*This is a non-final version of an article published in final form in *Journal of Strength & Conditioning Research**

The Reliability and Validity of Fatigue Measures during Multiple Sprint Work: An Issue Revisited

MARK GLAISTER, GLYN HOWATSON, JOHN R. PATTISON, AND GILL MCINNES

School of Human Sciences, St. Mary's University College, Strawberry Hill, Twickenham, United Kingdom

Corresponding Author:

Mark Glaister

School of Human Sciences

St Mary's University College

Waldegrave Road

Strawberry Hill

Twickenham, UK.

Phone: 0208 240 4012

Fax: 0208 240 4255

E-mail: Glaistem@smuc.ac.uk

Brief running head: Reliability and validity of fatigue

*This is a non-final version of an article published in final form in *Journal of Strength & Conditioning Research**

The Reliability and Validity of Fatigue Measures during Multiple Sprint Work: An Issue Revisited

ABSTRACT

The ability to repeatedly produce a high power output or sprint speed is a key fitness component of most field and court sports. The aim of this study was to evaluate the validity and reliability of eight different approaches to quantify this parameter in tests of multiple sprint performance. 10 physically active men completed two trials of each of two multiple sprint running protocols with contrasting recovery periods. Protocol 1 consisted of 12 × 30 m sprints repeated every 35 s; Protocol 2 consisted of 12 × 30 m sprints repeated every 65 s. All testing was performed in an indoor sports facility and sprint times were recorded using twin-beam photocells. All but one of the formulae showed good construct validity as evidenced by similar within-protocol fatigue scores. However, the assumptions upon which many of the formulae were based, combined with poor or inconsistent test-retest reliability (coefficient of variation range: 0.8 – 145.7%; intraclass correlation coefficient range: 0.09 – 0.75) suggested many problems regarding logical validity. In line with previous research, the results support the percentage decrement calculation as the most valid and reliable method of quantifying fatigue in tests of multiple sprint performance.

Key words: intermittent, repeated, test-retest, repeatability

Introduction

The term multiple sprint sports describes the intermittent activity patterns of many field and court sports. In an attempt to understand the physiological and experimental responses to this type of activity, several multiple sprint tests have been developed using various exercise modalities (3, 6, 14). Based on the results of time-motion analyses, these protocols have typically involved repeated bouts (≤ 20) of maximal work lasting ≤ 10 s, interspersed with relatively short (≤ 60 s) rest periods. The key performance outcomes derived from such tests are: a) an individual's maximum sprint speed or power output; and b) the ability to resist fatigue and maintain a high performance level throughout the test. Whilst the former can be directly measured and has been shown to have good test-retest reliability (6, 7, 9, 14, 15), the same is not true of the latter which relies on various formulae to determine its magnitude (10). In an earlier investigation we examined the validity and reliability of four different approaches to quantify fatigue in multiple sprint work, in an attempt to resolve previous discord (10). Whilst all gave some concern regarding validity, we concluded that the percentage decrement score (6) provided the most valid and reliable measure. Since then, three other approaches have come to light (1, 18) which have prompted us to revisit the issue. Moreover, since one of these approaches contains an obvious limitation regarding its validity, a revised formula has also been included for discussion.

The aim of the present study therefore was to assess the validity and reliability of eight different approaches to calculating fatigue in multiple sprint work. To improve the ecological validity of our previous study (multiple sprint cycling), multiple sprint running was chosen as the mode of exercise.

Methods

Experimental Approach to the Problem

To evaluate the construct validity of the various fatigue calculations used in this investigation, all subjects completed two trials of each of two multiple sprint running protocols. Protocol 1 consisted of 12 × 30 m sprints repeated at 35 s intervals. Protocol 2 consisted of 12 × 30 m sprints repeated at 65 s intervals. Although the absence of any learning effects from this type of multiple sprint protocol has recently been reported (7), all subjects completed a familiarization trial of Protocol 1 prior to the experiment as a precautionary measure. The order of the experimental trials was randomised and all trials were completed at approximately the same time of day, with a minimum of 48 hours between trials. Subjects were instructed to maintain their normal diet throughout the testing period, to avoid food and drink in the hour before testing, and to avoid strenuous exercise 24 hours before each trial.

Subjects

10 male sport science students volunteered for the study which was approved by St Mary's University College Ethics Committee. Prior to testing, subjects received written and verbal instructions regarding the nature of the investigation and completed a training history questionnaire which indicated that all had been actively involved in sport for approximately 16 years and that all regularly participated in some form of multiple sprint sport. Times spent training and competing each week were reported as 9.2 ± 5.4 hours and 3.4 ± 1.2 hours respectively. Prior to commencement, all subjects completed a health-screening questionnaire and provided written informed consent. Means ± standard deviation (SD) for age, height, body mass, and body fat (5) of the

subjects were: 23 ± 5 years, 178.7 ± 5.9 cm, 82.3 ± 13.4 kg, and 15.0 ± 4.4% respectively.

Testing Procedures

All testing was conducted indoors on a synthetic running surface. Prior to each multiple sprint test, subjects performed a standardised warm-up (approximately five-minutes) comprising 400 m of jogging (self-selected pace), a series of sprint drills (high-knees, heel-flicks, and walking lunges), and 3 practice sprints. Following the warm-up, subjects were given 5 minutes to stretch and prepare themselves for the multiple sprint test. Each sprint was initiated from a line 30 cm behind the start line (to prevent false triggering of the first timing gate) and times were recorded electronically via twin-beam photocells (Swift Performance Equipment, Lismore, Australia). Alternate sprints were performed in the opposite direction thereby enabling subjects to maximise the available recovery time between sprints. Computer-generated audio signals provided a 5 s countdown to the start of each sprint and subjects were verbally encouraged to give maximal effort. Fatigue during each test was calculated from sprint times using the following formulae:

Formula 1 (F1)

Fatigue = the percentage increase in time between the first and last sprints (4).

Calculation:

Fatigue = ((sprint 12 – sprint 1) ÷ sprint 1) x 100

Formula 2 (F2)

Fatigue = the percentage increase in time between the fastest and slowest sprints (12).

Calculation:

Fatigue = ((slowest sprint – fastest sprint) ÷ fastest sprint) x 100

Formula 3 (F3)

Fatigue = the back-transformation of the slope of the line of best fit for log-transformed sprint times over all sprints (17).

Calculation:

Fatigue = $(100 \times \mathrm{EXP}^{(\text{slope} \div 100)}) - 100$

Where:

Slope = (the slope of the line of best fit for: 100 x natural logarithm of sprint data) x (number of sprints – 1).

Formula 4 (F4)

Fatigue = the percentage decrement score (6).

Calculation:

Fatigue = (100 x (total sprint time ÷ ideal sprint time)) – 100

Where:

Total sprint time = sum of sprint times from all sprints.

Ideal sprint time = the number of sprints x fastest sprint time.

Formula 5 (F5)

Fatigue = the percentage increase in time between the average of the fastest two and slowest two sprints (1).

Calculation:

Fatigue = (((slowest two sprint times ÷ 2) – (fastest two sprint times ÷ 2)) ÷ (fastest two sprint times ÷ 2)) x 100

Formula 6 (F6)

Fatigue = the difference in speed between the mean of the first two and the mean of the last two sprints expressed as a percentage of the mean speed of the first two sprints (18).

Calculation:

Fatigue = ((((sprint 1 + sprint 2) ÷ 2) – ((sprint 11 + sprint 12) ÷ 2)) ÷ ((sprint 1 + sprint 2) ÷ 2)) x 100

Formula 7 (F7)

Fatigue = the ratio of mean speed to the average speed of the first two sprints (18).

Calculation:

Fatigue = ((total sprint speed ÷ 12) ÷ ((sprint 1 + sprint 2) ÷ 2)) x 100

Formula 8 (F8)

Fatigue = 100 minus the ratio of mean speed to the average speed of the first two sprints (amended version of F7).

Calculation:

Fatigue = 100 – (((total sprint speed ÷ 12) ÷ ((sprint 1 + sprint 2) ÷ 2)) x 100)

To provide additional information regarding construct validity, fatigue was also calculated over the first half of each trial.

Statistical Analyses

All statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS for Windows, SPSS Inc.). Measures of centrality and spread for each protocol were derived from the average of the performance measures over the two trials and are presented as means ± SD. A two-way ANOVA with repeated measures on both factors (protocol x trial) was used to evaluate any between-protocol differences in sprint 1 performance. Any within-subject pairs of negative fatigue scores were treated as positive scores for the purpose of the reliability analyses, whilst other negative scores were disregarded. Measures of reliability, presented as coefficients of variation (CV) and intraclass correlation coefficients (ICC), were derived from two-way ANOVA as described by Schabort *et al*. (20). Fatigue was the dependant variable in each model, with subject identity included as a random effect and trial number as a fixed effect. ICC were derived from the ANOVA using the method described by Bartko (2). Confidence limits (95%) for CV and ICC were calculated using chi square and McGraw & Wong (16) estimates respectively.

Results

Mean sprint times for each protocol are presented in Figure 1, with resultant fatigue scores and reliability data for Protocols 1 and 2 presented in Tables 1 and 2, respectively. 70% of fastest sprint times during Protocol 1 occurred in sprint 1, compared with 50% of cases during Protocol 2. There were no significant between-protocol differences in sprint 1 performance ($p = 0.274$). There were 17 instances of negative fatigue scores all resulting from Protocol 2, five of which resulted from F6, the remainder being evenly distributed across formulae F1, F3, and F8.

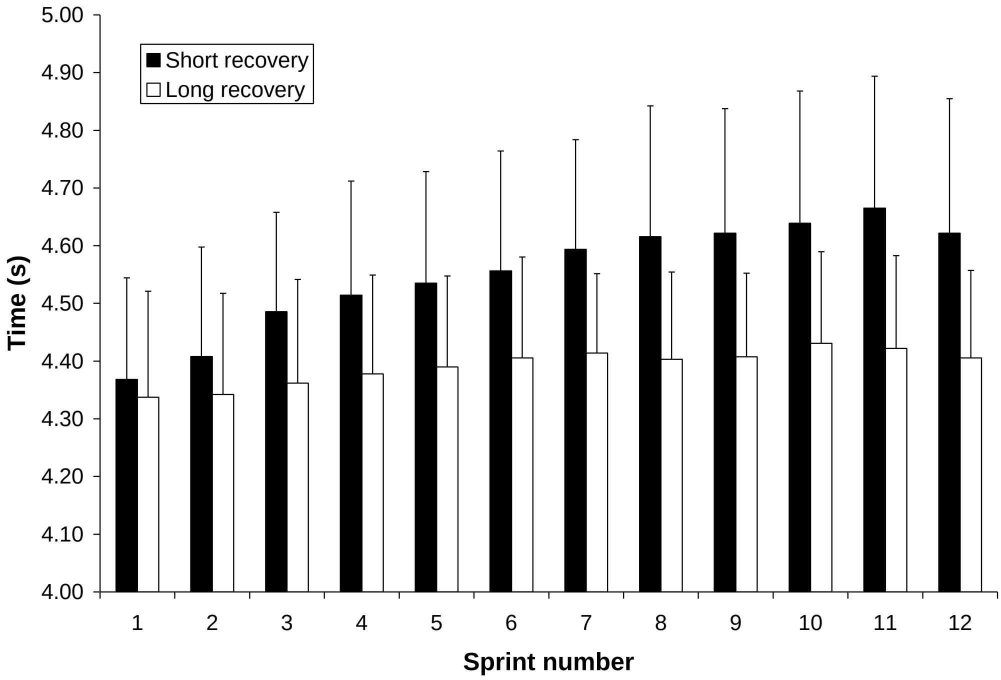

Figure 1. The influence of recovery duration on multiple sprint running performance (12 × 30 m; repeated at either 35 s or 65 s intervals). Values are means; bars are standard deviations.

Table 1. Fatigue scores and within-subject test-retest reliability of multiple sprint running trials (12 × 30 m; repeated at 35 s intervals)

	Formula							
	1	2	3	4	5	6	7	8
Fatigue (%)	5.81 ± 3.70	7.72 ± 3.22	6.13 ± 3.70	4.43 ± 1.79	6.72 ± 2.81	5.43 ± 2.71	96.47 ± 1.31	3.53 ± 1.31
Fatigue (1-6) (%)	4.31 ± 2.17	5.13 ± 1.90	4.41 ± 1.95	2.73 ± 1.06	3.96 ± 1.30	3.45 ± 1.35	98.02 ± 0.65	1.98 ± 0.65
CV (%)	107.2	25.3	52.8	31.7	24.8	45.2	0.9	26.2
Lower CL	73.8	17.4	36.4	21.8	17.0	31.1	0.6	18.0
Upper CL	195.8	46.3	96.5	57.9	45.2	82.5	1.6	47.8
ICC	0.17	0.66	0.54	0.51	0.65	0.47	0.62	0.57
Lower CL	-0.48	0.09	-0.10	-0.13	0.08	-0.18	0.03	-0.04
Upper CL	0.70	0.90	0.86	0.85	0.90	0.84	0.89	0.87

Fatigue (1-6) = fatigue over sprints 1 – 6; CV = coefficient of variation; CL = 95% confidence limits

Table 2. Fatigue scores and within-subject test-retest reliability of multiple sprint running trials (12 × 30 m; repeated at 65 s intervals)

	Formula							
	1	2	3	4	5	6	7	8
Fatigue (%)	1.62 ± 2.24	3.67 ± 1.38	1.96 ± 2.46	1.97 ± 0.86	3.27 ± 1.32	1.68 ± 2.02	98.82 ± 1.24	1.17 ± 1.24
Fatigue (1-6) (%)	1.59 ± 1.87	2.62 ± 1.13	1.68 ± 1.81	1.28 ± 0.50	2.05 ± 0.95	1.32 ± 1.40	99.33 ± 0.74	0.67 ± 0.74
CV (%)	145.7	25.3	65.5	37.4	27.8	62.7	0.8	141.2
Lower CL	100.2	17.4	45.1	25.7	19.1	43.1	0.5	97.1
Upper CL	266.0	46.3	119.6	68.3	50.7	114.5	1.4	257.7
ICC	0.30	0.60	0.70	0.44	0.60	0.75	0.65	0.09
Lower CL	-0.36	0.00	0.17	-0.22	0.00	0.27	0.08	-0.54
Upper CL	0.77	0.88	0.92	0.82	0.88	0.93	0.90	0.66

Fatigue (1-6) = fatigue over sprints 1 – 6; CV = coefficient of variation; CL = 95% confidence limits

Discussion

The purpose of this study was to evaluate the reliability and validity of eight different approaches to quantifying fatigue in multiple sprint work. In our previous article on this topic we examined formulae F1 to F4 and concluded that, despite limitations in the validity of all the formulae evaluated, F4 provided the best means of quantifying fatigue in this type of activity (10). Despite differences in the mode of exercise used, the results of the present study corroborate our previous findings, particularly when comparing those protocols with similar work and rest periods.

The main limitation of trying to evaluate fatigue in any type of exercise is that there is no 'gold standard' criterion with which to compare resultant values. However, the results of the present study suggest that with the exception of F7, which calculates the percentage of performance maintained rather than lost, all formulae were assessing a similar construct. Moreover, the fact that those same formulae produced higher values of fatigue in each trial than in the first half of each trial, coupled with the fact that they also produced higher fatigue scores for Protocol 1 than Protocol 2, supports a good level of construct validity. Nevertheless, there are a number of between and within-protocol anomalies in the pattern of the fatigue process during multiple sprint work that raise concerns regarding the logical validity of the various formulae. For instance, several studies have observed a potentiation of power output during the first few sprints in this type of work (4, 11, 12, 13, 19, 21). Furthermore, several studies, including this one, have observed an upsurge in power output or speed in the final sprint (8, 11, 19, 21). Whilst the mechanisms for the appearance, or lack of, these phenomena require further investigation, these effects have particular implications for F1 which is based on the assumption that the best and worst sprints occur at the extremes of a multiple sprint test. Although F6 attempts to control for this effect by using the means of the first and last two sprints, the degree of improvement in reliability is insufficient to support the validity of these approaches as a means of quantifying fatigue. Another effect which impacts on the validity of the various formulae is that caused by noise in the measurement due to technological and biological variability. In effect a participant could show no signs of fatigue in a multiple sprint test, yet it would be unusual to expect the same performance outcome from each sprint. This phenomenon has its greatest effect on F2 and F5 which rely on test extremes to determine fatigue and is evident in the higher than expected fatigue scores from both of those formulae. A final phenomenon which impacts on the validity of approaches to calculate fatigue in multiple sprint work is the fact that although the fatigue process appears to be fairly linear over the first few sprints, the effect tends to plateau as the number of sprints is extended (8, 11, 19). This effect has implications for F3 which calculates fatigue based on the slope of the fatigue process. All in all, the limitations described above combined with the number of negative fatigue scores, and the poor (or inconsistent) reliability demonstrated by many of the formulae suggest many problems regarding validity. Indeed, since a measure cannot be considered valid if it is not reliable, there are limitations in the validity of all the formulae evaluated. However, care must be taken when interpreting the test-retest reliability of various measures as evidenced by the considerable difference in CV between F7 and F8. CV is the typical (or standard) error in the measurement between trials expressed as a percentage of the mean score. In F7 the mean values of fatigue were much higher than those of F8 due to the fact that the formula was assessing a different construct. As such, despite the same error in the measurement, the CVs suggest a large difference in reliability.

Overall, since the assessment of fatigue is one of the key outcomes from a multiple sprint test, and since every reasonable approach to quantify this parameter appears to have been considered, we are left with no option but to recommend the most valid and reliable approach. As with our previous article, F4 seems to be the most suitable as it considers data from each sprint, provides consistent reliability, and shows good construct and logical validity. Although the results show a test-retest variability in fatigue of around 30%, when judged in context with the magnitude of the mean fatigue scores, it is possible that this degree of variability would still allow the effects of various experimental interventions to be evaluated effectively.

Practical Applications

Tests of multiple sprint work are becoming increasingly common as a means of evaluating the effects of various interventions on the capabilities of athletes involved in multiple sprint sports. Since the assessment of fatigue is one of the main performance outcomes from this type of test, the results of the present study support the use of the percentage decrement score as the most valid and reliable measure of this parameter.

Acknowledgements

The authors would like to express their gratitude to all the participating subjects for their enthusiasm and commitment, and to Peter Kenyon for writing the software to regulate the multiple sprint tests.

References

1. BAKER, J., R. RAMSBOTTOM, AND R. HAZELDINE. Maximal shuttle running over 40 m as a measure of anaerobic performance. *Br J Sports Med*. 27: 228–232. 1993.

2. BARTKO, J. J. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep*. 19: 3–11. 1966.

3. BISHOP, D., M. SPENCER, R. DUFFIELD, AND S. LAWRENCE. The validity of a repeated sprint ability test. *J Sci Med Sport.* 4: 19–29. 2001

4. BROOKS, S., M. E. NEVILL, L. MELEAGROS, H. K. LAKOMY, G. M. HALL, S. R. BLOOM, AND C. WILLIAMS. The hormonal responses to repetitive brief maximal exercise in humans. *Eur J Appl Physiol Occup Physiol*. 60: 144–148. 1990.

5. DURNIN, J. V., AND J. WOMERSLEY. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *Br J Nutr*. 32: 77–97. 1974.

6. FITZSIMONS, M., B. DAWSON, D. WARE, AND A. WILKINSON. Cycling and running tests of repeated sprint ability. *Aust J Sci Med Sport.* 25: 82–87. 1993.

7. GLAISTER, M., G. HOWATSON, R. A. LOCKEY, C. ABRAHAM, J. GOODWIN, AND G. MCINNES. Familiarization and reliability of multiple sprint running performance indices. *J Strength Cond Res*. In Press. 2007.

8. GLAISTER, M., R. A. LOCKEY, C. ABRAHAM, A. STAERCK, J. GOODWIN, AND G. MCINNES. Creatine supplementation and multiple sprint running performance. *J Strength Cond Res*. 20: 273–277. 2006.

9. GLAISTER, M., M. H. STONE, A. M. STEWART, M. HUGHES, AND G. L. MOIR. Reliability of power output during short-duration maximal-intensity intermittent cycling. *J Strength Cond Res.* 17: 781–784. 2003.

10. GLAISTER, M., M. H. STONE, A. M. STEWART, M. HUGHES, AND G. L. MOIR. The reliability and validity of fatigue measures during short-duration maximal-intensity intermittent cycling. *J Strength Cond Res.* 18: 459–462. 2004.

11. GLAISTER, M., M. H. STONE, A. M. STEWART, M. HUGHES, AND G. L. MOIR. The influence of recovery duration on multiple sprint cycling performance. *J Strength Cond Res*. 19: 831–837. 2005.

12. HAMILTON, A. L., M. E. NEVILL, S. BROOKS, AND C. WILLIAMS. Physiological responses to maximal intermittent exercise: differences between endurance-trained runners and games players. *J Sports Sci*. 9: 371–382. 1991.

13. HOLMYARD, D. J., M. E. CHEETHAM, H. K. A. LAKOMY, AND C. WILLIAMS. Effect of recovery duration on performance during multiple treadmill sprints. In: *Science and Football.* T. Reilly, A. Lees, K. Davids, and W. J. Murphy (eds.). London: E & FN Spon, 1988. Pp. 134–142.

14. HUGHES, M. G., M. DOHERTY, R. J. TONG, T. REILLY, AND N. T. CABLE. Reliability of repeated sprint exercise in non-motorised treadmill ergometry. *Int J Sports Med*. 27: 900–904. 2006.

15. MCGAWLEY, K., AND D. BISHOP. Reliability of a 5 x 6-s maximal cycling repeated-sprint test in trained female team-sport athletes. *Eur J Appl Physiol*. 98: 383–393. 2006.

16. MCGRAW, K. O., AND S. P. WONG. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1: 30–46. 1996.

17. PATON, C. D., W. G. HOPKINS, AND L. VOLLEBREGT. Little effect of caffeine ingestion on repeated sprints in team-sport athletes. *Med Sci Sports Exerc*. 33: 822–825. 2001.

18. PSOTTA, R., P. BLAHUS, D. J. COCHRANE, AND A. J. MARTIN. The assessment of an intermittent high intensity running test. *J Sports Med Phys Fitness*. 45: 248–256. 2005.

19. ROBINSON, J. M., M. H. STONE, R. L. JOHNSON, C. M. PENLAND, B. J. WARREN, AND R. D. LEWIS. Effects of different weight training exercise/rest intervals on strength, power, and high intensity exercise endurance. *J Strength Cond Res.* 9: 216–221. 1995.

20. SCHABORT, E.J., J. A. HAWLEY, W. G. HOPKINS, AND H. BLUM. High reliability of performance of well-trained rowers on a rowing ergometer. *J Sports Sci.* 17: 627–632. 1999.

21. STONE, M. H., K. SANBORN, L. L. SMITH, H. S. O'BRYANT, T. HOKE, A. C. UTTER, R. L. JOHNSON, R. BOROS, J. HRUBY, K. C. PIERCE, M. E. STONE, AND B. GARNER. Effects of in-season (5 weeks) creatine and pyruvate supplementation on anaerobic performance and body composition in American football players. *Int J Sports Nutr*. 9: 146–165. 1999.