2012

Bioharness multivariable monitoring device. Part II: Reliability

James A. Johnstone

Paul A. Ford

Gerwyn Hughes
University of San Francisco, ghughes@usfca.edu

Tim Watson

Andrew T. Garrett

©Journal of Sports Science and Medicine (2012) **11,** 409-417
http://www.jssm.org

Research article

Bioharness™ multivariable monitoring device. Part II: Reliability

James A. Johnstone [1]✉, Paul A. Ford [3], Gerwyn Hughes [1], Tim Watson [2] and Andrew T. Garrett [4]
[1] School of Life Sciences, and [2] School of Health and Emergency Professions, University of Hertfordshire, UK; [3] British Olympic Association, London, UK; [4] Department of Sport, Health and Exercise Science, University of Hull, UK

Abstract
The Bioharness™ monitoring system may provide physiological information on human performance but the reliability of this data is fundamental for confidence in the equipment being used. The objective of this study was to assess the reliability of each of the 5 Bioharness™ variables using a treadmill based protocol. 10 healthy males participated. A between and within subject design to assess the reliability of Heart rate (HR), Breathing Frequency (BF), Accelerometry (ACC) and Infra-red skin temperature (ST) was completed via a repeated, discontinuous, incremental treadmill protocol. Posture (P) was assessed by a tilt table, moved through 160°. Between subject data reported low Coefficient of Variation (CV) and strong correlations(r) for ACC and P (CV< 7.6; r = 0.99, p < 0.01). In contrast, HR and BF (CV~19.4; r~0.70, p < 0.01) and ST (CV 3.7; r = 0.61, p < 0.01), present more variable data. Intra and inter device data presented strong relationships (r > 0.89, p < 0.01) and low CV (<10.1) for HR, ACC, P and ST. BF produced weaker relationships (r < 0.72) and higher CV (<17.4). In comparison to the other variables BF variable consistently presents less reliability. Global results suggest that the Bioharness™ is a reliable multivariable monitoring device during laboratory testing within the limits presented.

Key words: Physiological technology, reproducibility of measurement, exercise.

Introduction

The development of new monitoring technology has assisted in allowing high-quality data to be recorded in a variety of free living active situations (Achten and Jeukendrup, 2003; Jobson et al., 2009). New measuring technology can collate information on multiple integrated physiological and activity variables which can be assessed in real-time or downloaded post-performance. Reproducibility, or repeatability, of data using new measuring technology is crucial if advancement of ecologically valid assessment of activity is to continue. The Bioharness™ (Version 1), can collate information on multiple integrated physiological and activity variables which can be assessed in real-time or post-performance. The Bioharness™ simultaneously measures five variables (i.e. heart rate, breathing frequency, skin temperature, activity and posture), which can be monitored wirelessly in real time or downloaded from the device after the activity. Previous literature supports the use of each individual variable which is integrated in to the device; Heart rate (HR) through chest mounted electrodes (Grossman et al., 2006; Kent et al., 2009; Leger and Thivierge, 1988; Macfarlane et al., 1989), Breathing Frequency (BF) through respiratory inductive plethysmograghy (Grossman, et al., 2006; 2010; Kent et al., 2009; McCool et al., 2002; Witt et al., 2006), infra-red Skin Temperature (ST) (Burnham et al., 2006; Gant et al., 2006; Hershler et al., 1992; Matsukawa et al., 2000), Tri axial Accelerometry (ACC) (Brage et al., 2005; Powell and Rowlands, 2004; Rowlands et al., 2003) and Posture (P) (inclinometry) (Hansson et al., 2001; 2006) both the latter variables using piezoelectric technology. The Bioharness™ device is being used within a variety of applied free living situations including the assessment of physical activity and within the emergency services for both rescuers and those being rescued. There is a lack of peer reviewed literature considering the reproducibility of the Bioharness™ device with only the breathing frequency variable tested (Hailstone and Kilding, 2011). Measurements made by new technology in any environment must have known clarity as to what variability may exist (Atkinson and Nevill, 1998; Welk et al., 2004) and to our knowledge there has been no peer reviewed paper published detailing the reliability of all five variables associated with the device. Therefore, the aim of this paper was to assess the reliability of each variable within the Bioharness™ device in an exercise based laboratory situation.

Methods

General design

To assess the reproducibility of the Bioharness™ variables appropriate assessment protocols were identified. A between (n = 10, using 1 Bioharness™ device) and within subject (n = 1, testing 4 different Bioharness™ devices) design, using a repeated treadmill protocol, allowed the assessment of ST, HR, BF and ACC, with the latter 3 variables being assessed at different velocities. P variable was assessed as a whole data set through a separate mechanical protocol. All data collection was synchronized to one timeline linked to a laptop computer. For consistency, a standardised technique for the fitting of all equipment was completed by one experienced researcher throughout the duration of the study.

Apparatus

The Bioharness™ (Version1) is worn against the skin by the participant via an elasticated strap attached around the chest (50 g, 50 mm width). The monitoring device (weight 35 g, 80x40x15mm), which attaches to the front of the chest strap, acts as a data logger or transmitter, has a memory of up to 480 hours and battery life of up to 10 hours in logging mode. Five variables are measured si-

Received: 30 March 2012 / Accepted: 04 May 2012 / Published (online): 01 September 2012

multaneously, time stamped and exportable to Excel. HR data is captured through electrode sensors housed within the chest strap (i.e. detecting R wave forms) at 250 Hz and reported as beats per minute ($b \cdot min^{-1}$). BF is provided using a capacitive pressure sensor (18 Hz) that detects circumference expansion and contraction of the torso producing an output as breaths per minute ($br \cdot min^{-1}$). Tri axial ACC, using piezoelectric technology (i.e. cantilever beam set up) samples at 18 Hz and reports 1Hz in counts per second ($ct \cdot sec^{-1}$). It is a micro electro-mechanical sensor accelerometer with a capacitive measurement scheme and is sensitive along 3 orthogonal axes (vertical (x), sagittal (z) and lateral (y)). Acceleration data is monitored in gravitational force (g) in a range of -3 to +3 g on each single axis or as Vector Magnitude Units (VMU) which is an integrated value over the previous 1 second epoch:

$$VMU = \sqrt{A_X^2 + A_Y^2 - A_Z^2}$$

The P variable uses similar piezoelectric technology as described. Acting as an inclinometer, data in angular degrees (°) ranges between -90° and +90°, it assesses the degrees the monitoring device is "off vertical". ST data is collected through an infrared sensor behind a clear window on the apex of the monitoring device. It records peripheral skin temperature at the inferior sternum. This sensor reports data (1 Hz) in degrees Celsius (°C).

Participants

After securing local institutional ethical agreement 10 male volunteers (age 20.5 ± 2.1yrs, body mass 70.4 ± 9.4 kg, body stature 1.77 ± 0.10 m) who were physically active, injury free and familiar with using a treadmill consented to participate. Participants were asked to refrain from consuming alcohol, caffeine, keep hydrated and rested 24 hours before testing. On arrival to the laboratory anthropometrical data were taken with stature (Seca 214, Birmingham, UK) and body mass (Seca 761, Birmingham, UK) measured (Stewart and Eston, 2007).

Test Procedures

Reproducibility of HR, BF, ACC and ST

Reproducibility of these variables were assessed by participants completing an adapted discontinuous incremental treadmill protocol (Rowlands et al., 2004). In a thermo-neutral laboratory environment (23.9 ± 1.7°C) the protocol consisted of 6 discontinuous incremental stages: rest ($0 km \cdot h^{-1}$), walking (4 and 6 $km \cdot h^{-1}$); and running (8, 10 and 12 $km \cdot h^{-1}$) performed on an electronically driven treadmill (HP Cosmos Mercury, Germany). Stages lasted a total of 8 minutes; 2 minutes rest, 4 minutes being active (i.e. walking or running) followed by 2 minutes recovery. Data was collected every 5 seconds for the last 90 seconds of each of the respective active stages. Participants were fitted with the Bioharness™ 15 minutes prior to test commencing and remained on the treadmill throughout. The retest was completed 5 days from the date of first test, at the same time of day as test 1, with participants adhering to the same 24 hour pre-test protocol described previously.

Reproducibility of P

In a controlled procedure, reproducibility of P data was tested by securing Bioharness™ devices to an inversion (i.e. tilt) table (F500III, STL International) was moved through 160° as noted elsewhere (Bernmark and Wiktorin, 2002). The tilt table was calibrated (to 0°) using a spirit level and then moved through a 160° (+80 to -80) at 10° intervals, pausing for 10 seconds, at each interval allowing data to be recorded. This process was then repeated.

Data Analysis

Data was exported to statistical software packages (Excel Microsoft Windows, USA; SPSS v17, SPSS Inc, Chicago, USA) for analysis. When assessing reproducibility a range of statistical procedures have been cited in combination with descriptive data are available for researchers providing a comprehensive summary (Atkinson and Nevill, 1998; Hopkins et al., 2009; Kent et al., 2009; Nunan et al., 2008; Sandercock et al., 2005). Reliability of the data was assessed through the use of descriptive statistics (mean ± standard deviation (S)), Change in mean, 95% Confidence Limits and reliability statistics, Coefficient of Variation, Inter Class Correlations. The change in mean and associated 95% confidence limits will provide an indication of absolute variation between the data sets. Coefficient of variation (CV) expresses the S as a proportion of the mean, is considered a dimensionless statistic and therefore easier to compare variation between protocols. A somewhat arbitrary CV acceptable boundary of < 10% for reliability has been cited in some papers though this is not accepted unanimously in the literature (Atkinson and Nevill, 1998; Currell and Jeukendrup, 2008; Hopkins, 2000). Information on the relationship between sets of data is provided by correlation coefficients. Intra class coefficients are more sensitive to systematic bias and also can be used for multiple retests so have been preferred within reliability studies (Atkinson and Nevill, 1998; Hopkins, 2000). Boundaries for correlation statistics are not confirmed though amalgamated thoughts of Leger and Thivierge (1988) and Hopkins (2000) suggest; $r > 0.9$ Excellent or very strong, $r = 0.7 – 0.9$ Very Large, $r = 0.7 – 0.5$ Good to moderate, $r < 0.5$ Moderate to minor. All of these statistics analysed collectively will provide a clear overview on the reproducibility of data.

Characteristics of the data set were considered and appropriate statistical procedures followed thereafter. After plotting the between subject predicted against the residuals for HR and BF (Figure 1), data were considered to be non-uniform (i.e. heteroscedastic or not normally distributed) so were logarithmically (log) transformed in order to provide a true interpretation (Atkinson and Nevill, 1998; Hopkins et al., 2000; 2009). Descriptive data for these variables were reported in absolute values and reliability statistics presented log transformed which was determined in order for comparison with other studies to occur, the majority of which report absolute data.

Previously research assessing reliability of a monitoring device has removed data sets when data is clearly erroneous in the belief that a technical breakdown has occurred with the system (Leger and Thivierge, 1988).

Analysis completed which includes highly erroneous data sets would possible reduce the practical usefulness of the results especially if this data was linked to a small clearly identifiable number of participants. The reporting of data removal (i.e. cleaning) has been used as additional evidence for reproducibility with high volumes of data being removed possibly reducing the reliability of the device. Therefore reporting of raw and clean data sets was completed on HR and BF data where some highly erroneous data was noted. Based on estimated maximal values of each physiological variable (McArdle et al., 2009), day-to-day biological variation (Achten and Jeukendrup, 2003) and considering other research (Hailstone and Kilding, 2011; Leger and Thivierge, 1988) the following data set removal criteria was established; If absolute mean of a data set presented a difference of ±20 b·min^{-1} for HR, ±7 br·min^{-1} for BF in comparison to equivalent data from the specific stage, the data was removed.

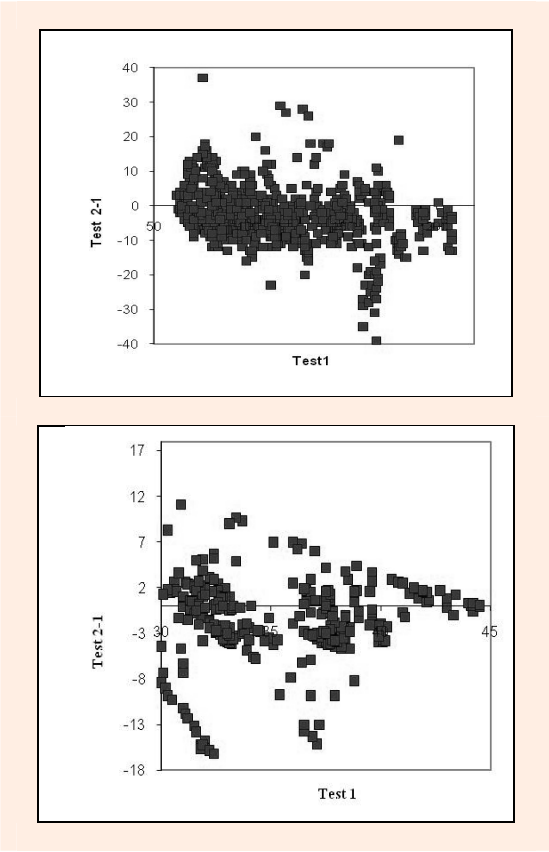

Figure 1. Residual versus predicted plot demonstrating the data spread for (a) HR and (b) BF.

Results

Overview of the reliability of the Bioharness™ (between subjects)

Between subject results (Table 1) for the whole data set note, low coefficient of variation (CV) and very strong relationships for P and ACC ($p < 0.01$). Less reliable variables are, ST, HR and BF, with ST variable having low coefficient of variation and moderate relationships. HR and BF present moderate relationships and a large CV.

Between subjects velocity specific HR reliability

HR results (Table 2) noted strong to very strong relationships ($r > 0.84$, $p < 0.01$), a lowering CV (< 6.2) and change in mean < 3.16 b·min^{-1} with velocity ≤8 km·h^{-1}. Reproducibility of data decreased at velocities at ≥ 10 km·h^{-1} with increases in change of mean (> 14.01 b·min^{-1}), CV > 24.7 and limited relationships in data.

Between subjects velocity specific BF reliability

BF data (Table 3) presents a weak relationship ($r < 0.51$), with elevated CV values (ranging 16.8 – 21.9). The change in mean remains < 1 br·min^{-1} and this value reduces from rest to the active stages.

Velocity specific results for HR and BF after erroneous data removed

Data considered to be highly erroneous was removed following the procedure described earlier. HR (n=6) and BF data (n = 8) produced data (Table 4) mirroring trends from statistics seen at lower velocities. Considering this clean data set with the other velocities, HR change in mean remained < 3.16 br·min^{-1}, CV < 6 and strong to very strong relationship ($r > 0.84$) were noted. BF data continued with < 1 br·min^{-1} for change in mean, CV presented its lowest values at 10 and 12 km^{-1} and a low to moderate relationship were identified.

Between subjects velocity specific ACC reliability results

At rest, ACC data (Table 5) presented the least reliable data with largest change in mean, largest CV and weakest relationship. As velocity increased, the change in mean reduced and became consistent, CV decreased (< 9.3) and moderate to strong ($r > 0.66$) relationship are reported.

Overview of within subject (intra and inter device) reliability of the Bioharness™

General findings for intra (Table 6) and inter (Table 7)

Table 1. Bioharness™ reproducibility across whole data set.

Variables	Descriptive Data		Reliability Data			
	Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
HR b·min^{-1}	120.6 ± 38.0	113.5 ± 35.1	-7.31	-9.07 to -5.54	19.4	.67*
BF br·min^{-1}	25.5 ± 8.1	26.0 ± 8.1	.51	.16 to .86	19.4	.71*
ACC ct·sec^{-1}	.71 ± .39	.71 ± 0.39	.002	-.001 to .006	6.5	.99*
ST degrees °C	32.5 ± 1.7	32.0 ± 2.0	-.5	-.61 to -.42	3.7	.61*
P degrees	45.1 ± 22.9	44.8 ± 23.9	-.34	-.31 to -1.00	7.6	.99*

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC). * $p < 0.01$

Table 2. Velocity specific reproducibility of HR (b·min^{-1}) data.

Velocity	Descriptive Data		Reliability Data			
	Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
0 km·hr^{-1}	80.2 ± 12.0	81.1 ± 11.7	.90	-.09 to 1.89	5.9	.84*
4 km·h^{-1}	89.5 ± 11.1	86.6 ± 11.4	-2.96	-3.72 to -2.71	4.5	.88*
6 km·h^{-1}	103.6 ± 11.6	100.9 ± 12.2	-2.66	-3.55 to -1.76	4.3	.87*
8 km·h^{-1}	135.7 ± 19.2	132.5 ± 18.5	-3.16	-4.05 to -2.27	3.4	.94*
10 km·h-^{1}	153.9 ± 23.7	138.3±33.9	-14.01	-19.73 to -8.28	24.7	.08
12 km·h^{-1}	160.4 ± 38.3	141.1 ± 42.6	-19.30	-28.43 to -10.16	30.5	.04

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC). * p < 0.01

device reliability presented mainly strong statistics for HR, ACC, P and ST (r > 0.89, p < 0.01; CV ≤ 10.1). BF variable performed less effectively in comparison (r < 0.72; CV 11.4 – 17.4). No data was considered highly erroneous so analysis includes all data.

Velocity specific intra and inter device reliability of the HR, BF and ACC variable of the Bioharness™

Further velocity specific intra and inter device results note low CV for HR (< 7.3) and ACC (< 10) with a general trend of decreasing variance with increasing treadmill velocity. At 0 km·h^{-1}, ACC presented high inter and intra device variance (CV range ~50 – 130) which reduced at the onset of activity. BF presented CV values < 11.4 with one exception during the inter device analysis (CV = 17.8). Correlation values for all variables were predominately low (r < 0.70; p < 0.05) with exception of one HR result (10 km·h^{-1}, r = 0.86).

Discussion

Reliability of the Bioharness™

Multivariable monitoring devices within sport and exercise can now provide time synchronised data which possibly could allow for further insights in to performance. Ensuring that a comprehensive precision of measurement assessment has occurred will allow for an understanding of the variability which exists and is a crucial step in achieving credibility in the market place (Welk, et al., 2004). The aims of the study were to assess the reliability of the Bioharness™ monitoring device due to limited information being available on this issue.

Overall results suggest that, the Bioharness™ produces adequately reliable data for HR, ST, ACC and P, with the latter two variables presenting the most accurate data. Erroneous data at higher velocities for HR and BF variables sets suggests caution should be applied to data collected during activities involving movement above 10 km·h^{-1}. For both BF and HR variable, between subject reliability improved after data cleaning at higher velocities. ST achieved the least test-retest reliability between subject, though produced stronger results in the within subject design.

Reliability of heart rate (HR)

HR data suggests adequate reproducibility across both testing designs at moderate velocity (≥ 8 km·h^{-1}) (Table 3). Considering the between-subject design, before data cleaning, there is weaker reproducibility in the data at higher velocities (≥ 10 km·h^{-1}) which is linked to an increase in highly erroneous data being produced. This latter phenomena was not apparent in the within subject data with reliability statistics being strong (r > 0.99; CV < 6.8) throughout all velocities. After erroneous data had been removed at ≥10km·h^{-1}, the between subject data presents equivalently strong results with slight trend of decreasing CV with increases in velocity, as previously reported elsewhere (Achten and Jeukendrup, 2003). In an arguably less intense yoga environment the results improve on correlations (r= ~.6) and match CV data (1.9 – 5.7) found for the Lifeshirt™, a multi variable assessment device (Grossman et al., 2006; Kent et al., 2009) and also is similar to unpublished data (CV 1.7 – 6.7) from our laboratory on the Polar HR monitor (T31, Polar Electro, Kempele, Finland). There is an expectation that this data should be within credible limits as monitoring HR telemetrically through electrodes housed within a chest strap has had over 20 years of development (Achten and Jeukendrup, 2003; Laukkanen and Virtanen, 1998).

Reliability of accelerometer (ACC)

Strong reproducibility data for ACC variable was noted in all testing scenarios and supports the notion that piezoelectric technology within the device can be deemed

Table 3. Velocity specific reproducibility of BF (br·min^{-1}) data.

Velocity	Descriptive Data		Reliability Data			
	Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
0 km·hr^{-1}	16.8 ± 4.2	17.7 ± 1.9	.89	.24 to 1.53	21.9	.06
4 km·h^{-1}	19.4 ± 4.5	20.3 ± 3.3	.87	.31 to 1.43	16.8	.51*
6 km·h^{-1}	22.3 ± 4.0	22.5 ± 5.4	.17	-.57 to .90	19.8	.49*
8 km·h^{-1}	27.5 ± 4.0	27.4 ± 4.2	-.16	-1.13 to 0.81	19.4	- .21
10 km·h^{-1}	31.7 ± 4.5	32.3 ± 5.4	.58	-.46 to 1.62	19.7	- .03
12 km·h^{-1}	35.5 ± 5.7	36.1 ± 6.5	.61	-.47 to 1.69	17.9	.23

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC) * p < 0.01

Table 4. Clean HR (b·min^{-1}) and BF (br·min^{-1}) data at 10 and 12 km·h^{-1}.

Velocity	Descriptive Data		Reliability Data			
	Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
Heart rate						
10 km·h^{-1}	155.4 ± 21.0	153.4 ± 23.3	-1.98	-3.30 to -0.66	3.5	.95*
12 km·hr^{-1}	168.9 ± 21.5	168.1 ± 20.7	-.77	-2.47 to 0.93	3.7	.91*
All data	116.2 ± 35.7	113.5± 34.6	-2.68	-3.15 to -2.22	4.8	.98*
Breathing Fr						
10 km·h^{-1}	32.7 ± 3.8	33.2 ± 3.3	.51	-.21 to 1.22	10.4	.22
12 km·hr^{-1}	35.9 ± 5.8	36.7 ± 5.9	.77	-.11 to 1.65	12.7	.54*
All data	25.1 ± 8.1	25.6 ± 7.9	.52	.21 to .82	17.7	.75*

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC) * p < 0 .01

reliable (Table 5, 6 and 7). A low change in mean, low CV (< 8) and very strong relationships (r > 0.99) match or exceed previous research suggesting the ACC provides reliable data within the testing environment (Brage et al., 2003; Powell et al., 2003; G. J. Welk et al., 2000). ACC data at rest (i.e. 0 km·h^{-1}) was not included in the overall analysis as during pilot testing this data was inconsistent. During the rest stage inevitable slight irregular motion of the subject was registered as an activity count. This erratic non-rhythmical data production led to spurious variance in comparison to the remainder of the ACC data set. Piezoelectric elements are more effective in dynamic rather than a static mode (Chen and Bassett, 2005) and the data notes a lowering of CV as treadmill velocity increases (4 – 12 km·h^{-1}) which also corresponds to findings for other reliable ACC such as the RT3 (Powell, et al., 2003) and Actiheart^{TM} device (Brage et al., 2005). Additional evidence from a study incorporating a free movement trial (e.g. sit-to-stand task) produced a wide range CV (8.7 – 25.6) between subjects which further corroborate this technical finding within the Bioharness^{TM} (Brage et al., 2005; Powell and Rowlands, 2004).

Reliability of posture (P)

P variable, as assessed by a tilt table, produced good reliability statistics (r > 0.99; CV < ~10). Additional analysis was completed during the treadmill activity comparing posture during the within subject protocol which produced additional evidence that this variable is reliable (Table 1, 6 and 7). There are other tools to measure angular degrees in humans and even though it has been reported digital inclinometers are more reliable than goniometers they are not extensively used due to the expense (Venturni et al., 2006). Other research on this variable using similar technology has also demonstrated good precision of measurement (Hansson, et al., 2006). The same piezoelectric technology is used within P and ACC variable and both have demonstrated good reproducibility data.

Reliability of breathing frequency (BF)

Across both experimental designs the wider statistical analysis suggests the BF variable produced less reliable data in comparison to the other variables (Table 1, 3, 4, 6 and 7). As with the HR variable, higher running velocities (≥ 10 km·h^{-1}) lead to an increase in erroneous data occurring. After data cleaning, variance during the active stages (> 4 km·h^{-1}) seems to remain constant with slight decreases in CV at the higher velocities. Weak relationships were identified and CV values ranged from 21.9 at rest to a low of 10.4 at 10 km·h^{-1}. Interestingly, in comparison Hailstone and Kilding (2011) noted higher absolute test-retest differences (< 2.8 br·min^{-1}) but found stronger relationships in data (r = 0.86 – 0.96) when they assessed reliability BF of the Bioharness^{TM}. A difference in methodology and analysis may explain some aspects of these results, also they fail to identify the Bioharness^{TM} version used. Later versions of this technology may well use improved or different data processing algorithms. Comparing the data to other corresponding respiratory inductive plethysmography technology, a non-active environment presented stronger test-retest relationships (r = ~ 0.8) (Grossman et al., 2006) and a repeated within subject treadmill test for the Lifeshirt^{TM} device reported CV ~10 (Kent et al., 2009). Weaker data from this study of the Bioharness^{TM} could be linked to the respiratory inductive plethysmography technical set up of the device. The Lifeshirt^{TM} adopts a 2 degree (i.e. 2 measuring band) model allowing thoracic and abdominal movements to be considered in producing respiratory data (McCool et al.,

Table 5. Reproducibility of Bioharness^{TM} ACC data (Vector Magnitude Units, ct·sec^{-1}).

Velocity	Descriptive Data		Reliability Data			
	Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
0 km·hr^{-1}	.04 ± .10	.02 ± .02	-.026	-.041 to -.011	131.1	- .03
4 km·h^{-1}	.17 ± .03	.18 ± .03	.007	.003 to .010	9.3	.66*
6 km·h^{-1}	.41 ± .22	.42 ± .23	.007	.002 to .012	5.9	.99*
8 km·h^{-1}	.86 ± .15	.86 ± .13	.002	-.009 to .13	6.5	.85*
10 km·h-^{1}	1.04 ± .09	1.03 ± .09	-.010	-.020 to .001	5.0	.68*
12 km·h^{-1}	1.12 ± .10	1.13 ± .09	.006	-.004 to .016	4.3	.75*

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC) * p < 0.01

Table 6. Overview of intra device reproducibility of Bioharness™ device.

Variables	Device No	Descriptive Data		Reliability Data			
		Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
HR (b·min^{-1})	Device 4	104.5 ± 28.4	98.1 ± 27.9	-6.4	-6.99 to -5.73	2.5	.99*
	Device 5	104.4 ± 30.7	106.0 ± 29.7	1.6	.54 to 2.58	4.7	.98*
	Device 6	100.7 ± 24.1	111.0 ± 29.2	10.3	8.73 to 11.86	5.4	.97*
	Device 7	103.1 ± 28.3	102.2 ± 24.2	-.9	-2.15 to 0.41	5.9	.97*
BF (br·min^{-1})	Device 4	24.6 ± 2.8	28.1 ± 8.1	3.4	2.39 to 4.47	13.7	.57
	Device 5	29.6 ± 8.3	24.6 ± 3.7	-4.7	-6.05 to -3.88	13.2	.59
	Device 6	25.1 ± 3.9	26.1 ± 4.0	1.0	.20 to 1.81	11.4	.41
	Device 7	29.7 ± 7.4	32.5 ± 9.4	2.9	1.64 to 4.09	13.2	.72*
ACC (ct·sec^{-1})	Device 4	0.77 ± 0.42	0.80 ± 0.42	.03	.01 to .04	7.0	.99*
	Device 5	0.81 ± 0.42	0.81 ± 0.43	.01	-.01 to .02	5.4	.99*
	Device 6	0.80 ± 0.42	0.77 ± 0.41	.02	-.04 to -.01	5.6	.99*
	Device 7	0.81 ± 0.43	0.81 ± 0.43	.01	-.01 to .02	5.7	.99*
ST (°C)	Device 4	30.8 ± 1.51	30.1 ± 0.66	-.69	-.81 to -.56	1.5	.90*
	Device 5	30.1 ± 1.76	30.5 ± 1.24	.36	.25 to .48	1.5	.92*
	Device 6	29.3 ± 1.24	30.2 ± 1.33	.89	.79 to .98	1.2	.92*
	Device 7	31.2 ± 1.51	29.2 ± 0.66	-1.94	-2.17 to -1.71	2.7	.48
P (°)	Device 4	46.9 ± 25.0	45.1 ± 25.7	-1.8	-.51 to -2.91	10.1	.99*
	Device 5	44.0 ± 24.4	45.4 ± 24.8	1.4	.61 to 2.14	1.6	.99*
	Device 6	44.8 ± 23.5	44.4 ± 24.9	-.4	-1.55 to .80	5.5	.99*
	Device 7	44.8 ± 23.3	44.1 ± 25.2	-.6	-2.29 to 1.04	8.8	.99*

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC) * p < 0.01

2002) in comparison to the one thoracic measuring band used within the Bioharness™. The BF CV results suggest quite high variance especially when considering respiratory values during calibration should be within ± 3% (Zeballos et al., 2003) and evidence has presented lower CV (9.1) within maximal testing (Garrard and Emmons, 1986). Though Kent et al. (2009) reported high CV (~17) for breath-by-breath data gained from a Cosmed metalyser which also corresponds to unpublished CV data from our laboratory using a Cortex 3B metalyser (Cortex Medical, Germany). It seems that BF may be a physiological variable with higher variance, especially if analysed breath-by-breath and discipline specific data processing methods with regards to data averaging are seemingly not standardised, so could influence outcomes presented elsewhere (Kent et al., 2009).

Reliability of skin temperature (ST)

Repeatability of ST in a thermo-neutral environment during the treadmill activity produced somewhat equivocal results (Table 1, 6 and 7). ST noted lower relationships in data for between subject design (r = 0.61) than reported in other research (Burnham, et al., 2006; Gant, et al., 2006) though a low CV (3.7) was maintained.

Table 7. Overview of inter device reproducibility of Bioharness™ device.

Variables	Device No	Descriptive Data		Reliability Data			
		Test 1 M ± S	Test 2 M ± S	Change in mean	95% CL	CV	ICC
HR (b·min^{-1})	Device 4	104.5 ± 28.4	-	-	-	-	104.5 ± 28.4
	Device 5	103.4 ± 30.6	-1.04	-1.85 to -.24	3.9	.99*	103.4 ± 30.6
	Device 6	100.3 ± 24.1	-3.02	-4.61 to -1.43	6.8	.95*	100.3 ± 24.1
	Device 7	103.6 ± 28.2	3.68	2.27 to 5.09	6.5	.96*	103.6 ± 28.2
BF (br·min^{-1})	Device 4	24.6 ± 2.8	-	-	-	-	24.6 ± 2.8
	Device 5	29.3 ± 8.2	4.7	3.67 to 5.74	12.3	.59	29.3 ± 8.2
	Device 6	25.0 ± 3.9	-4.4	-5.59 to -3.14	15.8	.48	25.0 ± 3.9
	Device 7	29.7 ± 7.3	4.6	3.34 to 5.95	17.4	.33	29.7 ± 7.3
ACC (ct·sec^{-1})	Device 4	.77 ± .42	-	-	-	-	.77 ± .42
	Device 5	.79 ± .42	.02	.01 to .04	6.2	.99*	.79 ± .42
	Device 6	.79 ± .42	-.01	-.02 to .01	5.3	.99*	.79 ± .42
	Device 7	.80 ± .42	.02	.01 to .04	5.9	.99*	.80 ± .42
ST (°C)	Device 4	30.8 ± 1.7	-	-	-	-	30.8 ± 1.7
	Device 5	30.2 ± 1.7	-.62	.72 to .52	1.2	.95*	30.2 ± 1.7
	Device 6	29.4 ± 1.2	-.81	.94 to .68	1.7	.89*	29.4 ± 1.2
	Device 7	31.2 ± 1.5	1.76	1.64 to 1.89	1.5	.89*	31.2 ± 1.5
P (°)	Device 4	46.9 ± 25.0	-	-	-	-	46.9 ± 25.0
	Device 5	44.0 ± 24.4	-2.9	-4.25 to -1.50	9.3	.99*	44.0 ± 24.4
	Device 6	44.8 ± 23.5	.8	-.22 to 1.72	8.0	.99*	44.8 ± 23.5
	Device 7	44.8 ± 23.3	.0	-.45 to .45	0.9	.99*	44.8 ± 23.3

Tabular report of reliability statistics: Descriptive statistics, Standard Deviation (S), 95% Confidence Limits (95% CL), Change in Mean, Coefficient of Variation (CV) and Intra Class Correlations (ICC). * p < 0.01

For the within subject design, except for one device, relationships were strong ($r > 0.89$) which coupled with a low CV (< 1.7) suggest the device attains good reliability. The difference in reliability between the two testing designs could be related to the positioning of the infra-red device relative to the subject. In the single subject (intra device) design, when the subject somatotype was standardised, data were more consistent. Previously the infra-red device placement, including lens angle and distance from skin, have been identified as important in attaining credible data and could have influenced the inter subject data collection (Hershler et al., 1992; Matsukawa et al., 2000).

Limitations

It is important to identify if technical breakdown of equipment occurs as this is noted as an additional indication of reliability (Leger and Thivierge, 1988; Terbizan et al., 2002). Failure to clean data with a transparent system may present skewed data. Between subject numbers reduced from n = 10 to n = 6 for HR and n = 8 for BF at the highest velocities (i.e. $\geq$ 10 km·h^{-1}) while in contrast no data was removed in the within subject testing. Though not formally assessed, the disparity between the two testing designs and number of useable data sets warrants further discussion. Increased number of errors for HR and BF variables between subject could have occurred due to the data signal that the monitoring device requires becoming corrupted by varying cross subject movement artefacts (Cho et al., 2011; Witt et al., 2006) such as; EMG activity (Boudet and Chamoux, 2000; McArdle et al., 2009), changes in the mechanics of breathing (McArdle et al., 2009; McCool et al., 2002) or movement of the monitoring device (Clarenbach et al., 2005; Leger and Thivierge, 1988). The full data set and stronger reliability results from the single subject design attained suggests that inter subject differences may influence the device's ability to collect precise information. Body type was not formally assessed though anecdotally the within subject participant, from which the full data set was attained, possessed ectomorph characteristics. Although firm conclusions cannot be drawn from this issue further work should be completed on the effects of body shape, generic user set-up information and data credibility. Moreover, in this study the participant sex was controlled therefore results are limited and further investigation in to the reliability of the device in other populations should be considered.

Velocity specific analysis allows for identification of micro level limits in the equipment though at times data sets begin to reduce in number and this can affect the statistical analysis. For example, low correlation values within inter and intra reliability velocity specific analysis, if reported out of context, could be misinterpreted though could attributed to low number of data points especially as when data was amalgamated *r* values were deemed strong.

Some of the variation in the data collected can be attributed to a number of sources and needs to be factored in to any analysis of new technology. Inter and intra subject biological variation (i.e. circadian rhythm, fatigue or subject motivation) and general "noise" from the testing environment (i.e. EMG) can influence reproducibility of data. Additionally some technical error will exist which is outside the control of the researcher all of which influence statistics outcomes and conclusions drawn (Achten and Jeukendrup, 2003; Hopkins et al., 2000; Massin et al., 2000).

Conclusion

The Bioharness^{TM} can be considered a reliable device within the limitations presented in this study. Within subject reliability data is very strong suggesting the fit of the device on different individuals could be an important factor in attaining consistent data especially for HR and BF. Even considering the latter, some caution regarding data quality should be noted when physical activity is > 10 km·h^{-1}. This increased variance at higher velocities may have implications for the devices use in sporting contexts though technical upgrades of newer versions of the Bioharness^{TM} may improve this issue . Being able to access a reliable and valid monitor which measures a range of physiological variables simultaneously in free living conditions will allow for further invaluable understanding of human performance in a variety of environments.

The Bioharness^{TM} device is designed to enable naturalistic physiologically based monitoring to occur across differing free movement scenarios without the need for obtrusive invasive equipment. The design limitations associated with incorporating multi-variable monitoring within a device which must be unobtrusive to the wearer may place some limitations on the effectiveness of the functioning of individual elements. Free movement physiological data, which the Bioharness^{TM} aims to capture, is inherently variable (Welk et al., 2004). Therefore, having established the validity and reliability of the Bioharness^{TM} device in a controlled laboratory environment the authors suggest that the next progression for the Bioharness^{TM} device is assessment in a less controlled, field based setting. This will allow for a more comprehensive understanding of its capacities in the mode of use it was intended for.

Acknowledgement

This research was supported with Bioharness^{TM} Version 1 equipment loaned from Zephyr Technology Ltd., MD, USA. The authors have no commercial or proprietary interest in this device.

References

Achten, J. and Jeukendrup, E. (2003) Heart rate monitoring: applications and limitations. *Sports Medicine* **33(7),** 517-538.

Atkinson, G. and Nevill, A.M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine* **26(4),** 217-238.

Bernmark, E. and Wiktorin, C. (2002) A triaxial accelerometer for measuring arm movements. *Applied Ergonomics* **33,** 541-547.

Boudet, G. and Chamoux, A. (2000) Heart rate monitors and abnormal heart rhythm detection. *Archives of Physiology and Biochemistry* **108(4),** 371-379.

Brage, S., Brage, N., Franks, P.W., Ekelund, U. and Wareham, N.J. (2005) Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical Nutrition* **59,** 561-570.

Brage, S., Wedderkopp, N., Franks, P.W., Anderson, L.B. and Froberg,

K. (2003) Reexamination of validity and reliability of the CSA monitor in walking and running. *Medicine & Science in Sport & Exercise* **35(8),** 1447-1454.

Burnham, R.S., McKinley, R.S. and Vincent, D.D. (2006) Three Types of Skin-Surface Thermometers. *American Jourbal of Physical Medicine & Rehabilitation* **85(7),** 553-558.

Chen, K.Y. and Bassett, D.J. (2005) The technology of Accelorometry-Based Activity Monitors: Current and Future. *Medicine & Science in Sport & Exercise* **37(11),** 490-500.

Cho, H., Koo, S., Lee, J., Cho, H., Kang, D., Song, H., Lee, J.W., Lee, K.H. and Lee, Y.J. (2011) Heart monitoring garments using textile electrodes for healthcare applications. *Journal of Medical Systems* **35(2),** 189-191.

Clarenbach, C.F., Senn, O., Brack, T., Kohler, M. and Bloch, K.E. (2005) Monitoring of ventilation during exercise by a portable respiratory inductive plethysmography. *American College of Chest Phsicians* **128,** 1282-1290.

Currell, K. and Jeukendrup, A.E. (2008) Validity. reliability and sensitivity of measures of sporting performance. *Sports Medicine* **38(4),** 297-316.

Gant, N., Atkinson, G. and Williams, C. (2006) The Validity and Reliability of Intestinal Temperature during Intermittent Running. *Medicine & Science in Sport & Exercise* **38(11),** 1926-1931.

Garrard, C.S. and Emmons, C. (1986) The reproducibility of respiratory responses to maximum exercise. *Respiration* **49,** 94 - 100.

Grossman, P., Spoerle, M. and Wilhelm, F.H. (2006) Reliability of Respiratory Tidal Volume Estimation by Means of Ambulatory Inductive Plethysmography. *Biomedical Instrumentation & Technology* **42,** 193-198.

Grossman, P., Wilhelm, F.H. and Brutsche, M. (2010) Accuracy of ventilatory measurement employing ambulatory inductive plethysmography during tasks of everyday life. *Biological Psychology* **84,** 121-128.

Hailstone, J. and Kilding, A.E. (2011) Reliability and validity of the Zephyr Bioharness to measure respiratory responses to exercise. *Measurement in Physical Education and Exercise Science* **15,** 293 - 300.

Hansson, G.A., Arvidsson, I., Ohlsson, K., Nordander, C., Mathiassen, S.E., Skerfving, S. and Balogh, I. (2006) Precision of measurements of physical workload during standardised manual handling. Part II: Inclinometry of head, upper back, neck and upper arms. *Journal of Electromyography and Kinesiology* **16,** 125-136.

Hansson, G.A., Asterland, P., Holmer, N.G. and Skerfving, S. (2001) Validity and reliability of triaxial accelerometers for inclinometry in posture analysis. *Medical and Biological Engineering and Computors* **39,** 405-413.

Hershler, C., Conine, T.A., Nunn, A. and Hannay, M. (1992) Assessmnet of an infrared non-contact sensor for routine skin temperature monitoring: a preliminary study. *Journal of Medical Engineering and Technology* **16(3),** 117-122.

Hopkins, W.G. (2000) A new view of statistics: Summarising data - Precision of measurement. Available from URL: http://www.sportsci.org/resource/stats/precision.html

Hopkins, W.G., Marshall, S.W., Batterham, A.M. and Hanin, J. (2009) Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sport and Exercise* **41(1),** 3-12.

Jobson, S.A., Nevill, A.M. and Atkinson, G. (2009).Choose your primary outcome variables with care. *Journal of Sports Sciences* **27(4),** 313-314.

Kent, L., O'Neill, B., Davidson, G., Nevill, A., Elborn, J.S. and Bradley, J.M. (2009) Validity and reliability of cardiorespiratory measurements recorded by the LifeShirt during exercise tests. *Respiratory Physiology and Neurobiology* **167,** 162-167.

Laukkanen, R.M.T. and Virtanen, P.K. (1998) Heart Rate Monitors: State of the Art. *Journal of Sports Science* **16,** 3-7.

Leger, L. and Thivierge, M. (1988) Heart rate monitors: Validity, stability and functionality. *The Physician and Sports Medicine* **16(5),** 143-151.

Macfarlane, D.J., Phill, D., Fogarty, B.A. and Hopkins, W.G. (1989) The accuracy and variability of commercially available heart rate monitors. *The New Zealand Journal of Sports Medicine* **17(4),** 51-53.

Massin, M., Maeyns, K., Withofs, N., Ravet, F. and Gérard, P. (2000) Circadian rhythm of heart rate and heart rate variability. *Archives of Disease in Childhood* **83,** 179 -182.

Matsukawa, T., Ozaki, M., Nishiyama, T., Imanura, M. and Kumazawa, T. (2000) Comparisonof infrared thermometer with thermocouple for monitoring skin temperature *Critical Care Medicine* **28(2),** 532-536.

McArdle, W.D., Katch, F.L. and Katch, V.L. (2009) *Exercise physiology: Nutrition, energy and human performance.* 7th edition. Philidelphia, Pen.: Lippincott Williams and Wilkins.

McCool, F.D., Wang, J. and Ebi, K.L. (2002) Tidal volume and respiratory timing derived from a portable ventilation monitor. *Clinical Investigations in Critical Care* **122(2),** 684-691.

Nunan, D., Donovan, G., Jakovljevic, D.G. and Hodges, L.D. (2008) Validity and reliability of short-term heart rate variability from the Polar S810. *Medicine & Science in Sport & Exercise* **41(1),** 243-250.

Powell, S.M., Jones, D.I. and Rowlands, A.V. (2003) Technical Variability of the RT3 Accelerometer. *Medicine & Science in Sports & Exercise* **35(10),** 1773-1778.

Powell, S.M. and Rowlands, A.V. (2004) Intermonitor Variability of the RT3 Accelerometer during Typical Physical Activities *Medicine & Science in Sports & Exercise* **36(2),** 324-330.

Rowlands, A.V., Thomas, P.W., Eston, R. and Topping, R. (2004) Validation of the RT3 Triaxial Accelerometer for the Assessment of Physical Activity. *Medicine and Science in Sport and Exercise* **36(3),** 518 - 524.

Rowlands, A.V., Thomas, P.W., Eston, R.G. and Topping, R. (2003) Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine & Science in Sports & Exercise* **36(3),** 518-524.

Sandercock, G.R., Bromley, P.D. and Brodie, D.A. (2005) The reliability of short-term measurements of heart rate variability. *International Journal of Cardiology* **103,** 238-247.

Stewart, A.D. and Eston, R. (2007) *Surface Anthropometry*. Oxon: Routledge.

Terbizan, D.J., Dolezal, B.A. and Albano, C. (2002) Validity of seven commercially available heart rate monitors. *Measurement in Physical Education and Exercise Science* **6(4),** 243-247.

Venturni, C., André, A., Aguilar, B., & Giacomelli, B. (2006). Reliability of two evaluation methods of active range of motion in the ankle of healthy individuals. *Acta Fisiatr, 13*(1), 39 - 43.

Welk, G.J., Blair, S.N., Wood, K., Jones, S. and Thompson, R.W. (2000) A Comparitive evaluation of theree accelerometry-based physical activity monitors *Medicine & Science in Sport & Exercise* **32(9),** 489-497.

Welk, G.J., Schaben, J.A. aand Morrow, J.R. (2004) Reliability of acceloremetry-based activity monitors: A generalizability study. *Medicine & Science in Sport & Exercise* **36(9),** 1637-1645.

Witt, J.D., Fisher, J.R.K.O., Guenette, J.A., Cheong, K.A., Wilson, B.J. and Sheel, A.W. (2006) Measurement of exercise ventilation by a portable respiratory inductive plethysmograph *Respiratory Physiology and Neurobiology* **154,** 389-395.

Zeballos, J., Cotes, J., Fairbarn, M.S. and Crapo, R.O. (2003) Reference values for cardiopulmonary exercise testing. *American Journal of Respiratory Critical Care Medicine*, **167 (2),** 211-277.

Key points

- Heart rate and breathing frequency data increased in variance at higher velocities (i.e. ≥ 10 km.h^{-1})
- In comparison to the between subject testing, the intra and inter reliability presented good reliability in data suggesting placement or position of device relative to performer could be important for data collection
- Understanding a devices variability in measurement is important before it can be used within an exercise testing or monitoring setting

AUTHORS BIOGRAPHY

James. A. JOHNSTONE
Employment
Senior Lecturer in Sport and Exercise Science, Sport, Health and Exercise Research Group, School of Life Sciences, University of Hertfordshire, UK.
Degree
MSc
Research interest
In-match physiological monitoring of performance, strength and conditioning, talent identification.
E-mail: j.a.johnstone@herts.ac.uk

Paul A. FORD
Employment
Performance Manager, British Olympic Association
Degree
PhD
Research interest
Paediatric exercise physiology, Elite sport development.

Gerwyn HUGHES
Employment
Senior Lecturer in Sport and Exercise Science, Sport, Health and Exercise Research Group, School of Life Sciences, University of Hertfordshire, UK
Degree
PhD
Research interest
Electrotherapy, tissue repair

Tim WATSON
Employment
Principal Lecturer, Health and Emergency Professions, University of Hertfordshire, UK.
Degree
PhD
Research interest
Electrotherapy, tissue repair

Andrew T GARRETT
Employment
Senior Lecturer in Exercise Science, Department Sport, Health and Exercise Science, University of Hull, UK.
Degree
PhD
Research interest
Environmental physiology

✉ **James A. Johnstone**
School of Life Sciences, University of Hertfordshire, AL10 9AB, UK