Assessing validity and reliability of Resting Metabolic Rate in six gas analysis systems

Jamie A. Cooper, M.S. [Graduate Student][1], **Abigail C. Watras, M.S. [Research Coordinator]** [1], **Matthew J. O'Brien, B.A. [Pulmonary Function Lab Manager]**[2], **Amy Luke, Ph.D [Associate Professor]**[3], **Jennifer R. Dobratz, M.S. [Graduate Student]**[4], **Carrie P. Earthman, Ph.D, R.D. [Assistant Professor]**[4], and **Dale A. Schoeller, Ph.D [Professor]**[1]

[1]*Department of Nutritional Sciences, University of Wisconsin-Madison, WI, USA, 1415 Linden St. Madison, WI 53706, ph: 608-261-1902 fax: 608-262-5860*

[2]*Pulmonary Lab-University of Wisconsin Hospital and Clinics, Madison, WI, USA. 600 Highland Ave. Madison, WI 53792, ph: 608-263-7001 fax: 608-263-7002*

[3]*Department of Preventive Medicine and Epidemiology, Loyola University, IL, USA, 2160 South First Avenue Maywood, IL 60153, ph: 708-327-9011 fax: 708-327-9009*

[4]*Department of Food Science and Nutrition, University of Minnesota, MN, USA, 1334 Eckles Avenue St. Paul, MN 55108-6099, ph: 612-624-9278 fax: 612-625-5272*

Abstract

The Deltatrac Metabolic Monitor (DTC), one of the most popular indirect calorimetry systems for measuring resting metabolic rate (RMR) in human subjects, is no longer being manufactured. This study compared five different gas analysis systems to the DTC. Resting metabolic rate was measured by the DTC and at least one other instrument at three study sites for a total of 38 participants. The five indirect calorimetry systems included: MedGraphics CPX Ultima, MedGem, Vmax Encore 29 System, TrueOne 2400, and Korr ReeVue. Validity was assessed using paired t-tests to compare means while reliability was assessed by using both paired t-tests and root mean square calculations with F tests for significance. Within-subject comparisons for validity of RMR revealed a significant difference between the DTC and Ultima. Bland-Altman plot analysis showed significant bias with increasing RMR values for the Korr and MedGem. Respiratory exchange ratio (RER) analysis showed a significant difference between the DTC and the Ultima and a trend for a difference with the Vmax (p = 0.09). Reliability assessment for RMR revealed that all instruments had a significantly larger coefficient of variation (CV) (ranging from 4.8% to 10.9%) for RMR compared to the 3.0 % CV for the DTC. Reliability assessment for RER data showed none of the instrument CV's were significantly larger than the DTC CV. The results were quite disappointing, with none of the instruments equaling the within person reliability of the DTC. The TrueOne and Vmax were the most valid instruments in comparison with the DTC for both RMR and RER assessment. Further testing is needed to identify an instrument with the reliability and validity of the DTC.

Introduction

Measurement of resting metabolic rate (RMR) is used in clinical and research settings. RMR measured by indirect calorimetry under standard conditions provides information at rest in the form of oxygen consumption (VO_2), carbon dioxide production (VCO_2), and respiratory exchange ratio ($RER = VCO_2 / VO_2$). Resting metabolic rate and RER are important for both clinical and research settings as they provide invaluable information regarding energy requirements and what fuels are being oxidized at rest. Recently, several gas analysis systems have been developed for the collection of metabolic information both at rest and during exercise. It is imperative that newly developed instruments used for metabolic measurements are valid and reliable. The validity of an instrument refers to its accuracy while the reliability refers to its reproducibility or repeatability. Information regarding both validity and reliability, however, is not available for many instruments currently being marketed. In addition, because few instruments are compared under standard conditions, it is difficult to determine how much of the total variation is due to physiologic rather than instrument variation. Because day-to-day subject variability of 5% is generally reported (8), an ideal gas analysis system should have a CV for RMR of less than 3% in order to not inflate the physiologic variation (determined by $CV_{Total} = (5^2 + 3^2)^{1/2} = 6\%$ variation).

The Deltatrac II (DTC) Metabolic Monitor (VIASYS Healthcare, Inc., SensorMedics, Yorba Linda, CA) has been well-established as a valid and reliable criterion reference system over the past 15 years (1). However, the DTC is no longer available. Thus, there is a need to validate new systems that closely mimic the validity and reliability of the DTC for nutrition studies, which is the aim of this study. Three study sites, the University of Wisconsin-Madison (UW), Loyola University Medical Center (LUMC) and the University of Minnesota (UMN), compared at least one gas analysis system to the DTC. Data was collected from all three study sites in 2005 and 2006.

Methods

University of Wisconsin (UW)

Subjects—Five men and seven healthy women with an average age of 24±11 years and body mass index (BMI) of 21.8±2.1 kg/m^2 were recruited from the UW-Madison through on-campus advertising to participate in a two-day RMR study. The study was approved by the Institutional Review Board (IRB) of the UW-Madison and informed written consent was obtained. Exclusion criteria included a history of metabolic or pulmonary disease, implanted electrical devices, and claustrophobia. All participants completed the two-day study.

Protocol—RMR was measured on four different instruments in a random order on two consecutive days, using a cross-over design. Testing was conducted after a 5h fast between 0900 and 1700 hours. Participant preparation and testing procedures have been described elsewhere (2). RMR was measured for 30 minutes on each instrument. Participants returned the following day and repeated the same testing procedures. For data analysis, the first five and last five minutes of testing were excluded once the subjects were in steady state.

Instruments—The DTC and the MedGraphics CPX Ultima (Ultima) (Medical Graphics Corporation, St. Paul, MN) respiratory systems were used at this site. The DTC incorporates a ventilated hood system and a mixing chamber. The Ultima uses a mouthpiece and generates

breath-by-breath data. Data generated from both instruments included VO_2 and VCO_2, which was used to calculate RMR and RER using the modified Weir equation (10). The DTC was calibrated to reference gases prior to each participant while representatives for the Ultima calibrated and conducted all testing with their instrument to ensure that it was operating correctly.

Alcohol burns were performed on the DTC prior to and after participant testing during each study day but the calibration was not changed daily. A known volume of alcohol was burned and the actual yield of VO_2 and VCO_2 was compared to theoretical values that had been calculated based on the moles of alcohol used.

University of Minnesota (UMN)

Subjects—Sixteen obese women (age 49±9 years and BMI 47.9±7.2 kg/m^2) enrolled in a longitudinal study underwent RMR testing before bariatric surgery. Exclusion criteria have been described previously (5). The study was approved by the IRB at the UMN, and subjects gave written, informed consent.

Protocol—Resting metabolic rate was measured by the DTC and MedGem (Microlife USA, Golden, CO) hand-held device in a random order, after an overnight stay (fasting) at the General Clinical Research Center (GCRC) at the UMN. Subject preparation and testing procedures have been described elsewhere (5). DTC measurements spanned 20 minutes and the first five minutes were discarded. The MedGem measurements span up to 10 minutes. The first 2 minutes are discarded, and with a rolling boxcar methodology, the device evaluates for steady state. If steady state is not achieved, the average of minutes two to 10 are utilized.

Instruments—The MedGem (with nose clip) measures VO_2 on a dual-channel oxygen sensor and does not measure VCO_2. RMR is calculated assuming an RER of 0.85 with the modified Weir Equation (10) (RMR = $(3.941 \times VO_2) + (0.85 \times 1.106 \times VO_2) - X$, with X being a factor for nitrogen excretion). The DTC was calibrated to reference gases and the MedGem was auto-calibrated prior to measurements.

Loyola University Medical Center (LUMC)

Subjects—Two men and eleven women (age 43±8 years and BMI 27.6±9.9 kg/m^2), and three boys and two girls (age 7±2 years and BMI of 19±6.1 kg/m^2) were recruited from the staff and their families of LUMC to participate. The study was approved by the IRB for the Protection of Human Subjects at LUMC. Exclusion criteria included active weight loss, history of metabolic or pulmonary disease, implanted electrical devices or claustrophobia. Nine of the 13 adults completed the two-day protocol while the remaining completed a one-day modified protocol.

Protocol—RMR was measured on four different instruments in a random order on two nonconsecutive days, in a cross-over design: the DTC, TrueOne 2400 (TrueOne) (Parvo Medics, Sandy, UT), the Vmax Encore 29 System (Vmax) (Viasys Healthcare, Inc., Yorba Linda, CA), and the Korr ReeVue (Korr) (Korr Medical Technologies, Salt Lake City, UT). Participant preparation and testing procedures have been described elsewhere (7). Fasted participants were measured for 20 minutes on each instrument and the first five minutes were excluded. For participants enrolled only in the one-day protocol, RMR was measured using the DTC, Vmax, and Korr. Enrollment in the one- or two-day protocol depended upon availability of the participant.

Instruments—Both the Vmax and TrueOne used a ventilated canopy and mixing chamber and generated VO_2 and VCO_2, which is used to calculate RMR and RER using the modified

Weir equation (10). The DTC, Vmax and TrueOne were all calibrated with reference gases prior to each participant. The Korr, which uses a mouthpiece and nose clip and measures VO_2 only, was auto-calibrated prior to each participant. An assumed RER of 0.85 is used in a modified Weir equation to calculate RMR (10). To ensure proper working condition, VIASYS and ParvoMedics representatives performed device set up, calibration, and actual testing of the Vmax and TrueOne instruments, respectively. For the DTC, alcohol burns were done three times throughout the study using a method similar to that of the UW site.

Statistical Analysis

Data were analyzed for all sites by UW using SAS version 8.2 package (SAS Institute Inc, Cary, NC). No statistical analyses were performed on a between- site basis. Statistical analyses were performed on RMR from all instruments and on RER from all instruments except for the MedGem and Korr. Paired t-tests were used to examine the reliability of the DTC by comparing means from Day 1 and Day 2 for UW and LUMC. To test validity, Day 1 values at all study sites were used. Mean differences between each instrument and the DTC were calculated using paired t-tests when only one instrument was being compared to the DTC at a study site. When multiple instruments were being compared to the DTC (LUMC), comparisons were done using a repeated measures ANOVA with contrast statements limiting the comparisons of each instrument to just the DTC. Within-subject differences, as well as within-subject CVs (percentage), were calculated for all instrument comparisons. To calculate the reliability of each instrument separate from any variability or uncertainty due to the DTC, the following formula was used: $RMS^2_{Total} = (RMS^2_{DTC})/2 + RMS^2_X$, where RMSx represents the root mean square of the instrument of interest. F-tests were used to test for significance. Finally, the Bland-Altman technique (3) was used to examine the differences between each gas analysis system and the DTC for both RMR and RER. Statistical significance was $p < 0.05$.

Results and Discussion

Internal Validity and Reliability of the DTC

Alcohol burn tests performed at UW and LUMC tested the accuracy of the DTC. For UW, burns revealed CO_2 at 98.9±1.2% of theoretical and O_2 at 98.9±1.1% of theoretical. LUMC burns yielded 101.9±2.1% and 100.3±0.9% of theoretical for CO_2 and O_2 respectively. Correction factors developed based on this data were applied to DTC data from UW and LUMC study sites.

The reliability of the DTC was also assessed at UW and LUMC from day-one and day-two measurements by calculating within-subject differences in kcal/d and as a CV percentage. RMR data for UW showed within-subject differences of 6±68 kcal/d and CV of 3.0%, while RER data yielded within-subject differences of 0.01±0.05 and CV of 4.0%. For LUMC, RMR within-subject differences were 4±74 kcal/d and CV of 3.6% while RER within-subject differences were -0.03±0.05 and 4.9% CVs. Paired t-test were used to compare the within-subject differences and were not significant (NS).

Validity

Validity comparisons between each instrument and the DTC for both RMR and RER at each site are listed in Table 1. For RMR, a paired t-test showed a significant difference between the Ultima and the DTC. No differences were seen between the Korr, Vmax, MedGem, or TrueOne versus DTC. Within-subject CVs were smallest for the TrueOne and largest for the Korr. Finally, a Bland-Altman analysis (Figure 1) showed significant bias with increasing RMR for the Korr and MedGem versus DTC. RER data was available on three instruments (Korr and MedGem measure VCO_2). A paired t-test showed a significant RER difference for the Ultima,

as well as a trend for a difference with the Vmax (p=0.09), versus the DTC. Within-subject CVs were smallest for Vmax and largest for Ultima. Bland-Altman analysis (Figure 2) resulted in no significant bias for RER. Together these results show that for the five instruments being compared to the DTC, the TrueOne and Vmax were valid for RMR, while only the TrueOne was valid for RER.

Reliability

The reliability of each instrument was determined by within-subject CV analyses (Table 2) removing the within-subject CV attributed to the DTC from the total CV. No repeated measures were made for the DTC at UMN, thus no within-subject CV were calculated for the MedGem. F tests showed that all instruments had significantly larger CVs for RMR compared to the RMR CV for the DTC. None of the RER CVs was significantly larger than the CV for the DTC. These results indicate that while RER data were reliable for the Ultima, Vmax, and TrueOne, none of the instruments had highly reliable measures of RMR although the TrueOne was the closest to being a reliable instrument for assessing RMR and RER.

The present study is the first to compare these five different gas analysis systems to the DTC. However, other studies have reported comparisons between individual instruments (1;4;6;8; 9). Our results on the validity of the TrueOne are in agreement with previous work although comparisons were made to the Douglas bag method and not the DTC (4). We found that the MedGem was not valid which has been reported previously (1;6). Conversely, others have found the MedGem to provide a valid measure of RMR (8;9). However, some limitations in these studies may exist based on a lack of randomization for instrument testing or testing two different methods simultaneously. Unfortunately, published data are lacking for the other instruments used in this study; therefore, comparisons of our results with other studies cannot be made.

With any study, certain limitations exist. Because this was a collaboration between three different institutions, there were some site-to-site variations in study protocol and subject characteristics. However, we employed a within-subject study design so that each instrument was compared to the DTC that was also used at that particular study site. This eliminated artifacts arising from variation between sites in terms of protocol and participants. Another limitation is that methanol burns were only performed on the DTC. Corrections were made on the DTC data because this is a standard calibration procedure for this instrument. While the DTC and all other instruments were calibrated daily with reference gases or auto calibration, these methods do not test flow rate of the instrument. We therefore recommend that all gas analysis systems incorporate an alcohol burn test or a comparable accuracy assessment on a frequent basis rather than relying solely on reference gas calibration to ensure accurate measurement and to maximize instrument performance.

Conclusion

Because the DTC is no longer being manufactured, another valid and reliable gas analysis system is needed. We compared five systems with the DTC and found that the TrueOne and the Vmax were the most valid gas analysis systems of those tested for measuring both RMR and RER relative to the DTC; however, none of the gas analysis systems tested can be considered adequately reliable for use in a research setting, although the TrueOne comes close. This is disappointing considering that the DTC has been shown to be a valid and reliable instrument. We can only assume that the development engineers of these other instruments are not aware that the ideal respiratory gas analyzer for nutritional studies should be accurate within several percent and have a within-person reliability of 3% or better. Clearly, the impetus is on the manufacturers to meet the requirements of the research community.

Reference List

1. Alam DS, Hulshof PJ, Roordink D, Meltzer M, Yunus M, Salam MA, van Raaij JM. 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;59:651–657. [PubMed: 15798776]

2. Blanc S, Schoeller DA, Bauer D, Danielson ME, Tylavsky F, Simonsick EM, Harris TB, Kritchevsky SB, Everhart JE. Energy requirements in the eighth decade of life. Am J Clin Nutr 2004;79:303–310. [PubMed: 14749238]

3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310. [PubMed: 2868172]

4. 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;98:139–151. [PubMed: 16896734]

5. Dobratz JR, Sibley SD, Beckman TR, Valentine BJ, Kellogg TA, Ikramuddin S, Earthman CP. Comparison of Handheld to Metabolic Cart Indirect Calorimetry for Resting Energy Expenditure Assessment in Extremely Obese Women. Topics in Clin Nutr 2007;22(2):115–129.

6. Hlynsky J, Birmingham CL, Johnston M, Gritzner S. The agreement between the MedGem indirect calorimeter and a standard indirect calorimeter in anorexia nervosa. Eat Weight Disord 2005;10:e83–e87. [PubMed: 16682866]

7. Luke A, Rotimi CN, Adeyemo AA, Durazo-Arvizu RA, Prewitt TE, Moragne-Kayser L, Harders R, Cooper RS. Comparability of resting energy expenditure in Nigerians and U.S. blacks. Obes Res 2000;8:351–359. [PubMed: 10968726]

8. St-Onge MP, Rubiano F, Jones A Jr, Heymsfield SB. A new hand-held indirect calorimeter to measure postprandial energy expenditure. Obes Res 2004;12:704–709. [PubMed: 15090640]

9. Stewart CL, Goody CM, Branson R. Comparison of two systems of measuring energy expenditure. JPEN J Parenter Enteral Nutr 2005;29:212–217. [PubMed: 15837782]

10. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol 1949;109:1–9. [PubMed: 15394301]
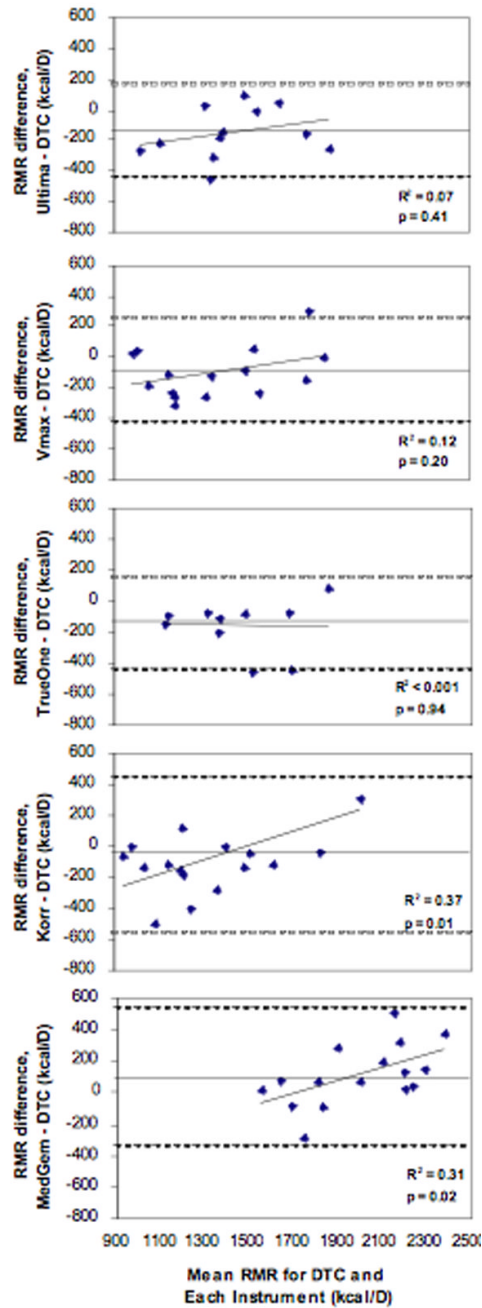
Figure 1.
Bland Altman plots of Resting Metabolic Rate (RMR) for each instrument compared to the DTC. Straight lines represent the average RMR difference while dashed lines represent ± 2SD.
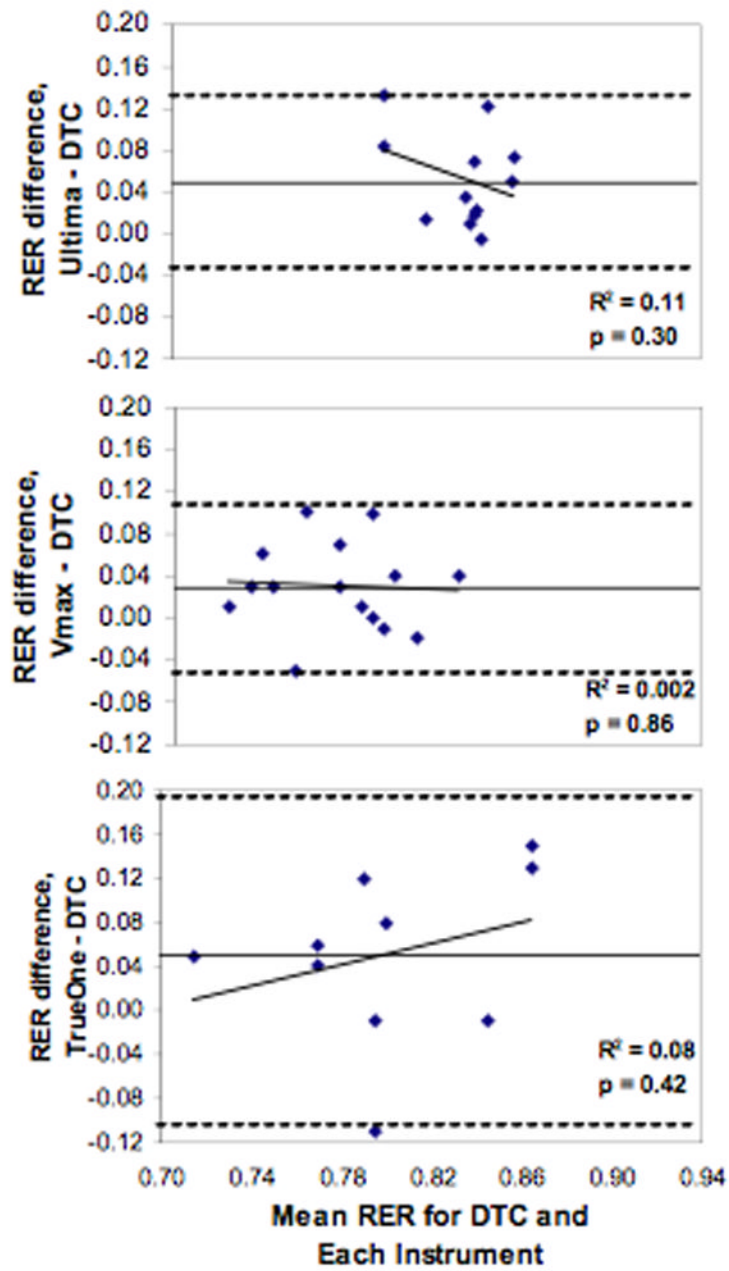
Figure 2.
Bland Altman plots of Respiratory Exchange Ratio (RER) for each instrument compared to the DTC. Straight lines represent the average RER difference while dashed lines represent ± 2SD.

Table 1

Validity of five respiratory gas analysis systems compared to the DTC

Gas Analyzer	Study Site	Sample Size	RMR[a]			RER[b]		
			Within Subject Difference (kcal/D)	T-test or Contrast statement (p value)	Within Subject CV[c] %	Within Subject Difference (kcal/D)	T-test or Contrast statement (p value)	Within Subject CV[c] %
Ultima	UW	n = 12	107 ± 180	2.7 (0.02)[*]	11.1	0.05 ± 0.05	-3.7 (0.004)[*]	5.5
Korr	LUMC	n = 17	14 ± 252	0.02 (0.89)	12.2	NA	NA	NA
Vmax	LUMC	n = 15	-26 ± 155	0.06 (0.8)	8.4	-0.03 ± 0.04	2.9 (0.09)	4.5
TrueOne	LUMC	n = 10	-6 ± 131	0.01 (0.96)	5.4	-0.01 ± 0.06	0.11 (0.7)	5.0
MedGem	UMN	n = 16	-121 ± 199	-1.0 (0.32)	7.7	NA	NA	NA

[*] Significant at $p < 0.05$, Paired t-test or repeated measures ANOVA

[a] RMR – Resting Metabolic Rate

[b] RER – Respiratory Exchange Ratio

[c] CV – Coefficient of Variation

Within Subject Difference reported as Mean ± SD

Table 2

Estimated CV of instruments independent of DTC CV

Gas Analyzer	RMR[a] Within Subject CV[c] (%)	RER[b] Within Subject CV[c] (%)
Ultima	10.9[†]	4.7
Korr	11.9[†]	NA
Vmax	8.0[†]	2.9
TrueOne	4.8[*]	3.6

[*] Significant at $p < 0.05$, F-test

[†] Significant at $p < 0.01$, F-test

[a] RMR – Resting Metabolic Rate

[b] RER – Respiratory Exchange Ratio

[c] CV – Coefficient of Variation