Nutrition in Clinical Practice
Volume 33 Number 2
April 2018 206–216

DOI: 10.1002/ncp.10070
wileyonlinelibrary.com
WILEY

Determining the Accuracy and Reliability of Indirect Calorimeters Utilizing the Methanol Combustion Technique

Sepideh Kaviani, BS[1]; Dale A. Schoeller, PhD[2]; Eric Ravussin, PhD[3]; Edward L. Melanson, PhD[4]; Sarah T. Henes, PhD, RD, LDN[5]; Lara R. Dugas, PhD, MPH[6]; Ronald E. Dechert, DrPH[7]; George Mitri, BS[7]; Paul F.M. Schoffelen, PhD[8]; Pim Gubbels, BS[9]; Asa Tornberg, PhD[10]; Stephen Garland, BS[10]; Marco Akkermans, BS[11]; and Jamie A. Cooper, PhD[1]

Abstract

Background: Several indirect calorimetry (IC) instruments are commercially available, but comparative validity and reliability data are lacking. Existing data are limited by inconsistencies in protocols, subject characteristics, or single-instrument validation comparisons. The aim of this study was to compare accuracy and reliability of metabolic carts using methanol combustion as the cross-laboratory criterion. *Methods:* Eight 20-minute methanol burn trials were completed on 12 metabolic carts. Respiratory exchange ratio (RER) and percent O_2 and CO_2 recovery were calculated. *Results:* For *accuracy*, 1 Omnical, Cosmed Quark CPET (Cosmed), and both Parvos (Parvo Medics trueOne 2400) measured all 3 variables within 2% of the true value; both DeltaTracs and the Vmax Encore System (Vmax) showed similar accuracy in measuring 1 or 2, but not all, variables. For *reliability*, 8 instruments were shown to be reliable, with the 2 Omnicals ranking best (coefficient of variation [CV] $<$ 1.26%). Both Cosmeds, Parvos, DeltaTracs, 1 Jaeger Oxycon Pro (Oxycon), Max-II Metabolic Systems (Max-II), and Vmax were reliable for at least 1 variable (CV $\le$ 3%). For *multiple regression*, humidity and amount of combusted methanol were significant predictors of RER ($R^2 = 0.33$, $P < .001$). Temperature and amount of burned methanol were significant predictors of O_2 recovery ($R^2 = 0.18$, $P < .001$); only humidity was a predictor for CO_2 recovery ($R^2 = 0.15$, $P < .001$). *Conclusions:* Omnical, Parvo, Cosmed, and DeltaTrac had greater accuracy and reliability. The small number of instruments tested and expected differences in gas calibration variability limits the generalizability of conclusions. Finally, humidity and temperature could be modified in the laboratory to optimize IC conditions. (*Nutr Clin Pract.* 2018;33:206–216)

Keywords

metabolic cart; reliability; accuracy; indirect calorimetry; methanol; energy metabolism

Oxygen consumption (VO_2) and carbon dioxide production (VCO_2) measurements obtained from indirect calorimetry (IC) in humans are used to measure energy expenditure (EE)[1-3] and calculate substrate utilization using the respiratory exchange ratio ($RER = VCO_2/VO_2$)[2,4] to calculate macronutrient oxidation.[5] Metabolic carts are mobile and have a small footprint so they are commonly used in research and clinical settings.[2] To measure fasting, exercise, and postprandial EE and substrate oxidation, it is crucial to have an instrument that is both reliable and accurate. Reliability refers to the extent to which an instrument is able to produce similar results on repeated measurements, whereas accuracy

---

From the [1]Department of Foods and Nutrition, University of Georgia, Athens, Georgia, USA; [2]Department of Nutritional Sciences, University of Wisconsin, Madison, Wisconsin, USA; [3]Pennington Biomedical Research Center, Baton Rouge, Louisiana, USA; [4]Division of Endocrinology, Metabolism, & Diabetes, University of Colorado Anschutz Medical Campus, Denver, Colorado, USA; [5]Department of Nutrition, Georgia State University, Atlanta, Georgia, USA; [6]Public Health Sciences, Loyola University, Chicago, Illinois, USA; [7]Pediatric Respiratory Care, University of Michigan Health System, Ann Arbor, Michigan, USA; [8]Department of Human Biology & Movement Sciences, NUTRIM School for Nutrition, Toxicology & Metabolism, Maastricht University Medical Centre, Maastricht, the Netherlands; [9]Topsport Expertise & Innovation Centre, Sittard, the Netherlands; [10]Department of Health Sciences, Lund University, Lund, Sweden; and [11]Center of Expertise for Chronic Organ Failure, Horn, the Netherlands.

Financial disclosure: This research was supported by grants from Colorado Nutrition and Obesity Research Center (P30 DK048520) and Colorado Clinical and Translational Science Institute (UL1 RR025780).

Conflicts of interest: None.

Corresponding Author:
Jamie A. Cooper, PhD, Department of Foods and Nutrition, University of Georgia, 305 Sanford Drive, Athens, GA 30602, USA.
Email: jamie.cooper@uga.edu

explains how close a measurement is to the actual or true value. A few previous studies have examined the accuracy and reliability of different metabolic carts by comparing resting metabolic rate (RMR) and RER measures using multiple metabolic carts with repeated tests performed on human participants.[4,6-8] This method of assessment considers 1 instrument as the most accurate, that is, the criterion. Then the accuracy of other instruments is tested based on how close to this criterion they can measure metabolic data. If the criterion itself is validated with methanol burns, it is considered accurate; however, this is not always done. By using human participants from varying populations, past studies have revealed that indirect calorimeters show a certain degree of inaccuracy in measuring RER. Further, they can have a large within-subject variability for resting energy expenditure (REE), even with interunit variability testing.[4,6-8] Moreover, these studies are often done first thing in the morning in fasting human subjects with no multiple measures within a day to investigate across-day variation.[4,9] Because of inconsistencies in previous methodological assessments and the absence of data on accuracy for several types of metabolic carts, research is needed to eliminate potential causes of subject-related variance by using a common chemical burn as the benchmark.

Alcohol combustion with methanol or ethanol is routinely used to validate the accuracy of metabolic carts, as well as correct data after they have been collected.[4,8,10-12] Methanol combustion has a well-defined theoretical value of O_2 and CO_2. It can therefore be used to determine accuracy and reliability of a metabolic cart while eliminating some of the aforementioned confounding variables. The purpose of this study was to determine the accuracy or reliability of several different metabolic carts using the methanol combustion technique. Because the DeltaTrac II (DTC) has previously been considered a gold standard instrument, we hypothesized that it would be the most accurate and reliable instrument.[4,13,14] We also hypothesized that there would be a high degree of variability indicating low reliability across 2 testing days.

Materials and Methods

Instruments and Collaborators

Researchers owning metabolic carts with methanol burn capabilities around the United States and Europe were contacted to collaborate. Two sites for each instrument were identified for all instruments except the Vmax Encore System (CareFusion, Yorba Linda, CA, USA) and Max-II Metabolic Systems (Max-II; AEI Technologies Pittsburgh, PA, USA) metabolic carts. Table 1 shows all instruments and study sites, along with an abbreviated name for each instrument that will be used throughout the manuscript. To perform the methanol burns, all of the instruments use a glass alcohol container with a wick, inside a ventilated glass canopy to burn methanol, except for the Max-II, the DTC, and the Cosmed Quark CPET (Cosmed), which use a crucible in the ventilated glass canopy. Table 2 shows more details about each instrument's characteristics.

Study Protocol

A 500-mL unopened bottle of methanol (A412500 Certified American Chemical Society, 0.1% maximum water; Fisher Scientific, Asheville, NC, USA) was sent by the lead researchers to each site. All testing was done with methanol from the same lot number. Before the 2-day testing protocol, each researcher with the glass canopy wick burning kit performed multiple methanol burns to achieve the wick height that equaled an EE of 1.0 kcal/min, which is a similar rate of human adult EE per minute at rest.[15-17] It was not possible to manipulate the burn rate with the crucible.

Testing was completed on 2 consecutive days, and methanol burns were performed at 7:00 AM, 10:00 AM, 1:00 PM, and 4:00 PM local time. Data were used to determine accuracy and reliability across and between testing days. Each instrument was calibrated before testing according to manufacturer's specifications. A fan was also kept running in the room to minimize variability in ambient air secondary to human activity. The parameters including room temperature, humidity, barometric pressure, and the amount of methanol burned were recorded for each burn. The flame was lit as soon as the program was started on the computer. The first minute of data collection was discarded because the combusted methanol had not yet filled the glass canopy of the burn kit and values were similar to room gas. At the 20th minute of methanol burning, the glass canopy of the burn kit was lifted, the flame was blown out, and the canopy was replaced as quickly as possible for the final minute of data collection. The methanol weight was not dynamically measured and recorded during the burn. Rather, methanol weight was recorded immediately before and immediately upon completion of the burn using a routinely calibrated gram scale. The methanol burn test for each site was performed by the same researcher using the aforementioned procedures. All data collected during the 20-minute burn were used for final data analysis.

Calculations

Percent recoveries of both O_2 and CO_2, as well as RER, were calculated by previously published formulas.[18] The theoretical value for percent O_2 and CO_2 recovery was 100%, and the theoretical value for RER was 0.667 based on the ratio of produced CO_2 to utilized O_2 in the burning of methanol from this equation: $2CH_3OH + 3O_2 \rightarrow 2CO_2 + 4H_2O$, which is the value regardless of fuel infusion rate and/or changes in fraction of inspired oxygen (FiO_2).[19]

Table 1. Sites and Characteristics of the 12 Instruments Tested at 11 International Study Sites.

Instrument	Center	Location	Age of Instrument (y)	Geographic Elevation (ft)	Burn Environment: Average Humidity (%)[a]	Burn Environment: Average Temperature (°C)[a]
Vmax Encore System (Vmax)[b]	CS Mott Children's Hospital	Ann Arbor, Michigan, USA	4	629 (191.7 m)	24.8	23.7
Cosmed Quark CPET (Cosmed1)	Georgia State University	Atlanta, Georgia, USA	4.5	1102 (335.9 m)	43.9	23
Cosmed Quark CPET (Cosmed2)	Loyola University of Chicago	Maywood, Illinois, USA	6	636 (193.9 m)	38.1	24.4
Max-II Metabolic Systems (Max-II)	Pennington Biomedical Research Center	Baton Rouge, Louisiana, USA	7	65 (19.8 m)	45	23
DeltaTrac II Metabolic Monitor (DTC1)			≥15		45.5	21
Parvo Medics trueOne 2400 (Parvo1)	University of Colorado-Denver	Denver, Colorado, USA	1	5367 (1635.9 m)	44.9	25.0
Parvo Medics trueOne 2400 (Parvo2)	University of Georgia	Athens, Georgia, USA	1	761 (232.0 m)	53	24.5
DeltaTrac II Metabolic Monitor (DTC2)	University of Wisconsin-Madison	Madison, Wisconsin, USA	16	892 (271.9 m)	26	22.8
Jaeger Oxycon Pro (Oxycon1)	Lund University	Lund, Sweden	4	167 (50.9 m)	35.9	22.6
Jaeger Oxycon Pro (Oxycon2)	CIRO–Center of Expertise for Chronic Organ Failure	Horn, the Netherlands	15	101 (30.8 m)	62.5	21.5
Omnical (Omnical1)	Topsport Expertise and Innovative Centre (TEIC)	Sittard, the Netherlands	2	153 (46.6 m)	64.1	17.1
Omnical (Omnical2)	Maastricht University Medical Center	Maastricht, the Netherlands	6	183 (55.8 m)	57.1	19.7

[a] The average humidity and temperature for each test site defines the average of humidity (%) and temperature (°C) during methanol burns done in all 8 time points of the 2 test days inside of the laboratories where testing occurred.

[b] For the purpose of easy reference to each of the metabolic carts in this article, each have been given a short name with or without a number in parentheses.

Statistical Analysis

Data were analyzed using SAS Version 9.4 (SAS Institute, Cary, NC, USA). No statistical analyses were performed on a between-site basis. To test accuracy, we compared average values of each of the 3 variables (RER, percent (%) recovery O_2 and % recovery CO_2) throughout the 8 time points obtained from each instrument with the earlier mentioned theoretical values for methanol. The percent of relative error (%RE) was calculated as $\left(\frac{\text{Average of measured values} - \text{theoretical value}}{\text{Theoretical value}}\right) \times 100$. The instruments were ranked from lowest to highest based on %RE. The use of %RE to show accuracy helps assess the distance of the average values measured by the instruments from the theoretical value, regardless of variability throughout multiple measurements. Eliminating the variability from assessments of accuracy allows the proper separation of accuracy and reliability given the previously mentioned differences in their definitions. Because there is no agreement on the constant value to be used in calculating RMR

Table 2. Instrument Characteristics.

Instrument	Calibration Gas Concentration: O_2	Calibration Gas Concentration: CO_2	Average Flow Rate (L/min)	Gas Analyzer: O_2	Gas Analyzer: CO_2	Flow Rate Calibration Method	Flow Rate Analyzer System
Vmax	16.0%	4.0%	46.7	Electrochemical fuel cell	Nondispersive infrared, thermopile	2-L syringe push and pull motion 10 times	Mass flow sensor
Cosmed1	16.02%	4.98%	30.97	Paramagnetic	Digital infrared	3-L syringe push and pull motion 10 times	ID18 turbine flowmeter; separate canopy unit
Cosmed2	16%	5%	28.02	Paramagnetic	Digital infrared	2-L syringe push and pull motion 10 times	Turbine 2000
Max-II	19.02%	0.028%	25.4	Paramagnetic	Infrared	3-L syringe with nonrebreathing valve, push and pull motion 5 times	Pneumotach (pressure transducer)
DTC1	Balance O_2	5%	42.10	Paramagnetic	Infrared	Flow rate assumed to be constant throughout the test using CO_2 recovery from ethanol	Not applicable
Parvo1	16%	4.05%	13.6	Paramagnetic	Digital infrared	3-L syringe push and pull motion 10 times	Rudolph heated Pneumotach
Parvo2	16%	4.05%	13.6	Paramagnetic	Digital infrared	3-L syringe push and pull motion 10 times	Rudolph heated Pneumotach
DTC2	96%	4%	36.2	Paramagnetic	Infrared	Methanol burns done weekly and averaged over 3 previous weeks	36.2 L/m STP; fixed flow rate using critical orifice design
Oxycon1	15.96%	4.95%	40	Paramagnetic	Infrared	Internal, automatic	Triple-V (flat fan)
Oxycon2	4.5 (16,000 vol%)	4.5 (5000 vol%)	38.88	Chemical fuel cell	Infrared absorption	Automated flow calibration at rates of 0.2 and 2.0 L/s	Triple-V (flat fan)
Omnical1	18%	0.8%	202.8	ABB H&B MAGNOS[a] dumbbell type paramagnetic	ABB H&B URAS[a] infrared	Periodical calibration with certified flowmeter in series	Unidirectional dry bellows flowmeter with digital counter
Omnical2	18%	0.8%	36.4	ABB H&B MAGNOS[a] dumbbell type paramagnetic	ABB H&B URAS[a] infrared	Periodical calibration with certified flowmeter in series	Unidirectional dry bellows flowmeter with digital counter

ID18, Inside Diameter 18 mm; STP, standard temperature and pressure; vol, volume.

[a] ABB H&B MAGNOS and ABB H&B URAS are brands/models of the analyzers.

Table 3. Accuracy Results of the 12 Instruments Tested Based on the Calculated Percent of Relative Error in Measuring the 3 Variables of Interest (Respiratory Exchange Ratio, O_2 Recovery, and CO_2 Recovery).

	O_2 Recovery (%)				**CO_2 Recovery (%)**				**Respiratory Exchange Ratio**		
Instrument	$\bar{X} \pm$ SD	%RE	Rank[a]	Instrument	$\bar{X} \pm$ SD	%RE	Rank[a]	Instrument	$\bar{X} \pm$ SD	%RE[a]	Rank
Omnical1	100.1 ± 0.8	0.094[b]	1	Parvo2	100.1 ± 3.2	0.144[b]	1	Omnical2	0.667 ± 0.002	0.001[b]	1
Parvo1	0.99 ± 0.02	−0.653[b]	2	Parvo1	0.99 ± 0.02	−0.525[b]	2	Parvo1	0.668 ± 0.006	0.126[b]	2
Cosmed2	99.8 ± 2.1	−0.250[b]	3	Cosmed2	100.7 ± 1.7	0.666[b]	3	Cosmed2	0.673 ± 0.005	0.927[b]	3
DTC2	100.8 ± 2.4	0.794[b]	4	DTC1	100.7 ± 1.5	0.702[b]	4	Parvo2	0.660 ± 0.005	−1.034[b,c]	4
Parvo2	101.2 ± 2.9	1.185[b]	5	Vmax	98.4 ± 2.1	−1.615[b]	5	Omnical1	0.677 ± 0.003	1.546[b]	5
Vmax	101.8 ± 5.1	1.804[b]	6	Omnical1	101.6 ± 0.8	1.640[b]	6	Max-II	0.652 ± 0.022	−2.226	6
Oxycon2	97.9 ± 7.2	−2.051	7	DTC2	98.1 ± 3.8	−1.874[b]	7	DTC2	0.649 ± 0.020	−2.649	7
Cosmed1	102.5 ± 2.2	2.469	8	Omnical2	103.6 ± 1.3	3.647	8	Oxycon1	0.649 ± 0.007	−2.705	8
DTC1	97.4 ± 1.7	−2.619	9	Cosmed1	95.7 ± 1.2	−4.293	9	Vmax	0.645 ± 0.023	−3.221	9
Omnical2	103.6 ± 1.3	3.647	10	Oxycon2	104.4 ± 6.7	4.368	10	DTC1	0.689 ± 0.008	3.420	10
Max-II	107.0 ± 3.9	6.970	11	Max-II	104.5 ± 2.2	4.493	11	Cosmed1	0.626 ± 0.014	−6.089	11
Oxycon1	91.4 ± 3.1	−8.615	12	Oxycon1	88.9 ± 2.6	−11.103	12	Oxycon2	0.712 ± 0.040	6.755	12

%RE, percent of relative error; Vmax, Vmax Encore System; $\bar{X}$, average of measured values by the instruments throughout the 8 burns.

[a] Instruments rankings (best to worst) are based on %RE ($\%RE = \frac{\text{(Average of measured values − theoretical value)}}{\text{Theoretical value}} \times 100$), which explains the difference of the average values measured by the instruments from theoretical. Theoretical values are 0.667 (for RER) and 100% (for O_2 and CO_2 recoveries).

[b] Denotes accurate instrument based on a %RE ≤2%.

[c] Negative %RE values on this table indicate that the calculated averages for the corresponding instruments were *less* than the theoretical value.

and substrate oxidation,[20] we elected to use 2% difference from the theoretical as the accuracy threshold for measured RER (RER range of 0.653–0.680) and gas recovery (recovery range of 98%–102%) because such an estimated, but defendable, approach implies that smaller differences are probably not biologically important. To obtain the cutoff point of 2% for RE, we pursued an empirical approach by calculating the median SD across instruments for each of the 3 variables and then a 95% confidence interval (CI) for standard deviation (SD). Further, a 95% CI for %RE was calculated for each variable. Those values were 0.94%, 2.21%, and 1.79% for RER, % recovery O_2, and % recovery CO_2, respectively. The smaller 95% CI for RER compared with O_2 and CO_2 suggests a flow rate calibration error, although incomplete methanol combustion is another possible explanation.

To measure reliability, we calculated the coefficient of variation (CV) for each laboratory–instrument pair to assess the amount of variability relative to the average values of RER, % recovery O_2, and % recovery CO_2. CV <3% was selected as the upper limit threshold of reliability.[4] Because a 4% variability from 1 day to another is reported for human subjects,[21] it is suggested that an ideal gas analysis system needs to have a CV <3% to not grossly inflate the day-to-day variation in RMR from a given individual.[4]

Finally, a stepwise multiple regression was used to assess which environmental or exogenous factors could explain the majority of the variance in our outcome measures. Factors used in the model included humidity, temperature, and the amount of methanol used in each burn. We also performed single regression to see how geographic elevation of the testing site and age of the instruments were correlated with outcome measures. The latter 2 variables were not included in the multiple regression model because they remained constant throughout the 8 time points of instrument testing, whereas the other factors showed variation from 1 burn to the next.

Results

Accuracy (% Recovery O_2 and CO_2)

Accuracy data for % recoveries can be found in Table 3 and Figures 1 and 2. Six out of the 12 instruments measured % recovery O_2 within the ±2% difference (%RE ≤ 2) from the true value (Omnical1 [O_2 recovery = 100.1% ± 0.8%; % RE = 0.1], Parvo Medics trueOne 2400 [Parvo1; O_2 recovery = 0.99 ± 0.02; %RE = 0.65], Cosmed2 [O_2 recovery = 99.8% ± 2.1%; %RE = −0.3], DTC2 [O_2 recovery = 100.8% ± 2.4%; %RE = 0.8], Parvo Medics trueOne 2400 [Parvo2; O_2 recovery = 101.2% ± 2.9%; %RE = 1.2], and Vmax [O_2 recovery = 101.8% ± 5.1%; %RE = 1.8]). Seven out of 12 instruments measured % recovery CO_2 within the ±2% difference (%RE ≤ 2) from the true value (Parvo2 [CO_2 recovery = 100.1% ± 3.2%; %RE = 0.1], Parvo1 [CO_2 recovery = 0.99 ± 0.02; %RE = −0.53], Cosmed2 [CO_2 recovery = 100.7% ± 1.7%; %RE = 0.7], DTC1 [CO_2 recovery = 100.7% ± 1.5%; %RE = 0.7], Vmax [CO_2 recovery = 98.4% ± 2.1%, %RE = −1.6], Omnical1 [CO_2 recovery = 101.6% ± 0.8%, %RE = 1.6], and DTC2 (CO_2 recovery = 98.1% ± 3.8%, %RE = −1.9]). Therefore, the Omnical1, Cosmed2, DTC2, Parvo2, and Vmax were

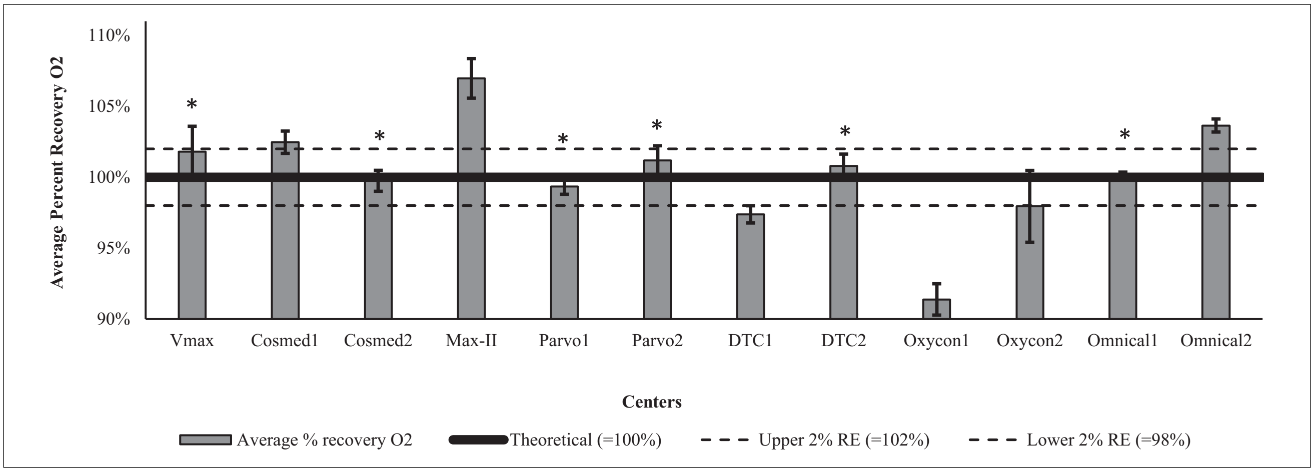

Figure 1. Six out of 12 instruments (Omnical1, Cosmed2, DTC2, Parvo1, Parvo2, and Vmax) measured % recovery O_2 with %RE ≤2% (inside the limits of upper and lower dashed lines). Asterisks denote accurate instrument based on a %RE ≤2%.

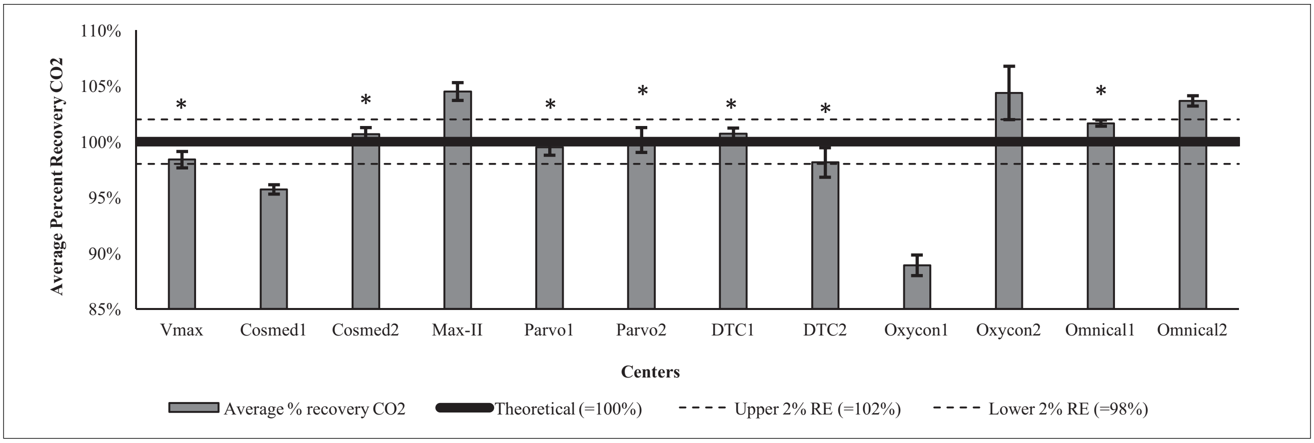

Figure 2. Seven out of 12 instruments (Parvo2, Parvo 1, Cosmed2, DTC1, Vmax, Omnical1, and DTC2) measured % recovery CO_2 with %RE ≤2% (inside the limits of upper and lower dashed lines). Asterisks denote accurate instrument based on a %RE ≤2%.

accurate for both O_2 and CO_2 recoveries, whereas DTC1 was accurate for only % recovery CO_2 and not for O_2.

Accuracy (RER)

The average RER across the 2 study days was compared with the theoretical value for methanol (0.667) to determine accuracy (Table 3 and Figure 3). Among all 12 instruments, only 5 instruments measured RER within the ±2% difference (%RE ≤ 2) from the true value (Omnical2 [RER = 0.667 ± 0.002; %RE = 0.0], Parvo1 [RER = 0.668 ± 0.006, %RE = 0.126], Cosmed2 [RER = 0.673 ± 0.005; %RE = 0.9], and Parvo2 [RER = 0.660 ± 0.005; %RE = −1.0], Omnical1 [RER = 0.677 ± 0.003; %RE = 1.5]). The other 7 instruments had %RE >2 for RER.

Finally, all of the instruments were rank ordered from closest to theoretical to the furthest from theoretical according to the %RE. These results can be found in Table 3. For % recovery O_2, Omnical1, Parvo1, and Cosmed2 had the least %RE, respectively; for % recovery CO_2, Parvo2, Parvo1, and Cosmed2 had the least %RE, respectively, indicating these to be the top/better metabolic carts tested for accuracy in this study. For RER, Omnical2, Parvo1, and Cosmed2 had the least %RE, respectively.

Reliability

For % recovery O_2, 8 out of 12 instruments (both Omnicals, Cosmeds, DeltaTracs, and Parvos) were shown to be reliable with Omnical1 (CV = 0.75%), Omnical2 (CV = 1.26%), and Parvo1 (CV = 1.56%) being the 3 most reliable

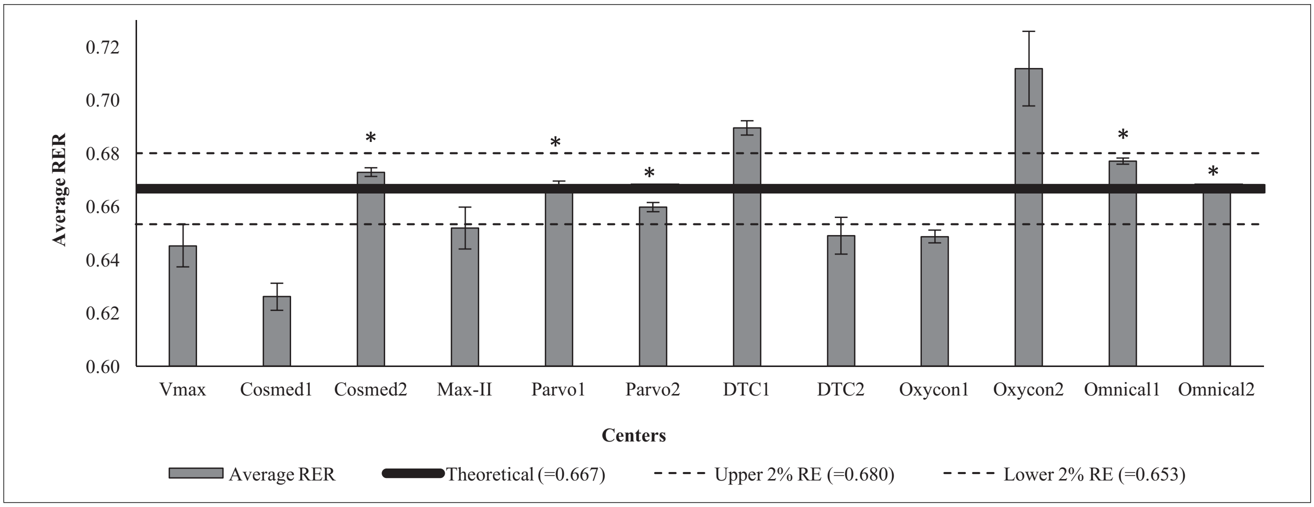

Figure 3. Five out of 12 instruments (Omnical2, Cosmed2, Parvo2, Omnical1, and Parvo1) measured RER with %RE ≤2% (inside the limits of upper and lower dashed lines). Asterisks denote accurate instrument based on a %RE ≤2%.

Table 4. Reliability Results of the 12 Instruments Tested Based on the Calculated Coefficient of Variation in Measuring the 3 Variables of Interest (Respiratory Exchange Ratio, O_2 Recovery, and CO_2 Recovery).

	% Recovery O_2			**% Recovery CO_2**			**Respiratory Exchange Ratio**	
Instrument	CV (%)	Rank[a]	Instrument	CV (%)	Rank[a]	Instrument	CV (%)	Rank[a]
Omnical1	0.75[b]	1	Omnical1	0.7[b]	1	Omnical2	0.23[b]	1
Omnical2	1.26[b]	2	Omnical2	1.21[b]	2	Omnical1	0.49[b]	2
Parvo1	1.56[b]	3	Cosmed1	1.22[b]	3	Cosmed2	0.68[b]	3
DTC1	1.76[b]	4	DTC1	1.47[b]	4	Parvo2	0.75[b]	4
Cosmed2	2.10[b]	5	Cosmed2	1.65[b]	5	Parvo1	0.86[b]	5
Cosmed1	2.18[b]	6	Parvo1	1.93[b]	6	Oxycon1	1.05[b]	6
DTC2	2.35[b]	7	Vmax	2.09[b]	7	DTC1	1.13[b]	7
Parvo2	2.89[b]	8	Max-II	2.14[b]	8	Cosmed1	2.28[b]	8
Oxycon1	3.42	9	Oxycon1	2.97[b]	9	DTC2	3.01	9
Max-II	3.69	10	Parvo2	3.18	10	Max-II	3.42	10
Vmax	4.97	11	DTC2	3.84	11	Vmax	3.50	11
Oxycon2	7.31	12	Oxycon2	6.46	12	Oxycon2	5.49	12

CV, coefficient of variation.

[a] Instruments rankings (best to worst) are based on CV, which was used to assess reliability of measurements by each instrument.

[b] Denotes reliable instrument based on a CV ≤3%.

carts. For % recovery CO_2, 9 out of 12 instruments (both Omnicals, Cosmeds, the DTC1, Parvo1, Vmax, Max-II, and Jaeger Oxycon Pro [Oxycon1]) were reliable with Omnical1 (CV = 0.76%), Omnical2 (CV = 1.21%), and Cosmed1 (CV = 1.22%) being the 3 most reliable metabolic carts with the smallest CVs. For RER, 8 out of 12 instruments (both Omnicals, Cosmeds, Parvos, the DTC1, and Oxycon1) were considered reliable (CV ≤ 3%), with Omnical2 (CV = 0.23%), Omnical1 (CV = 0.49%), and Cosmed2 (CV = 0.68%) being the 3 most reliable carts, respectively (Table 4). The CV calculation was based on all 8 tests for each instrument. Range of SD for RER was 0.002–0.040, for %O_2 was 0.008–0.139, and for %CO_2 was 0.008–0.144.

Correlation and Multiple Regression

The correlation matrix for outcome variables and exogenous factors (temperature, humidity, amount of methanol burned, age of instrument, and elevation) are shown in Table 5. Both RER and percent CO_2 recovery were negatively correlated with temperature and amount of methanol burned, and positively correlated with humidity. Percent O_2 recovery did not correlate with any exogenous factors. Geographic elevation of instrument was not correlated with any of the outcome variables. Age of instrument was significantly correlated with RER (P = .01), but not with gas recoveries. For the stepwise multiple regression,

Table 5. Correlation Matrix.

Correlation Coefficients	% O_2 Recovery	% CO_2 Recovery	Temperature	Humidity	Amount of Methanol Burned	Age	Geographic Elevation
RER							
R^2	−0.37	0.46	−0.29	0.37	−0.23	0.26	−0.01
P	<.001[a]	<.001[a]	.005[a]	<.001[a]	.02[a]	.01[a]	.94
% O_2 recovery							
R^2		0.66	0.02	0.09	−0.07	−0.07	−0.05
P		<.001[a]	.86	.36	.50	.49	.64
% CO_2 recovery							
R^2			−0.22	0.39	−0.27	0.14	−0.05
P			.03[a]	<.001[a]	<.01[a]	.19	.63
Temperature							
R^2				−0.65	0.50	−0.12	0.33
P				<.001[a]	<.001[a]	.23	.001[a]
Humidity							
R^2					−0.34	−0.03	−0.45
P					<.001[a]	.75	<.001[a]
Amount of methanol burned							
R^2						0.15	0.13
P						.15	.22
Age							
R^2							−0.36
P							<.001[a]
Geographic elevation							
R^2							
P							

RER, respiratory exchange ratio.
[a] Significant correlation based on a P value ≤.05.

humidity, temperature, and grams of methanol burned were included in the model with RER as dependent variable. Temperature could be removed from the model with backward elimination technique, leaving both humidity and amount of burned methanol as independent and significant positive predictors of RER, making the overall regression model significant [F (2, 92) = 10.91, $P < .001$, $R^2 = 0.3301$]. For % recovery O_2, humidity was removed from the model with backward elimination technique, leaving temperature and amount of burned methanol as significant predictors of % recovery O_2 (positive and negative predictors, respectively), making the overall regression model significant [F (2, 92) = 8.32, $P < .001$, $R^2 = 0.1822$]. Finally, for % recovery CO_2, both the amount of burned methanol and temperature were removed from the model, leaving only humidity as a significant remaining predictor [F (1, 92) = 21.10, $P < .001$, $R^2 = 0.1498$].

Discussion

In this study, 1 Cosmed, 1 Omnical, and 2 Parvos measured all 3 variables (RER and gas recoveries) within ±2% difference from the theoretical, whereas both DeltaTracs and the 1 Vmax showed similar accuracy in measuring 1 or 2, but not all, variables. The Omnicals, 1 Parvo, and 1 Cosmed were the highest ranked instruments for having the smallest %RE in measuring 1 or 2 of the variables. Because only 1 or 2 instruments at different study sites were used in this project, caution should be taken when attempting to make generalizations to all metabolic carts of a particular make and model. Further, because 2 instruments of the same model did not always reveal similar extents of accuracy, which could have been caused from environmental differences between the 2 sites, researchers need to periodically perform combustion burns as part of research purposes and calibration.

Reliability analysis demonstrated that both of the 2 Omnicals, Cosmeds, and 1 DTC were reliable for all 3 variables. Both Parvos, 1 Oxycon, and 1 DTC were reliable in measuring 1 or 2, but not all, variables. Further, Omnicals showed the least variability, and hence the most reliability, in measuring all variables. Note that these results are in part dependent on the calibration gas applied, which may explain a small portion of the error depending on absolute range within the gas certification.

Previous studies have assessed the validity or reliability of metabolic carts through widely different methods with some conflicting results.[4,18,22-30] Nearly all of those studies were done in human subjects with 1 instrument (DTC in some cases) chosen as the criterion.[4,18] Cooper et al[4] determined validity and reliability of 5 different IC systems (MedGraphics CPX Ultima, MedGem, Vmax Encore 29 System, TrueOne 2400, and Korr ReeVue) against DTC in human subjects. They found that the Parvo (TrueOne) and Vmax were the most valid instruments in measuring RMR and RER, whereas none of the instruments showed strong reliability. Conversely, other human studies on the Cosmed vs either the DTC[23] or Douglas bag system,[22] as well as studies on validity of Oxycon Pro against the Douglas bag method,[24,26-30] concluded that both instruments were equally as accurate and reliable. The fact that DTC was not the superior instrument in our experiment sheds light on the limitation of using a metabolic cart as the gold standard instrument when testing the validity of several different IC instruments because one cannot demonstrate superiority, just equivalence.

One question faced by researchers is whether RER or % recoveries are more important? We contend that the % recoveries for O_2 and CO_2, required for determining EE, are more important. However, both must be as reproducible as possible for supporting intervention studies. RER can be deemed less important because of the nature of the RER calculation being a ratio. If both O_2 and CO_2 are either overrecovered or underrecovered, this would not affect the ratio of these 2 gases (RER). In addition, the RER may be affected by calibration gas accuracy, specifically if O_2 and CO_2 have opposite deviation of certificate, though within certificate accuracy. Ultimately, O_2 has the largest impact on EE (4 times that of CO_2). Therefore, it may be important to evaluate the accuracy of IC systems primarily on gas recovery of O_2, then CO_2, then RER.

A second question facing researchers is how to determine whether IC outcomes will be able to detect the possibly small impact of an intervention where the effect size depends on the %CV of the IC. In this study, we used 2% as the cutoff for determining accuracy. To translate this into a biological context, we calculated what a 2% margin of error means for an example of RMR, RER, and diet-induced thermogenesis (DIT) in humans. If RMR is 1500 kcal/d, a 2% RE around this value equals the range of 1470–1530 kcal. Although seemingly small, this degree of an error (range of 30–60 kcal) could correspond to a considerably inaccurate estimation of an individual's total daily energy expenditure (TDEE) because RMR is then multiplied by an activity factor to estimate TDEE. Further, very small energy surpluses (as little as 8–9 kcal/d), if not compensated over time, have been shown to result in the average annual weight gain of 1 kg.[31] It is possible that some of the instruments with a 3% or 4% error in this study would also be acceptable for RMR measurements. Conversely, greater than a 2% RE (0.013 below or above the theoretical value which equals to the total of 0.026 difference around the theoretical value) for RER could be important because studies have reported significant treatment effects on changes of as low as 0.02 or 0.03.[32-34] A difference in RER of 0.02 can be detected in the accurate instruments; however, those with a %RE worse than 2% could easily mask that small biological difference. A similar case can be made for DIT because significant, but small, differences are often reported.[35]

Based on some of the variability in instrument performance at different sites, we examined possible environmental underlying factors that could influence accuracy or reliability. The results showing humidity and temperature as predictors for at least 1 outcome measure were somewhat surprising because the instruments are designed to account/correct for temperature and humidity. Therefore, the lack of accuracy and/or reliability for some of these instruments may lie partly in the inability to correctly account for humidity and temperature, and improvements on adjusting for these factors is warranted. We do want to note, however, that the magnitude of variance explained by these exogenous factors was fairly small (15%–33% depending on outcome variable), so these data should be interpreted with a small degree of caution. Because these small effects on instrument performance can add up to become biologically important, researchers should also consider regulating room humidity and temperature to improve instrument performance. In this study, we had a fairly large range of humidity and temperature during testing days across instruments (24.1%–68.2% for humidity; 16.9°C–26°C for temperature), which explained a small percentage of the variation, so if researchers can keep these variables somewhat consistent from day to day, they are less likely to negatively impact accuracy and reliability.

Study Limitations

We only tested 2 instruments of the same model, some of which varied drastically in their performance between 2 study sites, which indicates that >2 instruments per model might achieve more conclusive findings. We also used burn kits instead of the actual chambers/hoods that are used for human subjects. An attempt to have an open flame inside the plastic canopy hood instead of the glass hood, which is used for the methanol burns, resulted in melting of the plastic canopy or extremely poor gas recoveries, possibly because of protective layers and ice being used on the plastic hood. However, researchers that use methanol burn data from the glass hoods observe complete combustion of methanol and measure true combusted amounts by weight, thus the accuracy and reliability of that system was still important. We are operating under the assumption that 100% combustion occurred, which is the assumption routinely used for researchers who regularly perform alcohol burns for calibration purposes. Further, the alcohol combustion kit is often the only cost-effective option available, and is proposed as validation by manufacturers of breath-by-breath IC units. Another limitation could be the burn rate. We attempted to have all of the collaborators burn methanol at a certain rate to equal about 1.0 kcal/min, which is similar to human metabolic IC data; however, not every site was able to achieve this. Although the theoretical gas values from methanol combustion should be independent of the fuel infusion rate, the amount of methanol was shown to have a small impact on the % recovery O_2. This is a typical effect in IC as noted for whole room calorimetry by Murgatroyd and colleagues.[36]

Conclusions

Within the limits of the small number of instruments we tested, the Omnical, Parvo, Cosmed, and DeltaTrac were determined to be the most accurate and reliable instruments; however, accuracy was shown at only 1 of the 2 study sites, and not for all variables. Omnical and Cosmed showed reliability at both study sites for all 3 outcome measures. Exogenous factors such as humidity and temperature may influence instrument performance despite that IC systems are designed to correct for such environmental elements. Although this may have a relatively small effect on instrument performance, those small effects can add up and possibly become biologically important. Therefore, in addition to the typical requirement for stable background gas fractions, humidity and temperature could be modified in the laboratory to improve IC conditions. The fact that the comparisons made in this study are not generalizable to all the manufactured instruments of the earlier discussed models highlights the importance of cautious interpretations of the findings. Importantly, though, the findings do allow researchers to evaluate their own results in comparison with what was found here. The importance of performing periodic methanol burns in every laboratory as part of the instrument testing of accuracy and precision is a necessary step towards maximizing instrument performance and providing valid results in research. This relatively small sample provided by researchers in the field is only a first step in defining and clarifying validity and its problems in IC equipment. Future research will be needed using a similar or preferably larger scale approach of validating IC instruments, to further assess validity and reliability both in technical equipment and for measuring human participants.

Statement of Authorship

Sepideh Kaviani, Jamie A. Cooper, Dale A. Schoeller and Eric Ravussin contributed to the conception/design of the research; Sepideh Kaviani, Jamie Cooper, Dale A. Schoeller, Eric Ravussin, Edward L. Melanson, Sarah T. Henes, Lara R. Dugas, Ronald E. Dechert, George Mitri, Paul F.M. Schoffelen, Pim Gubbels, Asa Tornberg, Stephen Garland and Marco Akkermans contributed to the acquisition and analysis of the data; Sepideh Kaviani, Jamie Cooper, Dale A. Schoeller, Eric Ravussin and Sarah T. Henes contributed to the interpretation of the data; and Sepideh Kaviani and Jamie A. Cooper drafted the manuscript; All authors critically revised the manuscript, agree to be fully accountable for ensuring the integrity and accuracy of the work, and read and approved the final manuscript.

References

1. Weir JdV. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol*. 1949;109(1-2):1-9.
2. Matarese LE. Indirect calorimetry: technical aspects. *J Am Diet Assoc*. 1997;97(10):S154-S160.
3. Lam Y, Ravussin E. Indirect calorimetry: an indispensable tool to understand and predict obesity. *Eur J Clin Nutr*. 2017;71(3):318-322.
4. Cooper JA, Watras AC, O'Brien MJ, et al. Assessing validity and reliability of resting metabolic rate in six gas analysis systems. *J Am Diet Assoc*. 2009;109(1):128-132.
5. Frayn K. Calculation of substrate oxidation rates in vivo from gaseous exchange. *J Appl Physiol*. 1983;55(2):628-634.
6. Crouter SE, Antczak A, Hudak JR, DellaValle DM, Haas JD. Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. *Eur J Appl Physiol*. 2006;98(2):139-151.
7. Macfarlane D, Wu H. Inter-unit variability in two ParvoMedics TrueOne 2400 automated metabolic gas analysis systems. *Eur J Appl Physiol*. 2013;113(3):753-762.
8. Welch W, Strath S, Swartz A. Congruent validity and reliability of two metabolic systems to measure resting metabolic rate. *Int J Sports Med*. 2015;36(5):414-418.
9. St-Onge MP, Rubiano F, Jones A, Heymsfield SB. A new hand-held indirect calorimeter to measure postprandial energy expenditure. *Obes Res*. 2004;12(4):704-709.
10. Fraipont V, Preiser J-C. Energy estimation and measurement in critically ill patients. *JPEN J Parenter Enteral Nutr*. 2013;37(6):705-713.
11. Damask M, Weissman C, Askanazi J, Hyman A, Rosenbaum S, Kinney J. A systematic method for validation of gas exchange measurements. *Anesthesiology*. 1982;57(3):213-218.
12. White MS, Shepherd JA, McEniery JA. Energy expenditure in 100 ventilated, critically ill children: improving the accuracy of predictive equations. *Crit Care Med*. 2000;28(7):2307-2312.
13. Alam D, Hulshof P, Roordink D, et al. Validity and reproducibility of resting metabolic rate measurements in rural Bangladeshi women: comparison of measurements obtained by Medgem™ and by Deltatrac™ device. *Eur J Clin Nutr*. 2005;59(5):651-657.
14. Tissot S, Delafosse B, Bertrand O, Bouffard Y, Viale J, Annat G. Clinical validation of the Deltatrac monitoring system in mechanically ventilated patients. *Intensive Care Med*. 1995;21(2):149-153.
15. Júdice PB, Hamilton MT, Sardinha LB, Zderic TW, Silva AM. What is the metabolic and energy cost of sitting, standing and sit/stand transitions? *Eur J Appl Physiol*. 2016;116(2):263-273.
16. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. *Med Sci Sports Exer*. 2004;36(9):1625-1631.
17. Markwald RR, Melanson EL, Smith MR, et al. Impact of insufficient sleep on total daily energy expenditure, food intake, and weight gain. *Proc Natl Acad Sci U S A*. 2013;110(14):5695-5700.
18. Wells J, Fuller N. Precision and accuracy in a metabolic monitor for indirect calorimetry. *Eur J Clin Nutr*. 1998;52(7):536-540.

19. Miodownik S, Melendez J, Carlon VA, Burda B. Quantitative methanol-burning lung model for validating gas-exchange measurements over wide ranges of FIO_2. *J Appl Physiol*. 1998;84(6):2177-2182.
20. Elia M, Livesey G. Theory and validity of indirect calorimetry during net lipid synthesis. *Am J Clin Nutr*. 1988;47(4):591-607.
21. Ravussin E, Bogardus C. Relationship of genetics, age, and physical fitness to daily energy expenditure and fuel utilization. *Am J Clin Nutr*. 1989;49(5):968-975.
22. Nieman DC, Austin MD, Dew D, Utter AC. Validity of COSMED's quark CPET mixing chamber system in evaluating energy metabolism during aerobic exercise in healthy male adults. *Res Sports Med*. 2013;21(2):136-145.
23. Ashcraft CM, Frankenfield DC. Validity test of a new open-circuit indirect calorimeter. *JPEN J Parenter Enteral Nutr*. 2015;39(6):738-742.
24. Foss Ø, Hallen J. Validity and stability of a computerized metabolic system with mixing chamber. *Int J Sports Med*. 2005;26(7):569-575.
25. Rising R, Whyte K, Albu J, Pi-Sunyer X. Evaluation of a new whole room indirect calorimeter specific for measurement of resting metabolic rate. *Nutr Metab (Lond)*. 2015;12(1):46.
26. Carter J, Jeukendrup AE. Validity and reliability of three commercially available breath-by-breath respiratory systems. *Eur J Appl Physiol*. 2002;86(5):435-441.
27. Rietjens G, Kuipers H, Kester A, Keizer H. Validation of a computerized metabolic measurement system (Oxycon-Pro®) during low and high intensity exercise. *Int J Sports Med*. 2001;22(4):291-294.
28. Bassett DR, Howley ET, Thompson DL, et al. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J Appl Physiol*. 2001;91(1):218-224.
29. Gore CJ, Clark RJ, Shipp NJ, Van Der Ploeg GE, Withers RT. CPX/D underestimates VO (2) in athletes compared with an automated Douglas bag system. *Med Sci Sports Exerc*. 2003;35(8):1341-1347.
30. Salminen R, Aunola S, Mälkiä E, Vuori I. Computerized breath-by-breath analysis of respiratory variables during exercise. *Med Prog Technol*. 1981;9(1):27-32.
31. Schoeller DA. The effect of holiday weight gain on body weight. *Physiol Behav*. 2014;134:66-69.
32. Kien CL, Bunn JY, Ugrasbul F. Increasing dietary palmitic acid decreases fat oxidation and daily energy expenditure. *Am J Clin Nutr*. 2005;82(2):320-326.
33. Cooper JA, Watras AC, Shriver T, Adams AK, Schoeller DA. Influence of dietary fatty acid composition and exercise on changes in fat oxidation from a high-fat diet. *J Appl Physiol*. 2010;109(4):1011-1018.
34. Hawkins K, Hansen K, Schoeller D, Cooper J. Effect of exercise on the diurnal variation in energy substrate use during a high-fat diet. *Eur J Appl Physiol*. 2012;112(11):3775-3785.
35. Hibi M, Oishi S, Matsushita M, et al. Brown adipose tissue is involved in diet-induced thermogenesis and whole-body fat utilization in healthy humans. *Int J Obes*. 2016;40(11):1655-1661.
36. Murgatroyd P, Davies H, Prentice A. Intra-individual variability and measurement noise in estimates of energy expenditure by whole body indirect calorimetry. *Br J Nutr*. 1987;58(3):347-356.