $\dot{V}o_2$, CPF may possibly be used if the differences become larger and more relevant to clinical practice. Moreover, apart from being an estimate of differences in $\dot{V}o_2$, CPF is probably more related to the local muscle fatigue and the occurrence of wrist and shoulder pathology.[37] From this perspective and because of its high reproducibility, CPF is very relevant as an outcome measure in comparative trials on orthoses for people with SCI.

CONCLUSIONS

None of the outcome measures that were assessed in this study appeared to be cross-sectionally valid in predicting $\dot{V}o_2$. We conclude that only ($HR_{ss} - HR_{rest}$) can be used to detect small differences in $\dot{V}o_2$. Comparably, ($HR_{ss} - HR_{rest}$) normalized for walking speed (PCI) will be valid to detect longitudinal changes in Eo_2. In addition, it is expected that CPF may be used if the differences approach clinically relevant differences and thus larger than the test-retest differences in this study.

Interpretation of reproducibility and responsiveness is performed with reference to sample size required to detect a specified clinically relevant difference. It is concluded that, compared with Eo_2, PCI should not be the main outcome measure, because the limited number of patients who can be included in a comparative trial would result in too low statistical power.

References

1. Hirokawa S, Grimm M, Le T, Solomonow M, Baratta RV, Shoji H, et al. Energy consumption in paraplegic ambulation using the reciprocating gait orthosis and electrical stimulation of the thigh muscles. Arch Phys Med Rehabil 1990;71:687-94.
2. Isakov E, Douglas R, Berns P. Ambulation using the reciprocating gait orthosis and functional electrical stimulation. Paraplegia 1992;30:239-50.
3. Marsolais EB, Bennett G, Edwards CCT. Energy cost of walking and standing with functional neuromuscular stimulation and long leg braces. Arch Phys Med Rehabil 1988;69:243-9.
4. Nene AV, Patrick JH. Energy cost of paraplegic locomotion using the parawalker-electrical stimulation "hybrid orthosis." Arch Phys Med Rehabil 1990;71:116-20.
5. Winchester PK, Carollo JJ, Parekh RN, Lutz LM, Aston JW Jr. A comparison of paraplegic gait using two types of reciprocating gait orthoses. Prosthet Orthot Int 1993;17:101-6.
6. Astrand PO, Rodahl K. Textbook of work physiology. Physiological bases of exercise. 3rd ed. New York: McGraw Hill; 1986.
7. Wasserman K, Hansen JE, Sue DY, Whipp BJ. Principles of exercise testing and interpretation. 1st ed. Philadelphia: Lea & Febiger; 1987.
8. Huang C, Kuhlemeier KV, Moore NB, Fine PR. Energy cost of ambulation in paraplegic patients using Craig-Scott braces. Arch Phys Med Rehabil 1979;60:595-600.
9. Nene AV, Jennings SJ. Physiological cost index of paraplegic locomotion using the ORLAU parawalker. Paraplegia 1992;30: 246-52.
10. Engsberg JR, Herbert LM, Grimston SK, Fung TS, Harder JA. Relation among indices of effort and oxygen uptake in below-knee amputee and able-bodied children. Arch Phys Med Rehabil 1994;75:1335-41.
11. Stallard J, Rose GK, Tait JH, Davies JB. Assessment of orthoses by means of speed and heart rate. J Med Eng Technol 1978;2:22-4.
12. MacGregor J. The evaluation of patient performance using long-term ambulatory monitoring technique in the domiciliary environment. Physiotherapy 1981;67:30-3.
13. Stallard J, Major RE, Poiner R, Farmer IR, Jones N. Engineering design considerations of the ORLAU parawalker and FES hybrid system. Eng Med 1986;15:123-9.
14. Nene AV, Major RE. Dynamics of reciprocal gait of adult paraplegics using the parawalker (hip guidance orthosis). Prosthet Orthot Int 1987;11:124-7.
15. Yang L, Granat MH, Paul JP, Condie DN, Rowley I. Further development of hybrid functional electrical stimulation orthoses. Spinal Cord 1996;34:611-14.
16. Bar-on Z, Nene AV. Relationship between heart rate and oxygen uptake in thoracic level paraplegics. Paraplegia 1990;28:87-95.
17. IJzerman MJ, Baardman G, Hermens HJ, Veltink PH, Boom HBK, Zilvold G. The influence of the reciprocal cable linkage in the advanced reciprocating gait orthosis on paraplegic gait performance. Prosthet Orthot Int 1997;21:52-61.
18. Waters RL, Yakura JS, Adkins R, Barnes G. Determinants of gait performance following spinal cord injury. Arch Phys Med Rehabil 1989;70:811-8.
19. Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheumatol 1987;14 Suppl 15:6-10.
20. Katz DE, Haideri N, Song K, Wyrick P. Comparative study of conventional hip-knee-ankle-foot-orthoses versus reciprocating-gait orthoses for children with high-level paraparesis. J Pediatr Orthop 1997;17:377-86.
21. van Houwelingen JC, Stijnen T, van Strik R. Inleiding tot de medische statistiek. 1st ed. Utrecht: Wetenschappelijke uitgeverij Bunge; 1993.
22. Harvey LA, Davis GM, Smith MB, Engel S. Energy expenditure during gait using the walkabout and isocentric reciprocal gait orthosis in persons with paraplegia. Arch Phys Med Rehabil 1998;79:945-9.
23. Crosbie WJ, Nicol AC. Reciprocal aided gait in paraplegia. Paraplegia 1990;28:353-63.
24. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis 1985;38:27-36.
25. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol 1992;45:1341-5.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;8:307-10.
27. Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ 1996;313:41-2.
28. Streiner DL. Learning how to differ: agreement and reliability statistics in psychiatry. Can J Psychiatry 1995;40:60-6.

29. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health state measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12(4 Suppl):142S-158S.
30. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
31. Krebs DE. Intraclass correlation coefficients, use and calculation. Phys Ther 1984;64:1581-9.
32. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171-8.
33. Beckerman H. Efficacy of thermocoagulation and an ankle foot orthosis in stroke patients. A placebo controlled randomized clinical trial. [dissertation]. Amsterdam: Free University; 1996.
34. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of the generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther 1993;73:386-401.

35. Stratford PW. Reliability: consistency or differentiating among subjects? Phys Ther 1989;69:299-300.
36. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. 2nd ed. Belmont (CA): Duxbury Press; 1988.
37. Andrews BJ, Granat MH, Heller BW, MacMahon J, Keating L, Real S. Improved harness crutch to reduce upper limb effort in swing-through gait. Med Eng Phys 1994;16:15-8.

Suppliers