Within and between-day reliability of energetic cost measures during treadmill walking

Audra Davidson1, Emily S. Gardinier1 and Deanna H. Gates1,2*

Abstract: The efficacy of assistive devices used during walking is often measured as a reduction in metabolic cost. Metabolic cost is typically assessed within a day or on multiple days, yet the benefit of performing within-day vs. between-day metabolic assessments is unknown. The purpose of this study was to determine the within-day minimal detectable change of standard measures of physiologic performance using a conventional portable metabolic system (K4b² Cosmed, Rome, Italy), and compare these to between-day values. Twenty healthy adults completed two identical data collection sessions on separate days. In each session they performed three bouts of treadmill walking interspersed with three bouts of rest while oxygen consumption (\(\dot{VO}_2\)), carbon dioxide production, and heart rate were measured. Intraclass correlation coefficients (ICC) and minimal detectable change values were calculated for non-resting within-day, as well as all between-day comparisons. All within-day measures were clinically reliable (ICCs > 0.96), while between-day measures were generally less reliable (ICCs > 0.82). Within-day minimal detectable change values (walking heart rate = 4.9 bpm; gross \(\dot{VO}_2\) = 0.80 mL/kg/min; net \(\dot{VO}_2\) = 0.80 mL/kg/min; cost of transport = 0.022 J/Nm) were about half as large as between-day values. The results of this study suggest that, where possible, physiologic changes should be assessed within a single day of testing to maximize ability to detect small changes in performance.

ABOUT THE AUTHORS
Audra Davidson is a researcher and Dr Emily S. Gardinier is a postdoctoral fellow in The Rehabilitation Biomechanics Laboratory, run by Dr Deanna H. Gates. In the Rehabilitation Biomechanics Laboratory, we focus on the study of repetitive human movements such as walking and reaching. Throughout these studies, we aim to determine which aspects of movement a person actively controls and how this function can be modeled most effectively. Along with governing control strategies, these models can be used to design both passive and active devices, which can mimic biological function and restore or improve function in individuals with disability. Another focus of our research is determining appropriate outcomes to measure performance with new technology. Without such guidelines, it is difficult to assess the efficacy of assistive devices. The following work aids in this endeavor, providing additional tools for the evaluation of assistive devices.

PUBLIC INTEREST STATEMENT
A key component of the design and prescription of medical devices that assist in walking is a thorough understanding of how patients will benefit from their use. Reducing the energy patients must expend to walk is often a primary goal of such devices. In order to accurately determine if this goal is being met, it is important for researchers to separate everyday changes in energy expenditure from changes caused by using the device. This work aims to provide a standard against which changes in energy expenditure within one day of testing can be measured. With this benchmark, researchers will have a better understanding of which assistive devices result in meaningful reductions in energy expenditure. This will allow for more rigorous assessments of assistive walking devices, ultimately providing physicians and patients with a more thorough understanding of their benefits.
1. Introduction

A reduction in metabolic cost is one of the primary goals of assistive devices used during walking such as prostheses (Herr & Grabowski, 2012; Houdijk, Pollmann, Groenewold, Wiggerts, & Polomski, 2009; Major, Twiste, Kenney, & Howard, 2014; Segal et al., 2012; Zelik et al., 2011), orthoses (Balaban et al., 2007; Menotti, Laudani, Damiani, Mignogna, & Macaluso, 2014; Menotti, Laudani, Damiani, Orlando, & Macaluso, 2014; Zissimopoulos, Fatone, & Gard, 2007), or exoskeletons (Malcolm, Derave, Gale, & De Clercq, 2013; Sawicki & Ferris, 2008, 2009). The efficacy of such devices is often tested within a single day (Balaban et al., 2007; Houdijk et al., 2009; Malcolm et al., 2013; Menotti, Laudani, Damiani, Mignogna, et al., 2014; Menotti, Laudani, Damiani, Orlando, et al., 2014; Segal et al., 2012; Zelik et al., 2011; Zissimopoulos et al., 2007) or across multiple days where acclimation to the device is required (Major et al., 2014; Sawicki & Ferris, 2008, 2009). Typical changes observed in energetic measures range from 3 to 8% for prostheses (Herr & Grabowski, 2012; Major et al., 2014; Segal et al., 2012; Zelik et al., 2011; Zissimopoulos et al., 2007) or 2 to 12% for orthoses (Balaban et al., 2007; Menotti, Laudani, Damiani, Orlando, et al., 2014; Sawicki et al., 2007) or 7 to 15% for exoskeletons (Malcolm et al., 2013; Sawicki & Ferris, 2008, 2009).

Metabolic device evaluation studies have typically utilized portable metabolic units such as the Cosmed K4b2 system (Cosmed, Rome, IT). The between-day reliability and/or minimal detectable change (MDC) of several physiologic measures associated with the K4b2 system in various populations have been previously reported (Darter, Rodriguez, & Wilken, 2013; Stookey et al., 2013; Thomas, Buckon, Schwartz, Sussman, & Aiona, 2009; Welch, Strath, & Swartz, 2015). Although comparisons are often made within a single day of testing, the reliability and MDC of energetic measures collected within a single day during treadmill walking are not known. Given the benefit of collecting data on a single day where factors such as hydration and fitness are more controlled, we would expect the consistency of these measures to improve. Therefore, the purpose of this study was to calculate within-day MDCs for several physiological measurements, including heart rate, oxygen consumption, and cost of transport, using a K4b2 system, and compare these to between-day values. Establishing within-day MDCs provides a benchmark by which investigators can determine true changes in human performance and allow for more precise assessments of assistive devices used during walking.

2. Methods

2.1. Participants

Twenty healthy adults (9 females, 11 males) were recruited using flyers and an electronic community registry (UMClinicalStudies.org). Their mean (SD) age, height, weight and BMI were 30 (15) years, 1.74 (0.08) m, 72.4 (11.3) kg, and 23.8 (2.8) kg/m², respectively (see supplemental material for additional details). All subjects provided written informed consent prior to participation in this institutionally approved study.

2.2. Experimental protocol

Participants performed two identical test sessions in the morning at least one day apart (14 (13) days). Participants were instructed to abstain from caffeine, avoid strenuous exercise, and fast for at least 4 h prior to testing.

A K4b2 portable metabolic system was used to measure the rate of oxygen consumption ($\dot{V}O_2$, mL/kg/min), carbon dioxide production ($\dot{V}CO_2$, mL/kg/min), and heart rate. This system consists of a mask, heart-rate monitor, and a collection unit that wirelessly transmits data to a laptop where it can be observed in real time (Figure 1(A)). First, participants were asked to remain seated and still for
at least 5 min to obtain metabolic cost measures for seated rest. Then, they walked on a treadmill at a speed normalized to leg length and approximated a comfortable walking speed (average speed = 1.22 (0.03) m/s) (Gates, Wilken, Scott, Sinitski, & Dingwell, 2012). Participants walked for a minimum of 5 min, or until they reached steady-state, which was identified in real-time by visually observing a plateau in \( \dot{V}O_2 \) (average time = 9.1 (1.98) min; Figure 1(B)). Once this plateau was reached, participants walked for an additional 3 min to obtain metabolic cost measures for steady-state walking. This process was then repeated such that all participants completed three bouts of walking which were interrupted by 5 min bouts of seated rest.

Steady-state walking intervals were later objectively determined using MATLAB (MathWorks, Natick, MA, USA, v.R2014b). Steady-state was defined as a 3 min interval wherein the minute average \( \dot{V}O_2 \) and \( \dot{V}CO_2 \) varied by <10%, and the average respiratory quotient \( \dot{V}CO_2 / \dot{V}O_2 \) was between 0.7 and 1.0 and varied by <5% (McClave et al., 2003; Reeves, Davies, Bauer, & Battistutta, 2004). We began by assessing the last 3 min of walking. The interval was then iteratively shifted earlier by one breath until all steady-state criteria were met.

Performance measures included gross \( \dot{V}O_2 \) (mL/kg/min), net \( \dot{V}O_2 \) (gross \( \dot{V}O_2 \) − seated rest \( \dot{V}O_2 \): mL/kg/min), cost of transport (COT, J/Nm), and heart rate (HR, bpm). Seated rest \( \dot{V}O_2 \) was calculated as the average \( \dot{V}O_2 \) over the last 90 s of the first resting bout on each day and was included only in between-day analyses. Metabolic rate was estimated from \( \dot{V}O_2 \) and \( \dot{V}CO_2 \) according to standard formulae (Brockway, 1987). COT, defined as the energy cost of moving 1 kg of body mass 1 m (J/Nm), was calculated by dividing net metabolic rate by walking speed (Mian, Thom, Ardigo, Narici, & Minetti, 2006).

2.3. Statistical analysis

Intraclass correlation coefficients (ICCs) were calculated to assess reliability within- and between-days (Portney & Watkins, 2009). Specifically, we used an ICC\(_{3,k}\) model for absolute agreement. Within-day ICCs were calculated for the first day of testing. Between-day ICCs were calculated from averages across bouts for each day. Only variables with a significant F-test (p < 0.05) had a sufficiently heterogeneous group for the ICC to be valid (Portney & Watkins, 2009). ICC > 0.90 indicated reasonable agreement for clinical measures, ICC > 0.75 indicated good agreement, while ICC < 0.75 indicated poor to moderate agreement (Portney & Watkins, 2009). Pooled SD and ICCs were used to estimate the standard error of the measure (SEM) according to the equation:
where $SD_{pooled}$ is the pooled standard deviation across testing bouts or days.

Subsequently, minimal detectable change (MDC$_{95}$) values were estimated for each variable according to the equation (Weir, 2005):

$$MDC_{95} = SEM \times 1.96 \sqrt{2}$$

where 1.96 corresponds to a 95% confidence level. MDCs are defined as the smallest change in a measure that exceeds random variation (Lassere, van der Heijde, & Johnson, 2001), and are often used to quantify the limits of expected variation in a measure. Bland-Altman plots comparing daily average values were used to identify proportional bias (Bland & Altman, 1999). Pearson’s correlations were performed to determine if testing separation, subject age, or subject weight were correlated with percent difference in performance measures across days for each subject. All statistical analyses were performed in SPSS version 22 (IBM Corp., Armonk, NY, USA).

### 3. Results

All 20 subjects were included in the analysis. Data from one subject (S08) was lost for one session on day two of testing due to equipment malfunction. F-tests were significant for all variables ($p < 0.01$), indicating sufficient heterogeneity. No proportional bias was observed in any measure (see Supplemental material).

All within-day and one between-day measure (walking HR) indicated clinically meaningful reliability (ICC > 0.96; Table 1). Between-day measures had good reliability (ICC > 0.82), except for seated rest $\dot{V}O_2$ which had poor reliability (ICC = 0.65).

The within-day MDCs were about half the magnitude of between-day MDCs (Table 1). The MDC for walking heart rate exhibited the smallest comparative difference, with a within-day MDC of 4.9 bpm, which was about 30% less than its between-day MDC of 7.1 bpm (Table 1).

### Table 1. Physiologic measures between- and within-day for treadmill walking

<table>
<thead>
<tr>
<th>Physiologic measure</th>
<th>Mean (SD) Within-day reliability</th>
<th>Between-day reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>Resting heart rate (bpm)</td>
<td>69.0 (8.5)</td>
<td>69.9 (9.7)</td>
</tr>
<tr>
<td>Seated rest $\dot{V}O_2$ (mL/kg/min)</td>
<td>3.58 (0.55)</td>
<td>3.83 (0.66)</td>
</tr>
<tr>
<td>Walking heart rate (bpm)</td>
<td>94.2 (8.4)</td>
<td>96.6 (9.8)</td>
</tr>
<tr>
<td>Walking gross $\dot{V}O_2$ (mL/kg/min)</td>
<td>12.15 (1.64)</td>
<td>12.35 (1.4)</td>
</tr>
<tr>
<td>Walking net $\dot{V}O_2$ (mL/kg/min)</td>
<td>8.66 (1.56)</td>
<td>8.48 (1.32)</td>
</tr>
<tr>
<td>Cost of transport (J/Nm)</td>
<td>0.345 (0.052)</td>
<td>0.349 (0.045)</td>
</tr>
</tbody>
</table>
The percent change across days for energetic measures were not significantly correlated with the days between testing \((p \geq 0.24)\), subject age \((p \geq 0.17)\), or subject weight \((p \geq 0.54)\) for any performance measure (Table 2).

### 4. Discussion

The purpose of this study was to determine within-day MDCs for various energetic measures of walking performance and compare these to between-day values. All within-day measures were deemed clinically reliable \((ICC > 0.96)\). Between-day comparisons of physiologic measurements generally exhibited good agreement \((ICC > 0.82)\), but were less reliable than within-session measures. Therefore, consistent with our expectations, the K4b2 system was able to reliably detect smaller changes in physiologic measures for comparisons made within a day than comparisons made across different testing days.

The results for between-day reliability were consistent with those reported previously on healthy adults using the K4b2 system (Darter et al., 2013) and another portable metabolic system (Blessinger et al., 2009). While seated rest and walking HR MDCs in the present study were 3–4 bpm lower than those previously reported (Darter et al., 2013), between-day oxygen consumption MDCs were roughly 0.86 and 0.61 mL/kg/min higher than previously reported values (Darter et al., 2013) for gross and net \(\dot{VO}_2\), respectively. This discrepancy may be the result of a larger separation between collection days in the current study. In the present study, testing days were separated by an average of two weeks, compared to five days of average separation in other studies (Blessinger et al., 2009; Darter et al., 2013). The longer between-day interval in the present study may have allowed for greater changes in fitness, body weight, stress levels, or other factors known to influence energetic measurements (Donahoo, Levine, & Melanson, 2004).

Consistent with prior studies (Darter et al., 2013; Thomas et al., 2009), resting \(\dot{VO}_2\) was less reliable than walking \(\dot{VO}_2\). Littlewood et al. (2002) suggested that the K4b2 device is highly variable when acquiring resting energetic measurements, and may only be suitable for measuring the energetics of activity (Littlewood et al., 2002). When comparing the Cosmed system to the Deltatrac II, a metabolic cart specifically designed to measure resting energetics, Littlewood et al. (2002) revealed a mean bias for REE and \(\dot{VO}_2\) of 268 ± 702 kcal/day and 51.6 ± 125.6 mL/min, respectively. The exact reason for this discrepancy is not well understood, as its design is similar to devices that have been reported to produce consistently accurate resting measures (Littlewood et al., 2002). However, supine rest, longer and stricter fasting periods, and resting prior to data collection have been shown to improve the between and within-day reliability of resting measures obtained with the Cosmed K4b2 system (Welch et al., 2015). These stricter resting protocols should be used if resting measures are of particular importance to a study.

There are several limitations to this study. As mentioned above, the protocol may have limited our ability to determine accurate energetic measures for seated rest. To obtain more reliable measures we could have had participants rest for longer periods of time. This may have improved the reliability

### Table 2. Pearson’s correlations for energetic measures and subject details

<table>
<thead>
<tr>
<th></th>
<th>Walking heart rate (bpm)</th>
<th>Walking gross (\dot{VO}_2) (mL/kg/min)</th>
<th>Walking net (\dot{VO}_2) (mL/kg/min)</th>
<th>Cost of transport (J/Nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days between testing</td>
<td>-0.027</td>
<td>-0.094</td>
<td>-0.275</td>
<td>-0.113</td>
</tr>
<tr>
<td>Age</td>
<td>0.320</td>
<td>0.112</td>
<td>-0.036</td>
<td>0.095</td>
</tr>
<tr>
<td>Weight</td>
<td>0.147</td>
<td>-0.023</td>
<td>-0.042</td>
<td>-0.081</td>
</tr>
</tbody>
</table>
for both seated rest and net VO₂. Second, although participants were asked to abstain from exercise and to fast and avoid caffeine, we have no way of determining whether they actually did these things. Failure to adhere to these guidelines could cause increased variability from day-to-day. This would be true for any study unless participants were observed over extended periods of time in the laboratory. Third, there was a wide variation in subject characteristics and range of time between testing sessions. Testing days were separated by an average of two weeks and the age range of subjects tested was large (19–65 years). These factors were not correlated with changes in any measure across days, however. Although this may increase the variability of the data, testing a wide variety of subjects with varying lengths of testing separation increases the generalizability of the results.

In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day. Within-day MDCs were about half the size of between-day values. These results emphasize the benefit of using within-day methodological designs when feasible as the ability to determine differences in metabolic cost dramatically improves.

### Supplementary material

Supplementary material for this article can be accessed here http://dx.doi.org/10.1080/23311916.2016.1251028.

### Funding

This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health [grant number 2UL1TR000433]. Deanna H. Gates is supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health [grant number K12HD073945].

### Author details

Audra Davidson1
E-mail: davidsoa@umich.edu
Emily S. Gardinier1
E-mail: egardin@umich.edu
Deanna H. Gates1,2
E-mails: gatesd@umich.edu, deanna.h.gates@gmail.com

ORCID ID: http://orcid.org/0000-0002-6086-0913

1 School of Kinesiology, University of Michigan, Ann Arbor, MI 48109, USA.
2 Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

### Citation information

Cite this article as: Within and between-day reliability of energetic cost measures during treadmill walking, Audra Davidson, Emily S. Gardinier & Deanna H. Gates, Cogent Engineering (2016), 3: 1251028.

### Declaration

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### References


Donahoo, W. T., Levine, J. A., & Melanson, E. L. (2004). Reliability and minimum detectable change using the K4b2 system is able to detect small changes in physiologic measurements within a day. In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day. In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day. In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day. In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day. In conclusion, the K4b² system is able to detect small changes in physiologic measurements within a day.

### Supporting information

Supplementary material for this article can be accessed here http://dx.doi.org/10.1080/23311916.2016.1251028.


