©Journal of Sports Science and Medicine (2009) **8**, 528-532
http://www.jssm.org

Research article

Familiarisation and reliability of sprint test indices during laboratory and field assessment

James G. Hopker [1]✉, **Damian A. Coleman** [2], **Jonathan D. Wiles** [2] **and Andrew Galbraith** [1]
[1] Centre for Sports Studies, University of Kent, Chatham Maritime, Chatham, Kent, UK, [2] Department of Sport Science, Tourism and Leisure, Canterbury Christ Church University, Canterbury, Kent, UK

Abstract
The aim of the study was to assess the reliability of sprint performance in both field and laboratory conditions. Twenty-one male (mean ± *s*: 19 ± 1 years, 1.79 ± 0.07 m, 77.6 ± 7.1 kg) and seventeen female team sport players (mean ± *s*: 21 ± 4 years, 1.68 ± 0.07 m, 62.7 ± 4.7 kg) performed a maximal 20-metre sprint running test on eight separate occasions. Four trials were conducted on a non-motorised treadmill in the laboratory; the other four were conducted outdoors on a hard-court training surface with time recorded by single-beam photocells. Trials were conducted in random order with no familiarisation prior to testing. There was a significant difference between times recorded during outdoor field trials (OFT) and indoor laboratory trials (ILT) using a non-motorised treadmill (3.47 ± 0.53 *vs*. 6.06 ±1.17s; p < 0.001). The coefficient of variation (CV) for time was 2.55-4.22% for OFT and 5.1-7.2% for ILT. During ILT peak force (420.9±87.7N), mean force (147.2±24.7N), peak power (1376.8 ± 451.9W) and mean power (514.8 ± 164.4W), and were measured. The CV for all ILT variables was highest during trial 1-2 comparison. The CV (95% confidence interval) for the trial 3-4 comparison yielded: 9.4% (7.7-12.1%), 7.9% (6.4-10.2%), 10.1% (8.2-13.1%) and 6.2% (5.1-8.0%) for PF, MF, PP and MP and respectively. The results indicate that reliable data can be derived for single maximal sprint measures, using fixed distance protocols. However, significant differences in time/speed over 20-m exist between field and laboratory conditions. This is primarily due to the frictional resistance in the non-motorised treadmill. Measures of force and power during ILT require at least 3 familiarisations to reduce variability in test scores.

Key words: Non-motorised treadmill, force, power, familiarisation, sprint running.

Introduction

Sprint performance is a key component of many sporting disciplines. This component of fitness is often assessed by the sport scientist to indicate athletic ability (Bird and Davidson, 1997; Logan et al., 2000), to monitor changes in training status (Linossier et al., 1993; MacDougall et al., 1998; McManus et al., 1997), and to ascertain the effect of ergogenic supplementation on performance (Bell et al., 2001; Collomp et al., 1991; Green et al., 2001). To ensure confidence in interpreting data from new equipment or protocols, the random variation of a subject's repeated measurement must be evaluated (Coleman et al., 2005) rather than accepting the manufacturer claims (Davison et al. 2000). This data allows the exercise scientist to select appropriate tests or tools (Atkinson and Nevill, 1998) that have adequate precision. In turn, this will allow the detection of small, nevertheless, worthwhile changes in performance.

Many research laboratories use the Wingate cycle ergometer test to assess peak power and anaerobic capacity in athletes, regardless of their sporting discipline. The argument against using the Wingate cycle ergometer test is that it is not specific to running. Generally, non-specific testing also requires greater familiarisation before reliable results can be attained (Lakomy, 1987). The development of non-motorised treadmills enables the assessment of running capability in a stable laboratory environment. These devices have been developed to assess power and force as well as speed/time. However, the problem of familiarisation to sport-specific ergometry is still apparent. Therefore, ascertaining how many trials are required to collect reliable data is very useful information for the sport scientist. Such data can then be used when designing test protocols using non-motorized treadmills. Capriotti et al. (1999) indicated that there is minimal influence of a learning effect following two familiarisation trials during sprint cycling. However, there is currently no published data available for the length of familiarisation required during non-motorised treadmill sprinting. Tong et al. (2001) report reliability coefficients following familiarisation to non-motorised treadmill repeated sprint assessments. Although without full details of familiarisation procedures included in the paper, this data is somewhat limited. Familiarisation information is very useful to the sport scientist in designing experiments to ensure robust and reliable data is collected.

When collecting data in the laboratory on sports participants the true test of its value is by comparing scores with field-based data. To our knowledge there has been no comparison of the reliability of maximal sprint performance during field and laboratory assessment. Therefore, the aims of this investigation were; (1) to assess the reliability of sprint test indices during laboratory and field-based sprinting and; (2) to assess any changes that occur with familiarisation.

Methods

Twenty-one male (mean ± *s*: 19 ± 1 years, 1.79 ± 0.07 m, 77.6 ± 7.1 kg) and seventeen female team sport players (mean ± *s*: 21 ± 4 years, 1.68 ± 0.07 m, 62.7 ± 4.7 kg) participated in this study. These athletes were of mixed ability, ranging from collegiate to international standard in team sports. All experimental procedures were ap-

Received: 27 April 2009 / Accepted: 11 August 2009 / Published (online): 01 December 2009

proved by the University Ethics Committee, and written informed consent was obtained from all participants.

Protocol
Prior to all testing, the procedures were described and a demonstration of the testing protocol, using a video was shown to all participants. The aim of using a video demonstration was to give participants an idea of the testing protocol without having them set foot onto the treadmill. Participants then attended eight data collection sessions within a four-week experimental period. These consisted of four indoor sprint trials and four outdoor sprint trials over a distance of 20-metres. This was selected as a typical sprint distance in team sports (Barros et al., 1999). All experimental trials took place in a random order design (www.random.org). Participants arrived at the laboratory on each occasion refraining from exercise in the previous 24-hour period. Otherwise participants maintained their normal training routine and dietary practice, during the experimental period. The trials were performed at the same time of day to minimise the effect of diurnal variations.

A standardised warm up was administered prior to each testing session regardless of whether the participants were conducting an outdoor field trial (OFT) or indoor laboratory trial (ILT). The warm up took place outdoors and consisted of 8 minutes light running, stretching and five sprints (accelerations over 10-m). Following this, participants then completed a single maximal sprint trial (20-m). Participants were taken to a hard court surface for an OFT or escorted to the laboratory for an ILT.

Outdoor field trials
The OFT consisted of participants sprinting 20-m from a short rolling (walking) start (3-m). The purpose of this was to allow for a comparable sprint effort to that of ILT. Due to the frictional resistance within the non-motorised treadmill it is not possible to produce a maximal sprint effort from a standing start. The trial time was recorded using single-beam photocells (in house production). The participant was counted down to the start of the outdoor trial to replicate the starting conditions in the ILT. Outdoor Field Trials took place on a tennis court in dry, still conditions.

Indoor laboratory trials
The ILT took place on a commercially available non-motorised treadmill (Force 3.0, Woodway, Waukesha, WI), with a running surface 55 x 150cm in size. Participants were fitted with a waist belt which in turn was connected to a horizontal strain gauge in a similar manner to that described Tong et al. (2001). Participants were then given 20 seconds to walk on the treadmill and gradually accelerate it to 75% of maximum speed. This was designed to allow the participants to feel the amount of frictional resistance within the treadmill belt. Participants were then instructed to walk slowly to mimic the OFT starting protocol and were counted down into the trial. Time, peak force, mean force, peak power and mean power were recorded during the ILT. Strong verbal encouragement was given during all trials of this study.

Prior to each testing session the treadmill was calibrated using a measured mass of 100 kg (Kent Scientific Services, West Malling, England) and the distance was checked using a measuring wheel (Trumeter, Manchester, UK). Power output was calculated as the product of force and speed for mean and instantaneous readings during treadmill sprinting. All data collected using the non-motorised treadmill was recorded at 100 Hz.

Statistical analyses
Prior to all analysis, data was checked for the assumptions using parametric tests (Field, 2000). To derive the within-subject variation expressed as a coefficient of variation (CV) all data was log-transformed, change scores were then calculated (e.g. trial 2 - trial 1) for each participant. The standard deviation of the change scores for the sample was then divided by root2 in order to calculate the CV. Where reliability coefficients fell within the 95% confidence interval of the previous trial comparison a single CV was derived by log-transformed two-way analyses of variance (Atkinson and Nevill, 1998). Where this procedure was adopted for e.g. sprint time, it was assumed that the reliability coefficients reflected the error and variability of familiarised ILT and OFT assessment.

Intraclass reliability coefficients were also calculated for all parameters. Statistical difference between trials ($p < 0.05$) was assessed using a one-way repeated measures ANOVA. Where data did not meet parametric assumptions (specifically the criteria for normal distribution) the data was transformed, and if this process still did not satisfy the criteria a non-parametric test was used. Data analyses were conducted using SPSS (SPSS Version 14, Chicago, IL). All values were expressed as mean and standard deviation (mean ± *s*) unless otherwise stated.

Results

Preliminary data analysis indicated parametric assumptions were met for all parameters with the exception of the mean time to complete the sprint trials. These data were analysed using a Wilcoxon Signed Ranks test. The mean times recorded during the sprint trials were 3.47 ± 0.53 and 6.06 ± 1.17s for OFT and ILT sprints respectively ($p < 0.01$); with no significant change in time over the four trials for OFT (Table 1). However, during the ILT there was a significant reduction in time taken to complete the sprints from trial 1 to 2 (6.67 to 5.86s; $p < 0.01$), there were no further reductions in sprint time from trial 2-4 ($p = 0.91$).

The coefficient of variation (CV) for all measured variables can be seen in Table 2. With the exception of mean force, all parameters demonstrated a reduction in the CV across the trial comparisons. For the comparison of trials 3-4, peak power demonstrated a reduction in CV outside the 95% confidence interval for trials 2-3.

The average peak force recorded during ILT was 420.9 ± 87.7 N. There was no difference between any trials for peak force. For the consecutive trial comparisons, intraclass reliability coefficient was $r = 0.47$, $r = 0.56$, and $r = 0.74$ respectively. The mean force recorded during the indoor treadmill sprint trials was 147.2 ± 24.7 N. There was no difference between consecutive trials

Table 1. Mean (±s) values for the parameters measured during treadmill and field sprinting.

	1	2	3	4
OFT time (s)	3.5 (0.5)	3.5 (0.5)	3.48 (0.5)	3.47 (0.5)
ILT time (s)	6.7 (1.4) *	5.9 (1.0)	5.9 (1.0)	5.8 (1.0)
Peak Force (N)	415.6 (109.8)	414.6 (78.7)	415.5 (69.9)	467.8 (88.5)
Peak Power (W)	1288 (387) §	1310 (536)	1448 (429)	1464 (433)
Mean Force (N)	140.3 (24.6)	146.2 (25.4)	149.8 (23.9)	152.4 (24.0)
Mean Power (W)	475.4 (159.8) *	523.2 (172.9)	530.4 (167.3)	530.4 (157.6)

* denotes a significant difference between trial 1 and subsequent trials ($p < 0.05$). § denotes significantly lower than trials 3 and 4 ($p < 0.05$).

(p = 0.13, p = 0.25, and p = 0.99), with intraclass r ranging from 0.79-0.91; however, there was a significant increase in mean force from trial 1-3 (9.5 ± 17.4 N, p = 0.01) and trials 1-4 (12.1 ± 18.0 N, p < 0.01).

The mean power increased significantly (~2%) during ILT from trials 1-2, (47.8 ± 71.7 W, p = 0.01). Trial 1 also produced significantly lower power compared with trials 3 (-55 ± 76.5 W, p < 0.01) and 4 (-54.9 ± 80.4 W, p < 0.01), although there were no significant differences between other trials. Intraclass r ranged from 0.83-0.93 for the three consecutive trial comparisons.

The average peak power output recorded was 1376.8 ± 451.9 W. Trial 1 was significantly lower (~10%) than trials 3 (+159.9 ± 223.2 W, p < 0.01) and 4 (+177.6 ± 209.5 W, p < 0.01). This equated to an increase of ~14%. There were no other significant differences between trials (p > 0.05). Intraclass r for peak power ranged from 0.54-0.83.

Discussion

The aims of this investigation were to compare the reliability of laboratory and field based sprinting and to assess any changes that occur with familiarisation during the two protocols. To enable the comparison of ILT and OFT a fixed distance protocol was used. However, due to the frictional resistance of the non-motorised treadmill belt it took participants considerably longer to complete the fixed distance in the laboratory. This should be considered when using the non-motorised treadmill to formulate protocols, since fixed distance work is often stipulated by guidelines on the assessment of athletes (Gore, 2000). If the non-motorised treadmill is used, distances would need to be adjusted to reflect the predominant energy systems that the investigator intended to stress. Similarly, if a non-motorised treadmill is used to assess sprint athletes (e.g. 100-m sprinters), it would be more appropriate to fix the duration of work rather than replicate the competitive distance. Even though fixed duration assessment may not be specific to field races, the measurement of performance in this manner compares favourably to fixed distance work (Hopkins et al., 2001).

This study did not use any familiarisation trials prior to testing. The protocol was specifically formulated in this way to ascertain the learning effect associated with each measured parameter. Subsequent reliability analysis enabled us to quantify the changes in performance between repeated trials. Hopkins et al. (2001) provided an analysis of 17 studies which investigated power output using 3 or more trials. They suggest that the CV of the first two trials is 1.3 times that of the CV between subsequent trials. When compared to studies using only 2 trials the mean increase in performance was less than that suggested by the 3 trial experimental data. Hopkins et al. (2001) suggest this indicates that unreported familiarisation trails were undertaken prior to data collection where only 2 trials have been reported.

As the data presented in this current study is without familiarisation, it should allow experimenters using similar equipment to ascertain the number of trials to reduce variability in the data. Table 2 supports the analysis of Hopkins et al. (2001) by showing that all parameters in the current study demonstrated a reduction in CV when calculated between consecutive trials. The 95% confidence intervals indicate that in all instances the change in CV is likely to be real. Practice effects have been reported using the Wingate anaerobic cycling test (Barfield et al., 2002), with peak and mean power significantly increasing between two trials (~13% and ~5% respectively). In the current study the changes were 10% and ~2% for these two parameters from trials 1-3. Mean power recorded during trial 2 was significantly higher than trial 1 (p < 0.05). From trials 1-4 peak power increased significantly (~14%) indicating that more than one administration of the test is required to determine baseline measures.

The ranges of the CV's are similar to those recorded previously (Tong et al. 2001) despite differences

Table 2. CV (95% confidence interval) for all parameters measured.

	Trial 1-2	Trial 2-3	Trial 3-4	Trials 2-4
OFT time (s)	4.22 (3.4-5.6)			3.1 (2.7-3.7)
ILT time (s)	7.2 (5.9-9.3) *			5.5 (4.7-6.6)
Peak Force (N)	17.8 (14.5-23.0)	12.2 (10.0-15.8)	9.4 (7.7-12.1)	
Peak Power (W)	20.7 (16.9-26.8)	21.1 (17.2-27.3)	10.1 (8.2-13.1)	
Mean Force (N)	8.1 (6.6-10.5)	5.9 (4.8-7.6)	7.9 (6.4-10.2)	
Mean Power (W)	10.9 (8.9-14.1) *			7.4 (6.4-8.8)

* denotes a significant difference in the bias between consecutive trials ($p < 0.05$). Where a single figure appears in the Trials 2-4 column the CV between consecutive trials fell within the 95% confidence interval of the previous comparison so a single CV has been calculated. Where this is not the case trials 2-3 and 3-4 have been presented.

in the protocols used to collect the data (single vs. multiple sprints per session). Tong et al. (2001) used 4 x 6 second sprints and presented the maximum values from these four trials, with comparisons over two days of assessment. It appears that for time (or speed) their protocol derived a substantially lower CV (1.3 vs. 5.1%) than the one used in the current study. Despite these differences for time/speed, the values presented for peak and mean force by Tong et al. (2001) fall within or are above our study's 95% confidence intervals (using the trial 3-4 comparisons in Table 2). These similarities in variability for force and power occur despite the absolute values being lower in the current study.

There are substantial time/speed differences between the current study and the work of previous investigators, using a non-motorised treadmill (3.4 vs. 5.1-7.1 $m \cdot s^{-1}$) (Tong et al. 2001; Hamilton et al., 1991; Cheetham et al., 1985). This could be due to differences in commercially available treadmills between studies, or this factor, in combination with differences in the training status of the participants, and the walking (start) protocol used in the current study. In comparison during OFT participants recorded a mean speed of 5.8 $m \cdot s^{-1}$.

Atkinson and Nevill (2001) specify that in reliability experiments researchers should specify how reliability analysis influences the interpretation of individual responses. This is particularly pertinent to the sport scientist using non-motorised treadmill protocols for scientific support services with an individual athlete. Atkinson and Nevill (2001) reported that the International Standards Organisation (ISO) advocate using the 95% Limits of Agreement to indicate the acceptable boundaries that are represented by measurement error. If changes in e.g. sprint times are outside these limits, then they are likely to be real (Atkinson and Nevill, 2001). Hopkins (2000) suggested using 95% Limits of Agreement was too stringent. Instead he suggested that for elite athletes smaller changes are probably detectable. Thus, based on the calculations of Hopkins (2000), changes of >3.51% (±0.12 s) and >7.07% (±0.42 s) would be required to be certain that an individual's time had decreased for fixed distance sprints during OFT and IFT respectively. The changes required for peak and mean force; peak and mean power would be +57 and +16.7 N; +204.8 and +45.6 W respectively.

Conclusion

To our knowledge this is the first study to compare the reliability of OFT and ILT using a non-motorised treadmill. The findings of the study indicate that reliable data can be derived from both indoor and outdoor fixed distance protocols using a single maximal sprint measurement. Consistent data can be obtained after the second trial. However, a significant difference in time taken to complete the trials may exist between field and laboratory. This should be considered when using a non-motorised treadmill to test athletes. Measures of peak and mean force; peak and mean power during indoor trials demonstrate poorer reliability. Therefore, multiple trials may be required to reduce variability in test scores.

References

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

Atkinson, G. and Nevill, A.M. (2001) Selected issues in the design and analysis of sport performance research. *Journal of Sports Sciences* **19**, 811-827.

Barfield, J.P., Sells, P.D., Rowe, D.A. and Hannigan-Downs, K. (2002) Practice effect of the Wingate anaerobic test. *Journal of Strength and Conditioning Research* **16**, 472-473.

Barros, T., Valquer, W. and Sant'Anna, M. (1999) High intensity motion pattern analysis of Brazilian elite soccer players in different positional roles. *Medicine and Science in Sports and Exercise* **31(Supplement),** 260.

Bell, D.G., Jacobs, I. and Ellerington, K. (2001) Effect of caffeine and epinephrine ingestion on anaerobic exercise performance. *Medicine and Science in Sport and Exercise* **33**, 1399-1403.

Bird, S.R. and Davison, R.C.R. (1997) *Physiological testing guidelines.* 3rd edition. British Association of Sport and Exercise Sciences. Leeds, UK.

Capriotti, P.V., Sherman, M. and Lamb, D. (1999) Reliability of power output during intermittent high intensity cycling. *Medicine and Science in Sport and Exercise* **31**, 913-915.

Cheetham, M.E., Williams, C. and Lakomy, H.K. (1985) A laboratory running test; metabolic responses of sprint and endurance trained athletes. *British Journal of Sports Medicine* **19**, 81-84.

Coleman D.A., Wiles, J.D., Nunn, M. and Smith, M.F. (2005) Reliability of sprint test indices in well-trained cyclists. *International Journal of Sports Medicine* **26**, 383-387.

Collomp, K., Ahmaidi, S., Audran, M., Chanal, J.L. and Prefaut, C. (1991) Effects of caffeine ingestion on performance and anaerobic metabolism during the Wingate test. *International Journal of Sports Medicine* **12**, 439-443.

Davison, R.C.R., Coleman, D.A., Balmer J., Nunn, M, Theakston, S.C., Burrows, M. and Bird, S.R. (2000) Assessment of Blood Lactate: practical evaluation of the Biosen 5030 lactate analyzer. *Medicine and Science in Sport and Exercise* **32**, 243-247.

Field, A. (2000) *Discovering Statistics: Using SPSS for windows*. Sage, UK.

Green, J.M., McLester, J.R., Smith, J.E. and Mansfield, E.R. (2001) The effects of creatine supplementation on repeated upper-and lower-body Wingate performance. *Journal of Strength and Conditioning Research* **15**, 36-41.

Gore C.J. (2000) *Physiological tests for elite athletes*. Human Kinetics, Illinois.

Hamilton, A.L., Nevill, M.E., Brooks, S. and Williams, C. (1991) Physiological responses to maximal intermittent exercise: differences between endurance-trained runners and games players. *Journal of Sports Sciences* **9**, 371-382.

Hopkins, W.G. (2000) Measures of reliability in sports medicine and science. *Sports Medicine* **30**, 1-15.

Hopkins, W.G., Schabort, E.J., Hawley, J.A. (2001) Reliability of power in physical performance tests. *Sports Medicine* **31**, 211-234.

Lakomy (1987) The use of a non-motorised treadmill for analysing sprint performance. *Ergonomics* **30**, 627-637.

Linossier, M.T., Denis, C., Dormois, D., Geyssant, A. and Lacour, J.R. (1993) Ergometric and metabolic adaptations to a 5-s sprint training programme. *European Journal of Applied Physiology* **67**, 408-414.

Logan, P., Fornasiero, D., Abernethy, P. and Lynch, K. (2000) Protocols for the assessment of isoinertial strength. In: *Physiological tests for elite athletes.* Ed: Gore, C.J. Human Kinetics, Illinois. 220-222.

MacDougall, J.D., Hicks, A.L., MacDonald, J.R., McKelvie, R.S., Green, H.J. and Smith, K.M. (1998) Muscle performance and enzymatic adaptations to sprint interval training. *Journal of Applied Physiology* **84**, 2138-2142.

McManus, A.M., Armstrong, N. and Williams, C.A. (1997) Effect of training on the aerobic power and anaerobic performance of prepubertal girls. *Acta Paediatrica* **86**, 456-459.

Tong, R.J., Bell, W., Ball, G. and Winter, E.M. (2001) Reliability of power output measurements during repeated treadmill sprinting in rugby players. *Journal of Sports Sciences* **19**, 289-297.

Key points

- Reliable data can be derived from single maximal sprint measures in both indoor and outdoor environments using fixed distance protocols.
- There may be significant time differences to complete fixed distance trials between the two environments.
- Measures of mean force, peak force and peak power during indoor trials may require multiple trials to reduce variability in test scores.

AUTHORS BIOGRAPHY

James G. HOPKER

Employment

Lecturer, Centre for Sports Studies at the University of Kent, UK.

Degree

PhD

Research interests

Exercise physiology, the effect of training and training status on exercise efficiency

E-mail: j.g.hopker@kent.ac.uk

Damian A. COLEMAN

Employment

Principal Lecturer, Department of Sport Science, Leisure and Tourism at Canterbury Christ Church University, Kent, UK.

Degree

PhD

Research interests

Physiological assessment of athletes, energetics of competitive sport and the physiological evaluations of time-trial performance.

E-mail: damian.coleman@canterbury.ac.uk

Jonathan D. WILES

Employment

Senior Lecturer, Department of Sport Science, Leisure and Tourism at Canterbury Christ Church University, Kent, UK.

Degree

PhD

Research interests

Exercise physiology, isometric training and blood pressure responses to training.

E-mail: jim.wiles@canterbury.ac.uk

Andrew GALBRAITH

Employment

Technician, Centre for Sports Studies at the University of Kent, UK.

Degree

MSc

Research interests

Critical velocity and performance prediction in running.

E-mail: a.galbraith@kent.ac.uk

✉ **James G. Hopker**

Centre for Sports Studies, University of Kent, Chatham Maritime, Chatham, Kent, ME4 4AG, United Kingdom