

Research article

Bioharness™ multivariable monitoring device. Part I: Validity

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 there is limited information on its validity. The objective of this study was to assess the validity of all 5 Bioharness™ variables using a laboratory based treadmill protocol. 22 healthy males participated. Heart rate (HR), Breathing Frequency (BF) and Accelerometry (ACC) precision were assessed during a discontinuous incremental (0-12 km·h⁻¹) treadmill protocol. Infra-red skin temperature (ST) was assessed during a 45 min⁻¹ sub-maximal cycle ergometer test, completed twice, with environmental temperature controlled at 20 ± 0.1 °C and 30 ± 0.1 °C. Posture (P) was assessed using a tilt table moved through 160°. Adopted precision of measurement devices were; HR: Polar T31 (Polar Electro), BF: Spirometer (Cortex Metalyser), ACC: Oxygen expenditure (Cortex Metalyser), ST: Skin thermistors (Grant Instruments), P:Goniometer (Leighton Flexometer). Strong relationships ($r = .89$ to $.99$, $p < 0.01$) were reported for HR, BF, ACC and P. Limits of agreement identified differences in HR (-3.05 ± 32.20 b·min⁻¹), BF (-3.46 ± 43.70 br·min⁻¹) and P ($0.20 \pm 2.62^\circ$). ST established a moderate relationships (-0.61 ± 1.98 °C; $r = 0.76$, $p < 0.01$). Higher velocities on the treadmill decreased the precision of measurement, especially HR and BF. Global results suggest that the Bioharness™ is a valid multivariable monitoring device within the laboratory environment.

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

Introduction

Progress with new monitoring technology has assisted with the improvement of the collection of physiologically related data across a wide variety of free living situations. From everyday physical activity scenarios through to sporting performance new measuring technology now allows high-quality data to be recorded in increasingly ecologically valid situations (Achten and Jeukendrup, 2003; Grossman et al., 2010; Jobson et al., 2009; Soren Brage et al., 2005). A new measuring technology such as the Bioharness™ can simultaneously measure 5 physiological and activity related variables which can be monitored in real time, wirelessly, or downloaded from the device after the activity. Previous research supports the validity of each individual variable which is integrated in to the latter device; Heart rate (HR) through chest mounted electrodes (Leger and Thivierge, 1988; Macfarlane et al., 1989; Terbizan et al., 2002), Breathing Frequency (BF) through respiratory inductive plethysmography (Grossman et al., 2010; McCool et al., 2002; Witt et al., 2006), Skin Temperature (ST) using infra-red technology

(Burnham et al., 2006; Hershler et al., 1992; Matsukawa et al., 2000), Tri axial Accelerometry (ACC) (Powell and Rowlands, 2004; Rowlands et al., 2003) and Posture (P) (i.e. inclinometry) (Hansson et al., 2006; 2001) using a piezoelectric element. However, there is limited evidence linked to the precision of measurement for multi-variable monitoring devices, with only a single paper reporting on one variable of the Bioharness™ (Breathing frequency) (Hailstone and Kilding, 2011). Devices such as the Bioharness™ are being used within a variety of applied situations including physical activity and exercise monitoring, and also within the emergency professions. Measurements made by multi-variable devices in any environment must have known precision and clarity as to its validity (Atkinson and Nevill, 1998; Welk et al., 2004). The consistent agreement between the true (i.e. Criterion) and measured (i.e. Predictor) variable is the underlying principle of validity (Brunton et al., 2000; Currell and Jeukendrup, 2008). Any new technology which allows for data to be collected in free living situations must be rigorously assessed using controlled methodologies in order for precision of measurement to be known (Thomas et al., 2005; Welk, 2005). Therefore, the aim of this paper was to assess the validity of each variable measured in the Bioharness™ in relation to established criterion measures within a physically active laboratory situation.

Methods

General design

To assess the Bioharness™, appropriate established criterion measures and protocols were identified. In all testing scenarios a standardised technique for fitting all equipment was completed by one experienced researcher. Data from the adopted criterion and the Bioharness™ used one synchronized timeline linked to a laptop computer. A treadmill protocol assessed ACC, HR and BF with the latter two variables being analysed at specific velocities. ST, assessed during a cycle ergometry test, carried out in both hot and thermo-neutral conditions. P was validated using a tilt table protocol. Due to the experimental design it was only possible to analyse ACC and P as whole data sets.

Apparatus

Overview of the Bioharness™ monitoring device

The Bioharness™ (Version 1) is worn against the skin (Figure 1) by the participant via an elasticated strap attached around the chest (50 g, 50 mm width). The monitoring device (weight 35 g, 80x40x15 mm), which at-

taches to the front of the chest strap, acts as a data logger or transmitter, and has a memory of up to 480 hours and battery life of up to 10 hours in logging mode. Five variables are measured simultaneously, time stamped and exportable to Excel. HR data is captured through electrode sensors housed within the chest strap (i.e. detecting R wave forms) sampled at 250 Hz and reported as beats per minute ($\text{b}\cdot\text{min}^{-1}$). 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 ($\text{br}\cdot\text{min}^{-1}$). Tri axial ACC, using piezoelectric technology (i.e. cantilever beam set up) samples at 18 Hz and reports in counts per second ($\text{ct}\cdot\text{sec}^{-1}$; 1 Hz). 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 measured 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:

$$VMU = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

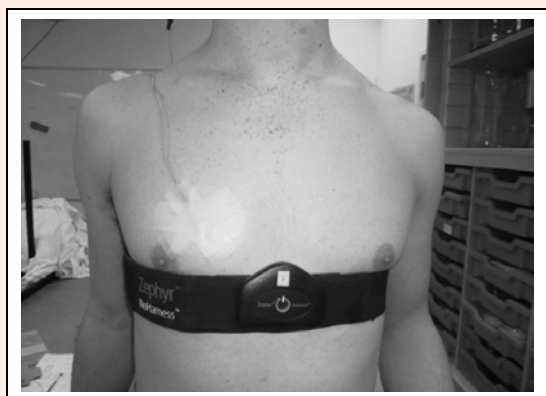


Figure 1. Picture of the Bioharness™ as worn by a subject participating in the testing process.

The P variable uses similar piezoelectric technology as described. Acting as an inclinometer, data is reported in angular degrees ($^{\circ}$), ranges between -90° and $+90^{\circ}$, monitoring how far the device is “off the vertical”. ST data is collected through an infrared sensitive 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 in degrees Celsius ($^{\circ}\text{C}$).

Participants

After securing local institutional ethical agreement 22 male volunteers (age 21.5 ± 2.8 yrs, body mass 71.4 ± 7.9 kg, body stature 1.79 ± 0.10 m) who were physical active, injury free and familiar with using a treadmill and/or cycle ergometer consented to participate. Participants were asked to refrain from consuming alcohol, caffeine, keep hydrated and rested 24 hours prior to testing. On arrival to the laboratory anthropometrical measures

(Stewart and Eston, 2007) were taken with stature (Seca 214, Birmingham, UK) and body mass (Seca 761, Birmingham, UK) measured.

Precision of Bioharness™

Heart rate (HR), Breathing Frequency (BF) and Accelerometry (ACC)

Using one standard Bioharness™ device, which was concurrently compared with adopted criterion measures, precision of the HR, BF and ACC were assessed by participants ($n = 12$) completing an adapted discontinuous incremental treadmill protocol (Rowlands et al., 2004). Adopted criterion measures within this procedure were the Polar T31 (Polar Electro, Kempele, Finland) for HR. For BF, a face mask (Hans Rudolf Inc, USA) was worn by participants in order to connect a Tripple-V spirometer which was attached to a metalyser (version 3B; Cortex Medical, Germany). Oxygen (O_2) expenditure was assessed for ACC also using the aforementioned metalyser which was calibrated prior to testing according to the manufacturers specifications. The latter criterion (O_2 expenditure) is considered an indirect measure of ACC so additionally a count of steps taken during each active stage was made for each participant. The right foot, observed by two data collectors, was counted each time it was placed on to the treadmill during a walking/running stride and the mean of the counts was used to relate to ACC.

In a thermo-neutral environment (24.1 ± 1.9 $^{\circ}\text{C}$) the protocol consisted of 6 discontinuous incremental stages (adapted from Rowlands et al 2004): rest ($0 \text{ km}\cdot\text{h}^{-1}$), walking (4 and $6 \text{ km}\cdot\text{h}^{-1}$); and running (8 , 10 and $12 \text{ km}\cdot\text{h}^{-1}$) 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 all the respective equipment 15 minutes prior to test commencing and remained on the treadmill throughout.

Infra-red skin temperature (ST)

Infra-red ST variable was assessed during an adapted version of a continuous submaximal cycle ergometer trial. Participants ($n = 10$) cycled (Monarch Ergomedic, Model 824E, Varberg, Sweden) at 60 rpm^{-1} in a University environmental chamber for 45 minutes against a resistance equivalent to 4% of body mass on two separate occasions, one week apart, in a randomised cross-over design. On one occasion the ambient temperature was set to 20 ± 0.1 $^{\circ}\text{C}$ on the other occasion set at 30.0 ± 0.1 $^{\circ}\text{C}$. A Bioharness™ device and the criterion measure, a separate skin thermistor (Type EUS-U-V5-V2; Grant Instruments, Cambridge, England), was secured on lower pectoral using medical grade tape (Hypafix, BSN Medical GmbH, Hamburg, Germany). Ambient temperature, thermistor temperature and Bioharness™ infra-red temperature were recorded at 1 minute intervals throughout the procedure.

Posture (P) (inclinometer)

This variable was assessed using reference data derived from a credible goniometry device (Daneshmandi et al., 2010), the Leighton Flexometer (Spokane, WA, USA). In a controlled procedure, both devices were secured to an inversion (i.e. tilt) table (F500III, STL International) which was moved through 160° as noted elsewhere (Bernmark & Wiktorin, 2002). The flexometer 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.

Data analysis

Data was exported to statistical software packages (Excel Microsoft Windows, USA; SPSS v17, SPSS Inc, Chicago, USA) for analysis. Concurrent validity for all variables were analysed against their respective criterion measures, identifying means and standard deviations ($M \pm S$) for the data. To fully understand the data generated a range of precision of measurement statistics in combination with descriptive data has been previously reported (Bland and Altman, 1986; Brunton, et al., 2000; Hopkins, 2000; Hopkins et al., 2009).

Characteristics of the data set were considered and appropriate statistical procedures were followed thereafter. After plotting the predicted against the residuals for HR, BF, ST and P (Figure 2), data was considered to be non-uniform (i.e. heteroscedastic) so was transformed logarithmically (log) in order to provide a true interpretation (Atkinson and Nevill, 1998; Hopkins, 2000; Hopkins, et al., 2009). It was decided that descriptive data for these variables would be reported in absolute values and validity statistics presented log transformed. The combined data presentation approach was determined in order for comparison with other studies to occur, the majority of which have report absolute data.

Adopting a composite of validity statistics may provide a more informed view to assess agreement between methods (Harper-Smith et al., 2010). The following statistical analysis was calculated for each variable; Descriptive statistics including absolute mean bias and 95% Confidence Intervals, Validity statistics (log transformed) including mean bias, 95% Limits of Agreement, Pearson's Product Moment Correlation Coefficient, Coefficient of Determination as described in previous literature (Hopkins, 2000).

Within the descriptive statistics, the mean bias and associated 95% Confidence Intervals provides an indication of raw difference between the data sets. Correlation coefficients, such as Pearson's (r), provide a good indication of the relationship between data sets. Coefficient of Determination (r^2), linked to the correlation analysis, express the variance in one variable that can be attributed to the second variable (Atkinson and Nevill, 1998; Bland and Altman, 2003; Brunton, et al., 2000; Winter et al., 2001). 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. Correlation statistics should not be reported in isolation as they can be blind to bias (Bland and Altman 2003). As

noted elsewhere (Finni et al., 2007), the limits of agreement method (Bland and Altman, 1986) is used to compare the agreement between methods. Summarising the differences between the two methods is a corner stone of the process. It is expected that the differences outside of ± 2 standard deviations (S) from the mean difference are not practically important. If 95% of data are within 2S it is considered an acceptable 'limit of agreement' and methods or equipment is thought to be interchangeable (Bland and Altman, 2003). Limits of agreement cannot be used when units between two methods are not comparable hence ACC data is not analysed in this way.

Previously precision of HR and BF measurement research has removed data sets when data is clearly

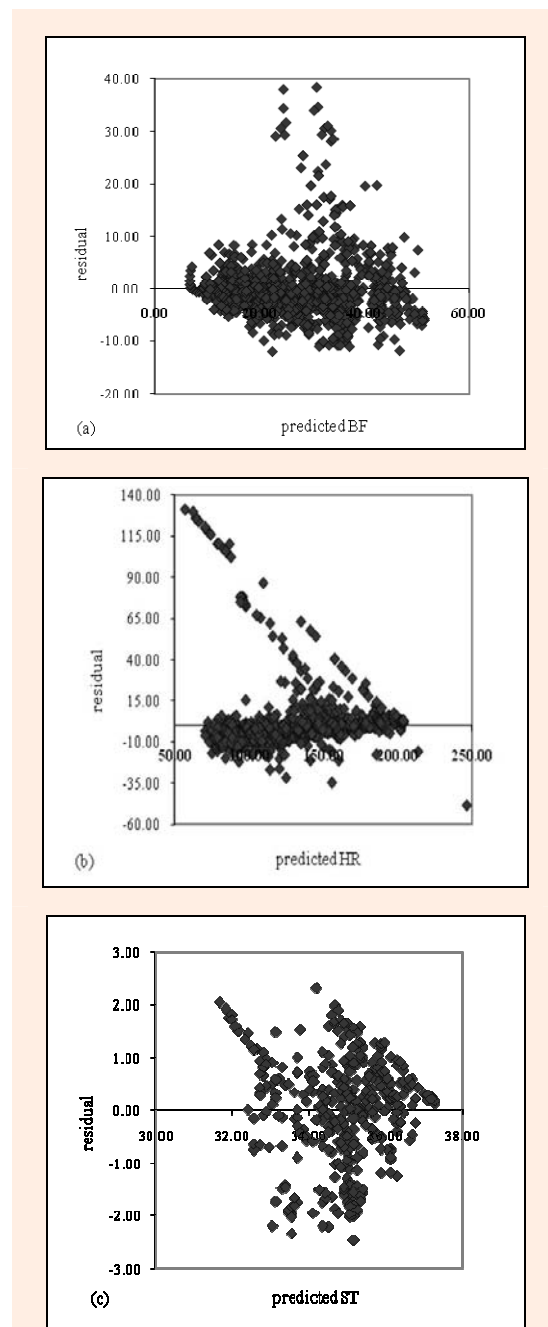


Figure 2. Residual versus predicted plot demonstrating the relationship for (a) BF, (b) HR, and (c) ST.

Table 1. Bioharness™ data in comparison to the respective criterion measure.

Variable	Descriptive Data			Validity Data (Log)		
	Predicted M ± S	Criterion M ± S	Mean Bias ±95% CI	Mean Bias ±95% LoA	PCC <i>r</i>	CoD <i>r</i> ²
HR (b.min ⁻¹)	122.6 ± 38.7	126.4 ± 39.0	-3.80 ± .93	-3.05 ± 32.20	.89 *	79%
BF (br.min ⁻¹)	24.5 ± 8.3	26.5 ± 11.9	-2.01 ± .32	-3.46 ± 43.70	.91 *	83%
ST (°C)	34.7 ± 1.4	34.9 ± 1.5	-.22 ± .10	-.61 ± 1.98	.76 *	58%
P (°)	42.4 ± 24.7	42.4 ± 24.8	.06 ± .32	.20 ± 2.62	.99 *	96%

Tabular report of validity statistics: Descriptive statistics, Standard Deviation (S), Mean Bias, 95% Confidence Intervals (CI), Log transformed mean bias, 95% Limits of Agreement (LoA), Pearson's Product Correlation Coefficient (PCC) and Coefficient of Determination (CoD) across whole data set. * $p < 0.01$

erroneous in the belief that a technical breakdown has occurred with the system (Hailstone and Kilding, 2011; Leger and Thivierge, 1988). Analysis completed, which includes erroneous data sets, would possibly reduce the practical usefulness of the results especially if the erroneous data was linked to only two or three participants. Previously, the reporting of data removal (i.e. cleaning) has been used as an additional validity statistic with high volumes of data being removed reducing the credibility of the device (Hailstone and Kilding, 2011; Leger and Thivierge, 1988; Terbizan et al., 2002). 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) 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 difference was ± 20 b.min⁻¹ for HR, ± 7 br.min⁻¹ for BF from the criterion the data set from the specific stage was removed.

Results

Overview of the validity of the Bioharness™

The results for whole data set (Table 1) indicate very strong to strong relationships for HR, BF and P ($p < 0.01$) with relatively small mean bias for each variable. In comparison, ST was less precise. Figure 4 presents a non-linear relationship for BF and HR. This aforementioned relationship starts for BF from ~ 45 br.min⁻¹ and for HR starts from ~ 175 b.min⁻¹, respectively. A very strong relationship for ACC was reported ($p < 0.01$) (Table 2) and trend lines for VMU and participants mean step

counts matched increments in intensity (Figure 3).

Table 2. Relationship of ACC data to the respective criterion measure (oxygen uptake, mL.kg⁻¹.min⁻¹) and mean step counts per stage.

	PCC (<i>r</i>)
Activity (VMU.ct.sec ⁻¹)	.97 *
Vertical peak (g.sec ⁻¹)	.95 *
Mean Step Counts (min ⁻¹)	.99 *

A tabular report of validity statistics: Pearson's Product Correlation Coefficient (PCC) for ACC Vector Magnitude Units (VMU) and Vertical peak output versus oxygen uptake, ml.min.kg⁻¹. * $p < 0.01$

Velocity specific validity results for HR

Very strong relationships ($r > 0.94$, $p < 0.01$) were noted until 10 – 12 km.h⁻¹ (Table 3). Improving precision

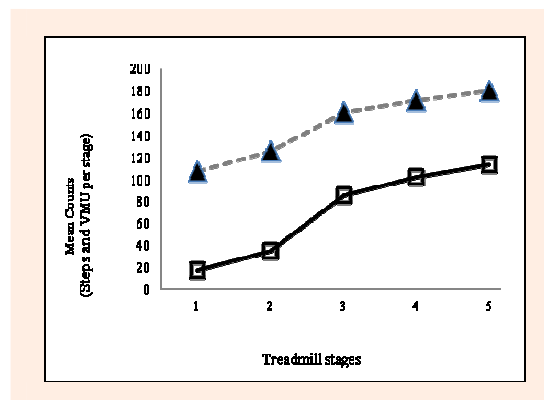


Figure 3. Trend in data between mean foot steps (- -▲- -) and mean VMU (-□-) per active stages of treadmill protocol (nb. Treadmill stages 1 to 5 refers to 4 km.h⁻¹ to 12 km.h⁻¹ respectively).

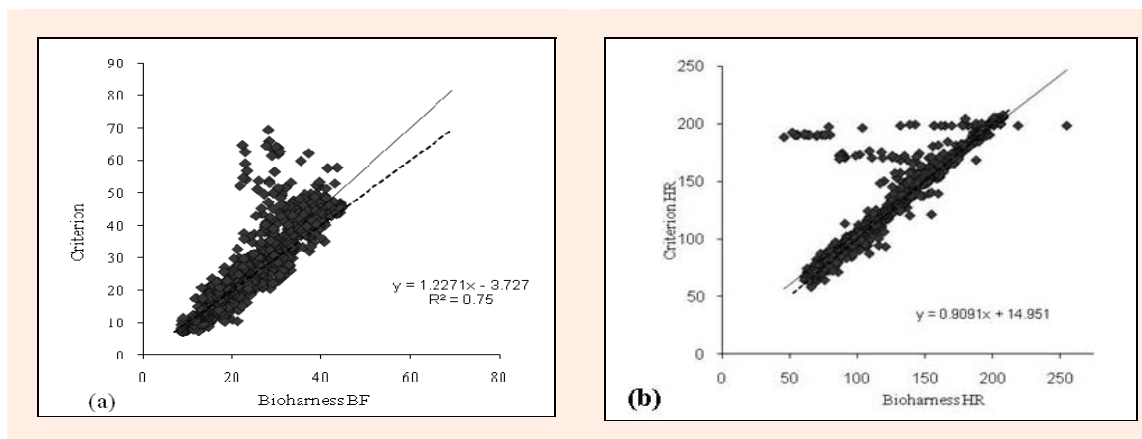


Figure 4. Scatter plot demonstrating the relationship between (a) Bioharness BF and Criterion and (b) Bioharness HR and Criterion across all velocities on treadmill. Nb. line of identity (- - -), regression line (—).

Table 3. Heart rate (b·min⁻¹) data at varying intensities.

Velocity	Descriptive Data			Validity Data (Log)		
	Predicted M ± S	Criterion M ± S	Mean Bias ±95% CI	Mean Bias ±95% LoA	PCC <i>r</i>	CoD <i>r</i> ²
0 km·h ⁻¹	81.6 ± 14.2	82.6 ± 14.3	-.98 ± .58	-1.21 ± 11.30	.94*	88%
4 km·h ⁻¹	90.9 ± 14.1	91.8 ± 13.8	-.83 ± .52	-.97 ± 9.00	.95*	90%
6 km·h ⁻¹	105.3 ± 14.0	105.0 ± 13.8	.32 ± .57	.28 ± 8.80	.94*	88%
8 km·h ⁻¹	142.2 ± 20.1	142.8 ± 19.6	-.60 ± .61	-.48 ± 7.20	.97*	94%
10 km·h ⁻¹	156.6 ± 24.9	161.0 ± 20.1	-4.44 ± 2.23	-3.32 ± 28.10	.63*	40%
12 km·h ⁻¹	160.4 ± 37.6	176.8 ± 18.4	-16.37 ± 4.66	-12.30 ± 72.90	.11	1%

Tabular report of validity statistics: Descriptive statistics, Standard Deviation (S), Mean Bias, 95% Confidence Intervals (CI), Log transformed mean bias, 95% Limits of Agreement (LoA), Pearson's Product Correlation Coefficient (PCC) and Coefficient of Determination (CoD) across whole data set. * $p < 0.01$

of measurement for HR data is seen from 0 km⁻¹, with absolute HR mean bias $< \pm 1$ b·min⁻¹ and 95% limits of agreement values reducing, until the higher velocities where accuracy of the device reduces.

Velocity specific validity results for BF

Between rest and 8 km·h⁻¹ consistent strong relationships were reported ($r > 0.81$, $p < 0.01$) with absolute mean bias remaining < 1.6 br·min⁻¹ (Table 4). Decreased precision is seen at the highest velocities with greater mean bias, weak relationships and high limits of agreement noted.

Velocity specific results for HR and BF after erroneous data removed

Erroneous data sets at the highest velocity were removed following a cleaning process described previously. Validity data was recalculated and improvement in accuracy of data is seen (Table 5). HR data for 10 km·h⁻¹ ($n = 12$) and 12 km·h⁻¹ ($n = 10$) presented very strong relationships, consistent limits of agreement and continued to underestimate HR which mirrors the data trends captured between 4 – 8 km·h⁻¹ in the raw data set. BF data for 10 and 12 km·h⁻¹ ($n = 10$) continued with similar trends seen from 4 km·h⁻¹ with strong relationships, increasing underestimation of BF (i.e. mean bias) and large but stabilising limits of agreement values

Temperature specific validity results for ST data

Results from the hot and thermo-neutral environments produced good to moderate ($r = 0.75$, $p < 0.01$) and weak ($r = 0.42$, $p < 0.01$) relationships respectively (Table 6). Mean bias was greater in hot conditions though limits of agreement were wider in thermo neutral conditions.

Validity of the ACC variable

Analysis was completed on the whole data set and illustrated a very strong relationship between VMU and rela-

tive oxygen uptake ($r = 0.97$, $p < 0.01$) (Table 7). Further relationships between relative oxygen uptake and the individual axis of the ACC are also presented with peak acceleration, vertical and lateral axis presenting strong correlations ($r > 0.84$, $p < 0.01$).

Table 7. Relationship of ACC data to the respective criterion measure (oxygen uptake mL·kg⁻¹·min⁻¹).

Accelerometer	PCC (<i>r</i>)
Activity (VMU/ct·sec ⁻¹)	.97 *
Vertical peak (g·sec ⁻¹)	.95 *
Lateral peak (g·sec ⁻¹)	.84 *

Tabular report of validity statistics: Pearson's Product Correlation Coefficient (PCC). * $p < 0.01$

Validity of the P variable

The P variable was tested using a tilt table protocol and data analysed as a whole (Table 1) with very strong relationship ($r = 0.99$, $p < 0.01$) and very small bias between measures identified.

Discussion

Main findings – validity of the Bioharness™

Multivariable physiological monitoring devices used within free living or sporting scenarios can now provide time synchronised data which may enable further insights in to day-to-day activity levels and athletic performance. Comprehensive assessment of the precision of all new measuring technology will allow for better understanding of its variability which exists and therefore allows for better interpretation of data collected (Welk, et al., 2004).

General results (Table 1 and 2) from this laboratory based study suggest that the Bioharness™ monitoring system is valid and demonstrates relatively accurate data in relation to the analysis completed. Collectively, with all data considered, the validity statistics for HR, BF, P and ACC suggest credible precision of measurement is

Table 4. Breathing frequency (br·min⁻¹) data at varying intensities.

Velocity	Descriptive Data			Validity Data (Log)		
	Predicted M ± S	Criterion M ± S	Mean Bias ±95% CI	Mean Bias ±95% LoA	PCC <i>r</i>	CoD <i>r</i> ²
0 km·h ⁻¹	15.9 ± 3.9	15.0 ± 4.5	.81 ± .35	7.31 ± 40.30	.81*	66%
4 km·h ⁻¹	18.9 ± 4.7	19.1 ± 5.8	.18 ± .39	.57 ± 35.20	.84*	71%
6 km·h ⁻¹	20.9 ± 5.9	21.0 ± 6.7	-.14 ± .48	.51 ± 42.40	.81*	66%
8 km·h ⁻¹	26.6 ± 5.8	28.1 ± 7.7	-1.57 ± .53	-4.13 ± 33.80	.83*	69%
10 km·h ⁻¹	29.5 ± 6.0	34.1 ± 8.3	-4.61 ± .99	-12.82 ± 55.80	.44*	19%
12 km·h ⁻¹	33.4 ± 5.9	40.8 ± 10.0	-7.41 ± 1.25	-16.96 ± 56.96	.35*	12%

Tabular report of validity statistics: Descriptive statistics, Standard Deviation (S), Mean Bias, 95% Confidence Intervals (CI), Log transformed mean bias, 95% Limits of Agreement (LoA), Pearson's Product Correlation Coefficient (PCC) and Coefficient of Determination (CoD) across whole data set. * $p < 0.01$

Table 5. Clean heart rate (b·min⁻¹) and breathing frequency (br·min⁻¹) data at 10 and 12 km·h⁻¹.

Velocity	Descriptive Data			Validity Data (Log)		
	Predicted M ± S	Criterion M ± S	Mean Bias ±95% CI	Mean Bias ±95% LoA	PCC <i>r</i>	CoD <i>r</i> ²
Heart rate						
10 km·h ⁻¹	159.6 ± 21.4	160.0 ± 20.6	-.45 ± .75	.40 ± 7.70	.96*	92%
12 km·h ⁻¹	174.3 ± 20.4	176.0 ± 19.1	-1.64 ± .89	-1.00 ± 7.40	.95*	90%
All data	122.2 ± 38.1	122.8 ± 38.2	-.65 ± .26	-.60 ± 8.88	.99*	96%
Breathing Fr						
10 km·h ⁻¹	30.4 ± 5.3	31.6 ± 7.0	-1.27 ± .51	-3.20 ± 23.90	.86*	75%
12 km·h ⁻¹	34.6 ± 5.4	37.8 ± 5.4	-3.14 ± .61	-7.40 ± 23.50	.86*	75%
All data	24.4 ± 8.4	25.3 ± 10.4	-.84 ± .22	-.60 ± 33.99	.94*	88%

Tabular report of validity statistics: Descriptive statistics, Standard Deviation (S), Mean Bias, 95% Confidence Intervals (CI), Log transformed mean bias, 95% Limits of Agreement (LoA), Pearson's Product Correlation Coefficient (PCC) and Coefficient of Determination (CoD) across whole data set. * $p < 0.01$

attained and limits for each variable have been established. When data is analysed at each velocity, even with a moderate/strong relationship and relatively small mean bias, HR and BF limits of agreement suggests some divergence of data at higher velocities in both absolute and log transformed values.

Velocity specific findings for heart rate (HR) and breathing frequency (BF) raw data

Velocity specific analysis for HR and BF identified differences in the precision of data. Relative to the respective criterions, there was a general trend of decreased accuracy as velocity increased (≥ 10 km·h⁻¹) which has been reported elsewhere for HR (Kingsley et al., 2004; Leger and Thivierge, 1988; Terbizan et al., 2002) and for BF (Grossman et al., 2010; Hailstone and Kilding, 2011; Witt et al., 2006).

Analysis of global HR results finds similar limits of agreement ($\sim \pm 6$ b·min⁻¹) as reported for the Polar heart rate monitor (Godsen et al., 1991) and Actiheart™ device (Soren Brage et al., 2005). HR validity data, specifically relationships, remained consistently very strong ($r > 0.94$) ≤ 8 km·h⁻¹ which would align it to the “excellent” category (Leger and Thivierge, 1988) and matches data noted in other research (Seaward et al., 1990; Wajciechowski et al., 1991). Improved accuracy of HR data from rest to 8 km·h⁻¹ could be attributed to accumulated physiological responses of exercise (i.e. perspiration/moisture) which may improve connectivity between the skin and the HR electrodes (Lopes and White, 2006; Powers and Howley, 2007). Evidence of decreasing precision of the device, specifically underestimation, with increasing velocity is further supported as the relationship of HR data becomes non-linear (Figure 4b) at ~ 175 b·min⁻¹, which corresponds to mean HR attained within the 12 km·h⁻¹ stage.

Interestingly, BF precision of measurement improves from rest to 6 km·h⁻¹, with strong relationships and

decreasing mean bias, but then the accuracy decreases rapidly through to the highest velocity. Moreover, Figure 4a suggests that the BF variable may have a threshold of accuracy at ~ 45 br·min⁻¹, which is the point where non-linear relationship in data becomes visible. A general trend of decreasing precision of measurement using similar respiratory inductive plethysmography technology has been noted elsewhere within another multivariable device (Grossman et al., 2010; Witt et al., 2006). In comparison, Hailstone and Kilding (2011) note somewhat stronger validity data for the Bioharness™ BF variable with good/strong correlations ($r > 0.86$) and absolute differences < 3.0 br·min⁻¹. A different subject specific treadmill protocol, a different data processing schedule and lack of clarity as to the version of the Bioharness™ being used limits any direct comparisons but, a general trend in agreement between these studies suggests the Bioharness™ under-estimates BF during activity. Another comparable system, the Lifeshirt™, presented stronger BF results from similarly active protocols but importantly this device uses 2 measuring bands in comparison to the Bioharness™ which uses one measuring band to assess this respiratory function (Witt et al., 2006). Two measuring bands will be able to capture both abdominal and thoracic respiratory related movements with McCool (2002) noting the more comprehensive data capture using 3 measuring bands incorporating changes in sterno-umbilical distance. Further research clarifying which thoracic landmarks influence the precision of respiratory inductive plethysmographic data may improve the accuracy of this variable, without losing the multi-functionality, unobtrusive (i.e. single measuring band) and portable nature of the device.

It is worth noting that the HR and BF non-linear scatter plot data (Figure 4) is attributed to specific participants at the highest velocities rather than a cross participant general data trend. One participant had erroneous

Table 6. Skin temperature (°C) data in hot, thermo-neutral conditions and combined data overview.

Temperature	Descriptive Data			Validity Data (Log)		
	Predicted M ± S	Criterion M ± S	Mean Bias ±95% CI	Mean Bias ±95% LoA	PCC <i>r</i>	CoD <i>r</i> ²
Hot 30°C	35.4 ± 1.0	35.9 ± 1.1	-.49 ± .10	-1.36 ± 4.14	.75*	56%
Neutral 20°C	33.8 ± 1.5	33.8 ± 1.2	.03 ± .19	.06 ± 9.12	.42*	18%
Combined	34.7 ± 1.4	34.9 ± 1.5	-.22 ± .10	-.60 ± 5.92	.75*	56%

Tabular report of validity statistics: Descriptive statistics, Standard Deviation (S), Mean Bias, 95% Confidence Intervals (CI), Log transformed mean bias, 95% Limits of Agreement (LoA), Pearson's Product Correlation Coefficient (PCC) and Coefficient of Determination (CoD) across whole data set. * $p < 0.01$

data in both variables but otherwise there was no consistency in this issue. A number of other participants presented $BF > 40 \text{ br}\cdot\text{min}^{-1}$ and $HR > 190 \text{ b}\cdot\text{min}^{-1}$ without consistent erroneous data being captured. Therefore threshold values intimated from the scatter plot should be used with caution and requires further investigation.

Velocity specific findings for heart rate (HR) and breathing frequency (BF) for cleaned data

A data cleaning process (Hailstone and Kilding, 2011; Leger and Thivierge, 1988) led to a decrease in data sets as velocity increased and can be used as evidence for credibility of a monitoring system. A total of 3 data sets (i.e. 3 participants) were removed from each variable at the higher velocities suggesting increasingly erroneous data is being captured at $\geq 10 \text{ km}\cdot\text{h}^{-1}$ a trend previously noted (Leger and Thivierge, 1988; Terbizan, et al., 2002).

Increasing errors with higher velocities in these variables can occur partly due to the data signal that the monitoring device requires becoming corrupted by movement artefacts (Cho et al., 2011; Witt et al., 2006) such as; EMG activity (Boudet and Chamoux, 2000; McArdle et al., 2009), movement of the monitoring device (Clarenbach et al., 2005; Leger and Thivierge, 1988) and specific to BF changes in the mechanics of breathing (McArdle et al., 2009; McCool et al., 2002). Additional cross technology (i.e. criterion versus predicted) data processing issues and discipline specific data handling methods could also influence the data output (Boudet and Chamoux, 2000; Kent et al., 2009).

Validity of accelerometry (ACC) variable.

The validation of the ACC variable against VO_2 ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$), which is considered an indirect criterion measure, and mean stride (step) counts per stage during the treadmill protocol, both have been noted elsewhere (McArdle et al., 2009; Rowlands et al., 2003). VMU was chosen as the main ACC unit since this is an integrated activity count providing an overall picture of activity commonly used in other research (Powell and Rowlands, 2004). Very strong relationships ($r > 0.95$) were noted for VMU suggesting the ACC demonstrates validity especially when compared against other devices which were deemed credible despite reporting weaker correlations ($r > \sim 0.80$) (Rowlands et al., 2003; G.J. Welk et al., 2004). Strong relationships and corresponding matching trend lines (Figure 3) for mean stride counts provide further confirmation for the direct validity of this variable against movement data. The ACC variable could have been compared against another peer reviewed ACC (e.g. RT3) providing an indication of data trends, however completing this analysis may have had limited results due to the lack of consensus as to how ACC counts are produced by individual devices. Assessment at different velocities in relation to VO_2 was not possible as subjects were not standardised for cardio-vascular fitness before the protocol. It is suggested that ACC could become less accurate at higher intensities due to technical limitations and as running mechanics alter (Brage et al., 2003; Powell and Rowlands, 2004) all of which should be investigated further for this device.

Validity of posture (P) variable

Due to difficulties assessing validity of the P variable against a criterion within the treadmill protocol, data was assessed against in a controlled procedure using a tilt table. Results present credible data with narrow limits of agreement (0.20 ± 2.62) and very strong relationship ($r > 0.99$) versus the criterion which mirrors other research using similar technology in the area (Bernmark and Wiktorin, 2002). The frequency of inclinometer devices using similar ACC technology is increasing with research from occupational studies being more common (Hansson, et al., 2001; 2006). Data from the P variable is generated from similar piezoelectric technology as seen within the ACC which has produced valid data within this research. The combined results associated with the ACC and P variable adds evidence to the credibility of the piezoelectric technical set up within the Bioharness™.

Validity of skin temperature (ST) variable

Infra-red ST global data set, validated against skin thermistors, initially suggests the Bioharness™ has less precision when compared to other equivalent research (Hershler et al., 1992). Moreover, Limits of agreement have not been extensively reported for infra-red temperature validity studies though the combined data note a tighter agreement when compared to previous research (Matsukawa et al., 2000). A consideration in this analysis of the Bioharness™ is that the latter two research papers were completed in a non-exercise environment which is arguably more controlled so it is expected that there is less variance in their data. Exercise adds another dimension to the validation of ST and there are few comparable data sets available in the literature. Other somewhat limited analysis reported no significant differences in infra-red temperature devices tested and strong correlations ($r > 0.95$) from a low intensity treadmill protocol incorporating an environmental chamber (Buono et al., 2007). Stronger correlations reported could be linked different methodological procedures and data analysis. For methods involving temperature measurement, a threshold of accuracy of 0.1°C has been proposed for systematic bias and $\pm 0.3^\circ\text{C}$ for 95% limits of agreement (Gant et al., 2006). These thresholds are not met by the Bioharness™ in any of the data sets collected. The weak relationships in data could possibly be explained by low number of data sets and limits of agreement analysis suggests relatively large discrepancies between the criterion and Bioharness™, especially when considering the narrow temperature data range. The equivocal results for the infra-red ST could be explained by the onset of sweating during exercise (Kistemaker et al., 2006), technical issues with the skin thermistors (Buono et al., 2007), changes in infra-red device angle to the body (Hershler et al., 1992) and distance from the skin surface (Matsukawa et al., 2000). Further examination of the ST precision of measurement should be considered.

Limitations

Reporting of absolute and/or logarithmically transformed HR and BF data relating to heteroscedasticity is highlighted within the paper. Even though absolute data is interpreted

more easily by the reader, log transformed data should be reported, if data fails to meet necessary criteria, in order for a full comprehension of the data. There is a lack of clarity as to the objective model to decide if data is heteroscedastic or not, also there are different log transformation models so further clarification on best practice should be investigated. Moreover, the sample size could be considered a limitation of the study though numbers of subjects in this research matches or even exceed other peer reviewed papers dealing with similar themes. Additionally, controlling for the participants sex added some further rigor to the research design, though in theory leaves 50% of the population untested on the Bioharness™ device. Further investigation of this device should be completed on wider population groups to fully understand its capacities.

Conclusion

The results suggest that, with prior understanding of data limitations, the Bioharness™ (Version 1) has proved to be a valid multivariable monitoring device within ambulatory laboratory testing. ACC and P variables presented strong data which relates to the advanced piezoelectric technology used. Using the device to capture HR and BF data during high intensity activities should be completed with the understanding that the validity of this data could be influenced by artefact at treadmill velocities of $\geq 10 \text{ km}\cdot\text{h}^{-1}$. Research on similar HR and BF devices report decreasing accuracy at higher activity levels therefore establishing a transparent data cleaning procedure should be considered or future technological development should amend data capture boundaries. Further development of infra-red ST technology within the device should be considered. In summary, having established the Bioharness™ is a valid multivariable monitoring device within ambulatory laboratory testing. It is suggested the next progression will be to assess the reliability of the multivariable Bioharness™ monitoring system in a laboratory environment.

Acknowledgement

This research was supported with Bioharness™ 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.
- Bland, J.M. and Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**(8476), 307 - 310.
- Bland, J.M. and Altman, D.G. (2003) Applying the right statistics: analysis of measurement studies. *Ultrasound in Obstetric and Gynecology* **22**, 85-93.
- 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.
- Brunton, A., Conway, J. H. and Holgate, S.T. (2000) Reliability: what is it, and how is it measured? *Physiotherapy* **86**(2), 94 - 99.
- Buono, M.J., Jechort, A., Marques, R., Smith, C. and Welch, J. (2007) Comparison of infrared versus contact thermometry for measuring skin temperature during exercise in the heat. *Physiological Measurement* **28**, 855-859.
- Burnham, R.S., McKinley, R.S. and Vincent, D.D. (2006) Three Types of Skin-Surface Thermometers. *American Journal of Physical Medicine & Rehabilitation* **85**(7), 553-558.
- 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 Physicians* **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.
- Daneshmandi, H., Rahmaninia, F., Shahrokh, H., Rahmani, P. and Esmaeili, S. (2010) Shoulder joint flexibility in top athletes. *J. Biomedical Science and Engineering* **3**, 811 - 815.
- Finni, T., Hu, M., Kettunen, P., Vilavou, T. and Cheng, S. (2007) Measurement of EMG activity with textile electrodes embedded in to clothing. *Physiological Measurement* **28**, 1405-1419.
- 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.
- Godsen, R., Carroll, T. and Stone, S. (1991) How well does Polar Vantage XL heart rate monitor estimate actual heart rate? *Medicine and Science in Sport and Exercise* **23**(4), S14.
- 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 Computers* **39**, 405-413.
- Harper-Smith, A.D., Crabtree, D.R., Bilzon, J.L.J. and Walsh, N.P. (2010) The validity of wireless ibuttons and thermistors for human skin temperature measurement. *Physiological Measurement* **31**, 95-114.
- Hershler, C., Conine, T.A., Nunn, A. and Hannay, M. (1992) Assessment 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. (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. (2000) 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.
- Kingsley, M., Lewis, M.J. and Marson, R.E. (2004) Comparison of Polar 810s and an Ambulatory ECG System for RR interval measurement during progressive exercise. *International Journal of Sports Medicine* **26**, 39-44.
- Kistemaker, J.A., Den Hartog, E.A. and Daanen, H.A.M. (2006) Reliability of an infrared forehead skin thermometer for core temperature measurements during progressive exercise. *Journal of Medical Engineering and Technology* **30(4)**, 252-261.
- Leger, L. and Thivierge, M. (1988) Heart rate monitors: Validity, stability and functionality. *The Physician and Sports Medicine* **16(5)**, 143-151.
- Lopes, P. and White, J. (2006) Heart rate variability: Measurement methods and practical implications. In: *Physiological Assessment of Human Fitness*. Eds: Mauds, P.J. and Foster, C. 2nd edition. Champaign, IL: Human Kinetics.
- 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.
- Matsukawa, T., Ozaki, M., Nishiyama, T., Imanura, M. and Kumazawa, T. (2000) Comparison of 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. Philadelphia, 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.
- 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.
- Powers, S.K. and Howley, E.T. (2007) *Exercise physiology: Theory and application to fitness and performance*. 7th edition. New York: McGraw-Hill Intl.
- 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.
- Seaward, B.L., Sleamaker, R.H., McAuliffe, T. and Clapp, J.F. (1990) The precision and accuracy of a portable heart rate monitor. *Biomedical Instrumentation & Technology* **24**, 37-41.
- 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.
- Thomas, J.R., Nelson, J.K. and Silverman, S. (2005) *Research methods in physical activity*. 5th edition. Champaign, IL: Human Kinetics.
- Wajciechowski, J., Gayle, R., Andrews, R. Dintiman, G. (1991) The accuracy of radio telemetry heart rate monitor during exercise. *Clinical Kinesiology* **45**, 9-12.
- Welk, G.J. (2005) Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine & Science in Sport & Exercise* **37(11)**, 501-511.
- Welk, G.J., Schaben, J.A. and Morrow, J.R. (2004) Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine & Science in Sport & Exercise* **36(9)**, 1637-1645.
- Winter, E.M., Eston, R.G. and Lamb, K.L. (2001) Statistical analyses in the physiology of exercise and kinanthropometry. *Journal of Sports Sciences* **19**, 761 - 775.
- 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.

Key points

- Different levels of precision exist for each variable in the Bioharness™ (Version 1) multi-variable monitoring device
- Accelerometry and posture variables presented the most precise data
- Data from the heart rate and breathing frequency variable decrease in precision at velocities ≥ 10 km·h⁻¹
- Clear understanding of the limitations of new applied monitoring technology is required before it is used by the exercise scientist

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 of Sport, Health and Exercise Science, University of Hull, UK.

Degree

PhD

Research interest

Environmental physiology

✉ James A. Johnstone

Sport, Health and Exercise Science Research Group, School of Life Sciences, University of Hertfordshire, AL10 9AB, UK.