Article title: Reliability of 5km running performance in a competitive environment

Running head: Reliability of a competitive 5 km time-trial

Abstract

The aim of this study was to examine the reliability of a 5 km time-trial during a competitive outdoor running event. Fifteen endurance runners (age = 29.5 ± 4.3 years, height = 1.75 ± 0.08 m, body mass = 71.0 ± 7.1 kg, 5 km lifetime personal best = 19:13 ± 1:13 minutes) completed two competitive, 5 km time-trials over two weeks. No systematic differences in run time between Trial 1 and Trial 2 were reported (Trial 1; 1217 ± 85 s, 95% CI [1170, 1264] and Trial 2; 1216 ± 79 s, [1172 to 1260], p =.855). Absolute reliability, expressed as the typical error (TE; 14.7s, 95% CI = 11.3 to 21.4 s) and coefficient of variation (CV; 0.95 ± 0.65%, [0.59 to 1.31]) confirms the reliability of 5 km running performance in a competitive time trial.

Key words

Ecological validity, reproducibility, time-trial, competition, performance

Researchers investigating the efficacy of an intervention must use a test that has high reliability (Currell & Jeukendrup, 2008). For this reason, numerous studies have examined the reliability of running time-trial performance in the laboratory. Laursen, Francis, Abbiss, Newton and Nosaka (2007) reported coefficient of variations (CV) of 3.3 and 2.0%, for 1500 m and 5 km running trials on a motorised treadmill respectively and Stevens et al. (2015) reported a similar CV of 1.2% during 5 km running on a non-motorised treadmill. The low CV for these measurements provides the researcher with confidence that any observed change in performance is attributed to the intervention and not to other extraneous variables (e.g. measurement error and inter-individual variation). However, the ecological validity of these performance measures are questionable, as performance tests conducted within the controlled laboratory environment are artificial and may not provide a true reflection of real-world outdoor events. If a performance measurement fails to adequately represent the target environment, then scientific experimental outcomes may not translate into practice and may lack true relevance and impact (Araújo & Davids, 2009).

The differences in performance between artificial (e.g. laboratory) and natural (e.g. outdoor) environments has been extensively investigated. Higher running velocities have been reported during field-based running at fixed blood lactate concentrations in comparison to laboratory based trials (Kunduracioglu, Guner, Ulkar, & Erdogan, 2007), and higher blood lactate concentrations have been reported during treadmill running compared to running on synthetic surfaces (Di Michele, Di Renzo, Ammazzalorso, & Merni, 2009). Others have reported different energetic/metabolic costs (Jones & Doust, 1996) and biomechanical differences (Ali, Caine, & Snow, 2007) between treadmill and outdoor environments. While more recent laboratory investigations have attempted to replicate the outdoor environment with the use of non-motorised treadmills (Stevens et al., 2015). Although authors have attempted to stimulate outdoor running performance in the laboratory with specialised equipment, the use of such protocols have shown differences in performance of 22% (Stevens et al., 2015) and suggest that the use of actual outdoor time-trials may be a more pragmatic and cheaper alternative. Furthermore, and from a psychological perspective, Terry, Karageorghis, Saha, and D'Auria (2012) proposed that the lack of visual stimulation within the laboratory environment may increase the tedium of the task compared to the outdoor external environment, whereas McAuley, Mihalko, and Bane (1997) purported the unfamiliarity and perceived threat of the laboratory environment, and/or testing equipment, may negatively influence anxiety and arousal. It is possible that these factors may negatively influence levels of athlete motivation, effort and perceived exertion, which may consequently influence performance and the inferences that can be made from interventions using these protocols.

Collectively, current research suggests that performance measured in the laboratory may not be an adequate representation of actual performance. Some studies have therefore investigated the reliability of time-trials outdoors. Hodges, Hancock, Currell, Hamilton, and Bruce (2006) and O'Rourke, Obrien, Knez, and Paton (2007) measured the reliability of 1500 m and 5000 m running and reported CVs of 0.8 and 1.4%, respectively. These results are similar, if not better, to the equivalent running time-trials performed in the laboratory (e.g. Laursen et al., 2007) and are more representative of actual running performance. However, the studies highlighted above did not investigate the effects of direct competition during performance and, like the laboratory protocols, may lack ecological validity. The effect of competition can have a significant impact on the physiology of the athlete and subsequently the performance. Pierce, Kuppart, and Hardy (1976) reported that adrenaline is significantly higher during competitions in comparisons to training for basketball and track and field athletes, whereas Viru et al. (2010) reported differences in the peak oxygen consumption ($\dot{V}O_{2peak}$) and performance between competitive and non-competitive situations in treadmill running. These studies highlight the physiological differences competitive and non-competitive environments can have, which can have a significant impact on the performance of the athlete. This illustrates the argument that for researchers to truly elucidate the efficacy of an intervention, the methods employed have to replicate the athletes' actual performance. That is, moving away from laboratory based measures to assessing actual performance in the field, and where possible, in a competitive environment.

Since 2004, weekly, open entry, free and timed 5 km road race events (parkrun®), have become increasingly popular throughout the United Kingdom (UK) and offer the opportunity for researchers to understand the efficacy of running interventions on a heterogeneous sample in a competitive environment. However, the reliability of these events has not been established. The aim of this study was therefore to assess the reliability of running performance during an outdoor running event in a competitive environment in trained athletes.

Method

Participants

Institutional ethical approval was obtained from the University of Sunderland. Fifteen, competitive, male, endurance runners (mean ± standard deviation [SD]; age = 29.5 ± 4.3 years, height = 1.75 ± 0.08 m, body mass = 71.0 ± 7.1 kg, 5 km personal best = 19:13 ± 1:13 minutes) were recruited following a call out for participants made through social media to Newcastle-upon-Tyne parkrunners. Participants trained regularly (>5 $d \cdot week^{-1}$) during the 6 months prior to the study and regularly participated in 5 km competitive races. All participants were habituated with the selected course and event, having each completed >10 parkruns® in Newcastle-upon-Tyne prior to the study. Written informed consent was obtained from all participants prior to participation. Participants were informed that they could withdraw from the study at any point in time should they wish to do so without reprisal.

Procedure

A within-participant study design was adopted. Participants completed two 5 km time-trial runs (Trial 1 and Trial 2) in a competitive environment within a 7-21 day period. The 5 km parkrun® trials took place in Newcastle upon-Tyne, UK. The Newcastle-upon-Tyne parkrun® is run on tarmac and has been accurately measured using a professional measuring wheel. The course is officially certified and is located 61-75 m above sea level. Approximately 500 runners compete weekly. For this reason, participants were asked to begin the trial at the front of the mass start to ensure times were not hindered by other runners. Participants were asked to prepare for and treat each run as they would for a competition. They were asked to maintain a similar diet for 48 hours, rest adequately (>8 hours of sleep) and maintain their pre-competition training routines before each trial. Participants performed individual warm up routines and were asked to keep this the same for subsequent trials. The 5 km runs started promptly at 09:00 and participants were instructed to complete the distance as fast as possible. Environmental conditions, wind speed (m/s), temperature (°C), relative humidity (%) and wind chill (°C) were recorded weekly using the Pasco weather sensor (PS-2174, Pasco, Roseville CA, USA) attached to the Xplorer GLX graphing data-logger (PS-2002, Pasco, Roseville CA, USA). Measures were taken at various points around the course and a mean value recorded. Time trials were not recorded on days when the wind speed exceeded ±2 m/s. Weather conditions remained stable (cool and dry) across all trials, with wind speed ranging from 0.9 to 1.8 m/s, temperature from 4 to 7 °C, relative humidity from 82 to 92% and wind chill from 3 to 4 °C.

Statistical analysis

Data are presented as mean ± standard deviation (SD) and 95% confidence intervals (95% CI) in brackets. Normality was assessed using the Shapiro-Wilk test for normality. Paired samples t-tests were conducted to determine any systematic difference in performance time between the two runs (Trial 1 - Trial 2). Cohen's d was calculated to determine the effect size (d) of the mean differences (Cohen, 1977) and interpreted using the modified scale proposed by Hopkins (2002): trivial ≤ 0.2; small 0.2-0.6; moderate, 0.6-1.2; and large, >1.2. Absolute reliability of 5 km performance time was determined using the within-participant coefficient of variation (CV) and typical error (TE) expressed in seconds. A CV ≤1.5% was set as a criterion for absolute reliability (Hopkins & Hewson, 2001). Within-participant CV's were calculated for individual participants by dividing the standard deviation of their Trial 1 and Trial 2 performances by their mean performance and multiplying by 100 (SD [Trial 1 and Trial 2] / mean [Trial 1 and Trial 2]*100). The mean CV is reported. Relative reliability was established using the intra-class correlation coefficient (ICC). TE and ICC were calculated using an online statistical spreadsheet (Hopkins, 2009). The precision of the ICC (95% CI) was established using the McGraw and Wong (1996) formula. The ICC was interpreted as follows: ICC <0.80 low reliability; ICC 0.80 to 0.90 moderate reliability; ICC >0.9 high reliability (Vincent & Weir, 2005). Statistical analyses were conducted using Microsoft Excel and SPSS for windows version 20.0 (SPSS Inc., Chicago, USA) software packages. Significance was accepted at $p < .05$.

Results

Twenty participants completed the first initial trial. Five participants withdrew from the study before completing the second trial (injury, n = 2; no reason provided, n = 2; non-availability, n = 1). Data analysis is based on the 15 participants who successfully completed both time-trials. No differences existed between participants who withdrew and the participants who remained for the demographic variables (i.e. age, height, weight, PB; $p > .05$).

Individual performances are presented in table 1. Performance times for the two 5 km time-trials were highly reproducible (mean ± standard deviation [SD]; 1217 ± 85 s, 95% CI [1170, 1264] and 1216 ± 79 s, [1172 to 1260] for Trial 1 and Trial 2 respectively). The mean difference in running performance between Trial 1 and Trial 2 was 1.0 ± 20.8 s [-10.5, 12.5]. A paired samples t-test revealed no differences between the two trials ($t_{(14)}$, $p = .855$, $d < 0.01$). The coefficient of variation (CV) was 0.95 ± 0.65% [0.59, 1.31], typical error (TE) = 14.7 s [11.3, 21.4] equating to approximately 1.2% of mean performance, and Intra-class correlation = 0.97 [0.93, 0.99].

Discussion

The aim of this present study was to evaluate the reproducibility of an outdoor, competitive time-trial. To our knowledge, this is the first study to assess the reproducibility of this type of measure. Results suggest that an outdoor 5 km time-trial within a competitive environment is highly reproducible in a population of trained athletes. Results have implications for future research that seek to understand the effects of interventions on endurance running performance. The use of a reliable competitive, outdoor time-trial could provide researchers with greater confidence that results of intervention studies can be extrapolated to real world environments.

The use of an indoor, laboratory based time-trial to distinguish the effects of an intervention for endurance performance has been frequently used within sport and exercise science (Stevens & Dascombe, 2015). This is commonly perceived to be a more reliable method compared to outdoor time-trials (Reilly, Morris, & Whyte, 2009). However, results from this study suggest that this may not be entirely accurate. The coefficient of variation (CV) of 0.95 ± 0.65%, 95% CI [0.59, 1.31] reported in this study is similar, if not better than indoor laboratory based time-trials. Russell, Redmann, Ravussin, Hunter, and Larson-Meyer (2004) reported CV of 1.0% for 10 km treadmill based time-trials and Laursen et al. (2007) reported CV of 3.3 and 2.0%, for 1500 m and 5 km treadmill based time-trials, respectively. In addition, the results of Stevens et al. (2015) who attempted to better simulate the outdoor environment with a non-motorised treadmill, reported a similar CV of 1.2% for 5 km time-trials. The use of an outdoor competitive time-trial is therefore comparable, if not more reliable than an indoor, laboratory based time-trial. This holds important implications and considerations for researchers aiming to establish the effectiveness of an intervention on running performance. For inferences to be extrapolated to performance, the use of a protocol that holds high reliability and validity should be used. If this isn’t achieved, the inferences reported may not be translated accurately to actual performance. We therefore encourage the use of the 5 km outdoor, competitive time-trial (i.e. parkrun®) as a means of confidently assessing the efficacy of running interventions.

The typical error (TE; 1.2% or 14.7 s) reported is also lower compared to previously reported variability in non-elite distance runners and corroborates a similar TE (1.3%) observed in trained endurance athletes over a 3000 m indoor time-trial (Durussel et al., 2013) and a 1.4% TE for distances between 3000 m and 10000 m in elite athletes (Hopkins & Hewson, 2001). The low values for TE and CV in the current study may be attributed to the level of participant familiarisation with the competitive, parkrun® time-trial adopted for the study. Stevens et al. (2015) emphasised that to minimise the test-retest variation, running time-trials should include participants that are familiar with the testing procedures. For this study, prior to their recruitment, participants had completed on average 51 ± 38 5 km parkrun® time-trials (range = 14 to 144). This highlights a strength of the current study and of parkrun®, as it ensures that the participants were well versed and familiar with the course and distance. The parkrun® events can provide a unique advantage for future research, as it allows the researcher the opportunity to access a population of athlete that are experienced with the running protocol already, without having to include familiarisation trials prior to the study. In short, this would save time and resources during repeated measure, cross-over design studies.

The results of this study should take into consideration a number of potential limitations. First, the use of pacing and drafting practices were not fully explored during this study. We acknowledge that time-trials by design, permit athletes to alter their pace, which may influence reproducibility of performance. However, we, like others (Hampson, Gibson, Lambert, & Noakes, 2001; Laursen et al., 2007), argue pacing strategies are an integral component of real-life, competitive performance. To maximise ecological validity, athletes in this study were permitted to alter their race-pace to suit the interactions between their perceptions of fatigue and external motivational cues (Hampson et al., 2001). Secondly, it may be argued the ‘competitive’ environment in this study would be better described as semi-competitive as it does not mimic the atmosphere, pressures and demands, arousal or anxiety experienced by elite level athletes competing for podium status at international athletic competitions.

These limitations notwithstanding, the results provide valuable information for researchers wishing to ascertain the effect of interventions on outdoor endurance running performance. Empirical investigations establishing ecologically valid research protocols, in a competitive environment, have been difficult to design. The timing of competition events often differs between venues and are rarely held at the same time of the day. The parkrun® event used in this research eliminates such confounds as parkrun® events are scheduled weekly at the same time of day.

To conclude, mean performance time was found to be highly reproducible over repeated competitive, outdoor-based 5 km time-trials. Results are similar, if not better, than indoor based, treadmill and outdoor, track based time-trials. Results provide a useful platform from which to measure the magnitude of performance changes following future interventions.

References

Ali, A., Caine, M. P., & Snow, B. G. (2007). Graduated compression stockings: physiological and perceptual responses during and after exercise. Journal of Sports Sciences, 25(4), 413-419. doi:10.1080/02640410600718376

Araujo, D., & Davids, K. (2009). Ecological approaches to cognition and action in sport and exercise: Ask not only what you do, but where you do it. International Journal of Sport Psychology, 40(1), 5-37.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. Sports medicine, 38(4), 297-316. doi:10.2165/00007256-200838040-00003

Di Michele, R., Di Renzo, A. M., Ammazzalorso, S., & Merni, F. (2009). Comparison of physiological responses to an incremental running test on treadmill, natural grass, and synthetic turf in young soccer players. The Journal of Strength & Conditioning Research, 23(3), 939-945. doi:10.1519/JSC.0b013e3181a07b6e

Durussel, J., Daskalaki, E., Anderson, M., Chatterji, T., Wondimu, D. H., Padmanabhan, N., ... & Pitsiladis, Y. P. (2013). Haemoglobin mass and running time trial performance after recombinant human erythropoietin administration in trained men. PloS one, 8(2), 1-8. doi:10.1371/journal.pone.0056151

Hampson, D. B., Gibson, A. S. C., Lambert, M. I., & Noakes, T. D. (2001). The influence of sensory cues on the perception of exertion during exercise and central regulation of exercise performance. Sports Medicine, 31(13), 935-952. doi:10.2165/00007256-200131130-00004

Hodges, K., Hancock, S., Currell, K., Hamilton, B., & Jeukendrup, A. E. (2006). Pseudoephedrine enhances performance in 1500-m runners. Medicine and Science in Sports and Exercise, 38(2), 329-333. doi:10.1249/01.mss.0000183201.79330.9c

Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. Retrieved from http://www.sportsci.org/resource/stats/effectmag.htm

Hopkins, W. G. (2009). Calculating the reliability intraclass correlation coefficient and its confidence limits (Excel spreadsheet). Retrieved from http://www.sportsci.org/resource/stats/relycalc.html

Hopkins, W. G., & Hewson, D. J. (2001). Variability of competitive performance of distance runners. Medicine and Science in Sports and Exercise, 33(9), 1588-1592. doi:10.1097/00005768-200109000-00023

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. Journal of Sports Sciences, 14(4), 321-327. doi:10.1080/02640419608727717

Kunduracioglu, B., Guner, R., Ulkar, B., & Erdogan, A. (2007). Can heart rate values obtained from laboratory and field lactate tests be used interchangeably to prescribe exercise intensity for soccer players? Advances in Therapy, 24(4), 890-902. doi:10.1007/BF02849982

Laursen, P. B., Francis, G. T., Abbiss, C. R., Newton, M. J., & Nosaka, K. (2007). Reliability of time-to-exhaustion versus time-trial running tests in runners. Medicine and Science in Sports and Exercise, 39(8), 1374-1379. doi:10.1249/mss.0b013e31806010f5

McAuley, E., Mihalko, S. L., & Bane, S. M. (1997). Exercise and self-esteem in middle-aged adults: multidimensional relationships and physical fitness and self-efficacy influences. Journal of Behavioral Medicine, 20(1), 67-83. doi:10.1023/A:1025591214100

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intra-class correlation coefficients. Psychological Methods, 1(1), 30-46. doi:10.1037/1082-989X.1.1.30

O'Rourke, M. P., O'Brien, B. J., Knez, W. L., & Paton, C. D. (2008). Caffeine has a small effect on 5-km running performance of well-trained and recreational runners. Journal of Science and Medicine in Sport, 11(2), 231-233. doi:10.1016/j.jsams.2006.12.118

Pierce, D., Kupprat, I., & Harry, D. (1976). Urinary epinephrine and norepinephrine levels in women athletes during training and competition. European journal of applied physiology and occupational physiology, 36(1), 1-6. doi:10.1007/BF00421628

Reilly, T., Morris, T., & Whyte, G. (2009). The specificity of training prescription and physiological assessment: A review. Journal of Sports Sciences, 27(6), 575-589. doi:10.1080/02640410902729741

Russell, R. D., Redmann, S. M., Ravussin, E., Hunter, G. R., & Larson-Meyer, D. E. (2004). Reproducibility of endurance performance on a treadmill using a preloaded time trial. Medicine and Science in Sports and Exercise, 36(4), 717-724. doi:10.1249/01.MSS.0000121954.95892.C8

Stevens, C. J., & Dascombe, B. J. (2015). The Reliability and Validity of Protocols for the Assessment of Endurance Sports Performance: An updated review. Measurement in Physical Education and Exercise Science,19(4), 177-185. doi:10.1080/1091367X.2015.1062381

Stevens, C. J., Hacene, J., Sculley, D. V., Taylor, L., Callister, R., & Dascombe, B. (2015). The reliability of running performance in a 5 km time trial on a non-motorized treadmill. International Journal of Sports Medicine, 36(9), 705-709. doi:10.1055/s-0034-1398680

Terry, P. C., Karageorghis, C. I., Saha, A. M., & D'Auria, S. (2012). Effects of synchronous music on treadmill running among elite triathletes. Journal of Science and Medicine in Sport, 15(1), 52-57. doi:10.1016/j.jsams.2011.06.003

Vincent, W.P. & Weir, J.P. (2005). Statistics in Kinesiology. Leeds: Human Kinetics.

Viru, M., Hackney, A., Karelson, K., Janson, T., Kuus, M., & Viru, A. (2010). Competition effects on physiological responses to exercise: performance, cardiorespiratory and hormonal factors. Acta Physiologica Hungarica, 97(1), 22-30. doi:10.1556/APhysiol.97.2010.1.3

Table captions

Note: CV = coefficient of variation

Tables

Table 1. Individual differences for outdoor competitive 5km time-trial

Participant	**Trial 1 (s)**	**Trial 2 (s)**	**Differences (s)**	**Percentage Change in Performance**	**CV (%)**
1	1173	1170	3	0.26%	0.18
2	1157	1150	7	0.61%	0.43
3	1328	1289	39	3.03%	2.11
4	1363	1343	20	1.49%	1.05
5	1121	1154	-33	-2.86%	2.05
6	1180	1176	4	0.34%	0.24
7	1070	1078	-8	-0.74%	0.53
8	1127	1111	16	1.44%	1.01
9	1152	1156	-4	-0.35%	0.25
10	1212	1227	-15	-1.22%	0.87
11	1286	1254	32	2.55%	1.78
12	1261	1250	11	0.88%	0.62
13	1282	1296	-14	-1.08%	0.77
14	1271	1301	-30	-2.31%	1.65
15	1270	1283	-13	-1.01%	0.72
Mean ± SD	1216.9 ± 84.7	1215.9 ± 79.2	-1.0 ± 20.8	0.08 ± 0.02	0.95 ± 0.65