Kyprianou, E, Lolli, L, Haddad, HA, Di Salvo, V, Varley, MC, Mendez Villanueva, A, Gregson, W and Weston, M

A novel approach to assessing validity in sports performance research: integrating expert practitioner opinion into the statistical analysis

Abstract

Purpose: Maximal sprinting speed is decisive in soccer, placing great importance on the valid measurement of this variable. Through equivalence testing, we used expert practitioner opinion to evaluate 10-Hz Global Positioning System (GPS) validity for measuring maximal sprinting speed. **Methods**: We surveyed practitioners on issues related to the measurement of maximal sprinting speed and also assessed twelve elite youth soccer players performing two maximal 40 m sprints, measured by 10-Hz GPS units and a criterion measure (100-Hz Laser). Setting equivalence bounds as practitioner opinion of the practically acceptable amount of measurement error for maximal sprinting speed, we assessed agreement between GPS and Laser. **Results**: Survey respondents (n=50) reported using a combination of methods for deriving maximal sprinting speed (tests, training, match) but the majority did not assess system validity. The median value of practically acceptable amount of measurement error for maximal sprinting speed was 0.20 m/s. Maximal sprinting speed was 8.79 ± 0.33 m/s (Laser) and 8.75 ± 0.32 m/s (GPS) and the mean difference was 0.04 (90% confidence interval -0.03 to 0.11) m/s. Equivalence testing using 0.2 m/s as lower (-0.2 m/s) and upper (+0.2 m/s) thresholds, or as a range (-0.1 to +0.1 m/s), showed Radar Gun and GPS as most likely and likely equivalent measures, respectively. **Conclusions**: Assessed against our expert-informed equivalence thresholds, GPS-measured maximal sprinting speed is equivalent to that recorded by a criterion. When measuring maximal sprinting speed over 40 m, GPS can be used with confidence.

Key Words: Validation; Equivalence Testing; GPS; Soccer; Peak Velocity

Introduction

In soccer, maximal sprinting represents the most infrequent match activity recorded by elite male youth soccer players.[1,2] Despite this, the practical importance of sprinting is shown via straight-line sprints preceding a high percentage of goals scored and match sprint distance being greater for successful compared to unsuccessful teams.[3] As such, players' ability to sprint at high velocities is decisive and therefore practitioners regularly use sprint tests to inform training prescription and manage player physical preparation.[4] Furthermore, the measurement and interpretation of training and match distances in speed zones defined relative to players' maximal sprinting speed, opposed to arbitrary zone classification, could help practitioners prescribe an appropriate training stimulus that minimises negative consequences of an inaccurate prescription and quantification of workload.[5-7]

Fully automatic timing systems, laser guns and high-speed video are considered to be gold standards for measuring sprinting speed,[8] yet Global Positioning Systems (GPS) are more accessible and easier to use in daily practice. As such, GPS are now frequently used in team sports to measure and monitor player running velocities during training and matches.[9] However, it is important that practitioners have confidence in systems used to measure maximal sprinting speed, especially when systems are noncriterion measures.[6] Validity studies are therefore fundamental in the development of alternate measures that save costs, facilitate analyses, and enable data field-based collection.[10]

Validity studies compare a new, or more practically feasible measure against a gold standard (criterion), and if the difference in measures is sufficiently small, validity is assumed. For example, the difference in 40 m maximal sprinting speed measured via 10-Hz GPS and a radar gun was trivial (-0.8%; 90% confidence interval -1.1 to -0.4%) and so GPS was concluded to provide a valid measure of maximal sprinting speed.[6] While this study and others[11] clearly represent a valued additions to the literature, between-system differences were interpreted against standardised thresholds which are influenced by heterogeneity.[12-14] Furthermore, effect (e.g., difference) magnitude should be evaluated according to its practical relevance and a standardised scale may not be relevant to the research question.[14] Indeed, team sport researchers and practitioners should not be constrained by interpreting practical relevance via standardised thresholds. An alternate approach here could be the gathering of information on what constitutes the smallest important difference through gauging expert/end-user opinion[15] as practitioner insight can represent a catalyst for external validity.[16]

Recently, equivalence testing has been suggested to have potential for advancing measurement research in exercise science.[10] This approach assesses whether two measurement systems are statistically equivalent by comparing the differences against a pre-determined 'area of equivalence'. The concept of statistical equivalence is, however, heavily influenced by the choice of the equivalence region[10] and here the use of standardised thresholds as benchmarks is considered a last resort.[17] Relying on standardised effect sizes as justification from the smallest effect size of interest should therefore be avoided. What may be of more[13] relevance to practitioners and researchers in sport and exercise science is setting equivalence thresholds around the smallest numerical value, in raw units, that experts perceive practically relevant. As such, the aims of the present study were twofold, 1) to survey expert opinion on issues surrounding the measurement of maximal sprinting speed in elite soccer, and 2) to assess the validity of GPS as a measure of maximal sprinting speed using equivalence testing informed by surveyed expert opinion.

Methods

Maximal Sprinting Speed Survey

To obtain information on issues related to the maximal sprinting speed measurement, we conducted a short cross-sectional survey. Here, practitioners (sport scientists, strength and conditioning coaches, and fitness coaches) currently working in elite soccer, were asked about perception and practices of their teams maximal sprinting speed measurement. The survey was circulated privately to known contacts with data collected using an online survey platform (Online Surveys, formerly Bristol Online Surveys [BOS]). The survey consisted of ten questions, covering two main areas: 1) introduction/ informed consent and background information (Questions 1 to 5), and 2) issues related to the measurement of maximal sprinting speed (Questions 6 to 10), of which all were multiple choice questions.

Participants and Study Design for the Maximal Sprinting Speed Assessment

Twelve full-time male youth soccer players (age 16.3 ± 0.8 years, body mass 54.5 ± 1.2 kg, height 173.9 ± 6.2cm) were recruited from an elite academy. All players were participating in ~8 training sessions per week, combining soccer, strength and conditioning training, and competitive play. This observational study conformed to the Declaration of Helsinki and received ethics approval from the Aspire Zone Research Committee and the Anti-Doping Laboratory Institutional Review Board, Qatar (approval number E20140000012).

Methodology

Validity of 10-Hz global positioning systems (GPS) units against a criterion measure (100-Hz Laser) was tested for maximal sprinting speed. All testing was undertaken on an outdoor natural grass pitch and all players wore their regular soccer boots. Participants performed two maximal 40-m sprints (Trial 1, Trial 2) with three minutes rest between efforts. Typical errors for the between-trial differences were 0.13 (90% confidence interval 0.10 to 0.20) m/s for Laser and 0.07 (0.06 to 0.11) m/s for GPS, and intraclass correlation coeffcients were 0.85 (0.64 to 0.95) and 0.95 (0.88 to 0.98), respectively. Maximal sprinting speed was assessed simultaneously via 10-Hz GPS (Catapult Optimeye S5, version 7.32) and Laser (Laveg LDM 300C, Jenoptik, Germany). Each sprint was recorded using a hand-held digital video recorder (SONY AX53 4K) to allow precise time alignment between GPS and Laser. Each GPS unit was inserted into the manufacturer provided vest that was fitted tightly to the players, holding the receiver between the scapulae. All devices were activated 15 min before data collection to allow acquisition of satellite signals in accordance with the manufacturer's instructions.[18] The average horizontal dilution was 0.68 ± 0.04 and the average number of satellites per unit was 12.0 ± 0.0. Laser was calibrated with zero showing the start of the 40 m measured sprint and was centred on the middle of the running lane. Laser height was 1.2 m and all measurements were taken from the centre of the lens which was 3.1 m behind the starting line. The laser beam was directed at the lower part of the players back. After recording, GPS data were downloaded to a computer and analyzed using the manufacture's software (Catapult Openfield Software, version 1.21.1).[6] The raw GPS velocity data are calculated using the Doppler-shift method.[19] Laser data were processed using the software associated with the device (das3e). Displacement-time data were captured at 100-Hz and analyzed with a 51-point moving average, and from this an instantaneous velocity trace was derived. The velocity trace was used to establish the maximal velocity that occurred within the 40 m measured sprint.

Statistical Analysis

All survey data are presented as response frequency (expressed as a percentage) or where appropriate, the median and interquartile range (IQR). The peak value attained from either Trial 1 or Trial 2 was used as the maximal sprinting speed recorded by the two different measurement systems. Using the *TOSTER* package,[13] we assessed for statistical equivalence between our two measurement systems using two one-sided tests (TOST), as per the guidelines for assessing agreement between a surrogate measure (GPS) with a known criterion measure (Laser).[10] For equivalence testing, users need to define the targeted region,[10] so we set the lower and upper equivalence bounds from the median value that experts surveyed perceived as the acceptable amount of measurement error. Here, the median value of 0.20 m/s was represented by the upper end of the response category given that question on the acceptable amount of measurement error contained categories encompassing a range of measurement error (e.g., 0.15 – 0.20 m/s). As such, our equivalence bounds were specified before results are known.[13] Given that measurement error is random (i.e., + or -), we acknowledge the potential for ambiguity when asking survey respondents on the practically acceptable amount of measurement error. As such, we assessed for equivalence using the median value as the lower and upper equivalence bounds (-0.20 m/s, +0.20 m/s) and also as a range spanning 0.2 m/s, giving lower and upper equivalence bounds of -0.10 m/s and +0.10 m/s, respectively. Results of equivalence tests can be obtained by mere visual inspection of the confidence interval,[13] with statistical equivalence between the two measures concluded when the 90% confidence interval around the mean difference excludes the lower and upper equivalence bounds.[13] However, to avoid test interpretation via the dichotomy of null hypothesis significance testing,[20,21] we assessed equivalence on a continuous scale. This was done via conversion of the t statistics from both one-sided tests to a probability (via the t-distribution), with subsequent equivalence probability interpreted using a one-sided calibrated Bayes.[22-24] Here, probabilities were interpreted using the following scale: 75–95%, likely; 95-99.5%, very likely; >99.5%, most likely,[25] and equivalence was indicated by the lower probability.[10,13] Analyses were performed in R (version 3.4.1, R Foundation for Statistical Computing, Vienna, Austria). Uncertainty in all estimates is presented as 90% confidence intervals.

Results

Median time (min:sec) to complete the survey was 02:57 (02:08,4:27) and of 50 respondents, 60% were sports scientists, 32% fitness coaches, and 8% strength and conditioning coaches (Question two). Respondents had a median of 8 (5,12) years' experience working in elite soccer (Question three) and worked at predominantly European soccer clubs (76%) (Question five). Where respondents selected a combination of methods for deriving maximal sprinting speed (Question nine), the majority of responses were for the combination of match and training (55%). The median value for the practically acceptable amount of measurement error for maximal sprinting speed (Question ten) was 0.20 (0.10,0.25) m/s. For this question, two respondents chose 'Other' and provided exact values of 5% and 0.6 km/h, respectively; the latter of these values was included in the appropriate answer category giving a total of 49 answers for this question (Figure 1).

Sprint times were 8.72 ± 0.34 m/s (Trial 1) and 8.71 ± 0.28 m/s (Trial 2) for Laser, and 8.69 ± 0.32 m/s (Trial 1) and 8.72 ± 0.29 m/s (Trial 2) for GPS. The mean of the players' fastest sprint from either Trial was 8.79 ± 0.33 m/s (Laser) and 8.75 ± 0.32 m/s (GPS) and the mean difference was 0.04 (90% confidence interval -0.03 to 0.11) m/s. Equivalence of maximal sprinting speeds measured by Laser and GPS was most likely (probability 100%) when using 0.2 m/s as the lower and upper thresholds (Figure 2a), and likely (probability 93.7%) when using 0.2 m/s as a range (Figure 2b).

Discussion

Maximal sprinting speed is key in soccer, so measurement validity is needed, especially when systems are not a gold standard measure. Therefore, in this study we employed a novel approach for assessing GPS validity as a measure of maximal sprinting speed by using equivalence testing informed by expert practitioner opinion. We found that GPS-measured maximal sprinting speed was likely to very likely equivalent to a criterion, gold standard measure and therefore practitioners should have confidence in GPS as a measure maximal sprinting speed.

Additionally, our survey results provide valuable insights into current practices surrounding the measurement of maximal sprinting speed in elite soccer.

While previous work has shown validity of 10-Hz GPS for measuring maximal sprinting speed,[6] criterion comparison was made via standardisation. Standardised scales, however, lack practical context and may therefore not be relevant to the research question.[14] This is by no means a criciticism as establishing externally valid minimum important differences represents a huge challenge to sport and exercise science as changes in one variable need to be assessed against subsequent changes in a relevant anchor such as performance.[15] Use of expert opinion therefore represents a credible approach to informing the definition of practically important differences or, in the context of our study, an acceptable amount of measurement error for measures relevant to sports performance.[26] Generally, reliability studies in this research domain have entailed, to a great extent, the indiscriminate calculation of Pearson's correlation coefficients,[27] or the definition of the typical error of the estimate expressed in percentage points whose magnitude assessment may be irrelevant for both the researcher and practitioner.[28] Notwithstanding the deceptive simplicity and specious practicality of calculating these common statistics, failure to express the actual amount of measurement error adopting a meaningful metric may limit practitioner definition of what represents a true population increase in the response of interest deemed substantially greater than a predefined practically important difference.[26] Further, tests of mean difference are common in agreement research but may not neccessarily represent the best statistical approach.[10] Equivalence testing has been proposed as a more appropriate method for evaluating agreement among measures than mean difference tests; however, choosing and justifying equivalence regions is a difficult aspect of this approach.[10] Indeed, previous studies using equivalence testing have reported, yet not justified the smallest effect size of interest for the equivalence bounds.[13] Therefore, we attempted to overcome these methdological concerns by setting our equivalence bounds on what a relatively large sample of experienced practitioners perceived to be an acceptable amount of measurement error when measuring maximal sprinting speed. This novel and rigourous approach enabled us to conclude that the GPS-measured maximal sprinting speed is likely to most likely equivalent to the speed recorded by a gold standard measure (Laser).

For accurate assessment of maximal sprint speed, fully automatic timing systems represent the gold standard with dual-beamed photocells, laser guns and high-speed video timing representing cheaper, more practical tools with acceptable accuracy.[8] The results presented in this study lend the first empricial support of this observation given that 84% of survey respondents perceived either laser/ radar guns, fully automatic timing systems and timing gates as gold standard measures. Only 16% of respondents regarded GPS as gold standard despite these systems being the most frequent system (34%) used to measure maximal sprinting speed in the field. Our findings help to address this apparent disconnect as practitioners can be assured of the validity of maximal sprinting speeds recorded by 10-Hz GPS. The infrequent nature of system validity checks observed in our study possibly reflects a lack of available time given that practitioners are indeed cognisant of the need for validity assessments.[29]

The most common single method to derive maximal sprinting speed in our survey was fitness tests. The neeed for sprint testing was recently questioned as peak speeds recorded during matches were faster than when recorded during a 40-m maximal running test, albeit in semi-professional senior players.[9] These findings contrast with previous work whereby highly trained youth footballers' maximal match speeds were ~90% of the speed attained on a 40-m sprint test.[30 31] In light of these equivocal findings, it is encouraging that survey respondents derived maximal sprinting speeds from a variety of scenarios (e.g., tests, training, matches). Such an approach will help to ensure an on-going calibration of maximal speeds, which is of vital importance if these speeds are used to inform the classification of relative speed zones.[6]

Practical Applications and Conclusion

Despite not being perceived as a gold standard measure of maximal sprinting speed by the experts we surveyed, speeds recorded by 10-Hz GPS were equivalent to a gold standard measure, thereby supporting validity. Utilising an approach that overcomes methodological concerns with traditional validation studies, our data therefore strengthen the confidence[6] practitioners can take from using GPS to assess maximal sprinting speed. Furthermore, using GPS to measure maximal sprinting speed during fitness tests negates the need for more expensive and less accessible testing equipment, resulting in less time burdensome tests. Whether or not practitioners continue to use dedicated sprint tests to assess maximal sprinting speeds may well depend on the purpose of the test. For example, dedicated sprinting tests clearly have worth if used to benchmark physical progression but may well be unnecessary if the sole purpose is to establish maximal speeds to inform relative training and match activity zones. Indeed, our survey shows that fitness tests are no longer the sole method used by practitioners for measuring maximal sprinting speeds. As training and match data are now used by practitioners to assess maximal speeds, future research should build on our findings by examining the whether maximal speeds are more frequently occuring during training or matches. Such research would have important implications for informing player preparation and performance evaluation strategies.

Acknowledgments

We would like to acknowledge the expertise of Dr Philip Graham-Smith and also the cooperation of the players, coaches and survey respondents, without whom the study would not have been possible.

Figure Legends

Figure 1. Responses (n=49) for the practically acceptable amount of measurement error for maximal sprinting speed (Question ten)

Figure 2a. Mean difference (m/s) and uncertainty for the difference (90% confidence interval) in maximal sprinting speed measured by Laser and GPS. The black vertical dashed lines represents the expert-informed statistical equivalence region of 0.2 m/s, expressed as the lower and upper threshold.

Figure 2b. Mean difference (m/s) and uncertainty for the difference (90% confidence interval) in maximal sprinting speed measured by Laser and GPS. The black vertical dashed lines represents the expert-informed statistical equivalence region of 0.2 m/s, expressed as a range.

Reference List

1. Harley JA, Barnes CA, Portas M, et al. Motion analysis of match-play in elite U12 to U16 age-group soccer players. *J Sports Sci* 2010;28(13):1391-7. doi: 10.1080/02640414.2010.510142 [published Online First: 2010/10/23]

2. Varley M, Gregson W, McMillan K, et al. Physical and technical performance of elite youth soccer players during international tournaments: influence of playing position and team

success and opponent quality. *Sci Med Footb* 2017;1(1):18-29. doi: 10.1080/24733938.2018.1427883

3. Yang G, Leicht AS, Lago C, et al. Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Res Sports Med* 2018;26(2):158-67. doi: 10.1080/15438627.2018.1431539 [published Online First: 2018/02/01]

4. Pyne DB, Spencer M, Mujika I. Improving the value of fitness testing for football. *Int J Sports Physiol Perform* 2014;9(3):511-4. doi: 10.1123/ijspp.2013-0453 [published Online First: 2013/11/16]

5. Murray NB, Gabbett TJ, Townshend AD. The use of relative speed zones in australian football: are we really measuring what we think we are? *Int J Sports Physiol Perform* 2018;13(4):442-51. doi: 10.1123/ijspp.2017-0148 [published Online First: 2017/09/06]

6. Roe G, Darrall-Jones J, Black C, et al. Validity of 10-HZ GPS and timing gates for assessing maximum velocity in professional rugby union players. *Int J Sports Physiol Perform* 2017;12(6):836-39. doi: 10.1123/ijspp.2016-0256 [published Online First: 2016/10/14]

7. Mendez-Villanueva A, Buchheit M, Simpson B, et al. Match play intensity distribution in youth soccer. *Int J Sports Med* 2013;34(2):101-10. doi: 10.1055/s-0032-1306323 [published Online First: 2012/09/11]

8. Haugen T, Buchheit M. Sprint running performance monitoring: methodological and practical considerations. *Sports Med* 2016;46(5):641-56. doi: 10.1007/s40279-015-0446-0 [published Online First: 2015/12/15]

9. Massard T, Eggers T, Lovell R. Peak speed determination in football: is sprint testing necessary? *Sci Med Footb* 2018;2(2):123-26. doi: 10.1080/24733938.2018.1427883

10. Dixon PM, Saint-Maurice PF, Kim Y, et al. A primer on the use of equivalence testing for evaluating measurement agreement. *Med Sci Sports Exerc* 2018;50(4):837-45. doi: 10.1249/mss.0000000000001481 [published Online First: 2017/11/15]

11. Beato M, Coratella G, Stiff A, et al. The validity and between-unit variability of GNSS Units (STATSports Apex 10 and 18 Hz) for measuring distance and peak speed in team sports. *Front Physiol* 2018;9 doi: 10.3389/fphys.2018.01288

12. Hopkins WG. Improving meta-analyses in sport and exercise science. *Sportscience* 2018;22:11-17.

13. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. *Adv Methods Pract Psychol Sci* 2018;1(2):259 - 69. doi: 10.1177/2515245918770963

14. Pek J, Flora DB. Reporting effect sizes in original psychological research: a discussion and tutorial. *Psychol Methods* 2018;23(2):208-25. doi: 10.1037/met0000126 [published Online First: 2017/03/10]

15. Thorpe RT, Atkinson G, Drust B, et al. Monitoring fatigue status in elite team-sport athletes: implications for practice. *Int J Sports Physiol Perform* 2017;12(Suppl 2):S227-s34. doi: 10.1123/ijspp.2016-0434 [published Online First: 2017/01/18]

16. Esculier JF, Barton C, Whiteley R, et al. Involving clinicians in sports medicine and physiotherapy research: 'design thinking' to help bridge gaps between practice and evidence. *Br J Sports Med* 2018 doi: 10.1136/bjsports-2018-100078 [published Online First: 2018/10/29]

17. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013;4:863. doi: 10.3389/fpsyg.2013.00863 [published Online First: 2013/12/11]

18. Duffield R, Reid M, Baker J, et al. Accuracy and reliability of GPS devices for measurement of movement patterns in confined spaces for court-based sports. *J Sci Med Sport* 2010;13(5):523-5. doi: 10.1016/j.jsams.2009.07.003 [published Online First: 2009/10/27]

19. Varley MC, Jaspers A, Helsen WF, et al. Methodological considerations when quantifying high-intensity efforts in team sport using global positioning system technology. *Int J Sports Physiol Perform* 2017;12(8):1059-68. doi: 10.1123/ijspp.2016-0534 [published Online First: 2017/01/05]

20. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31(4):337-50. doi: 10.1007/s10654-016-0149-3 [published Online First: 2016/05/23]

21. Rothman KJ. Disengaging from statistical significance. *Eur J Epidemiol* 2016;31(5):443-4. doi: 10.1007/s10654-016-0158-2 [published Online First: 2016/06/09]

22. Little R. Calibrated Bayes. *Am Stat* 2006;60(3):213-23. doi: 10.1198/000313006X117837

23. Hopkins WG, Batterham A. The vindication of magnitude-based inference. *Sportscience* 2018;22:19-29.

24. Little RJ. Calibrated Bayesian inference: a comment on The Vindication of Magnitude-Based Inference. *Sportscience* 2018;22

25. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform* 2006;1(1):50-57.

26. Lassere MN, van der Heijde D, Johnson KR. Foundations of the minimal clinically important difference for imaging. *J Rheumatol* 2001;28(4):890-1. [published Online First: 2001/05/01]

27. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26(4):217-38. [published Online First: 1998/11/20]

28. Atkinson G. Does size matter for sports performance researchers? *J Sports Sci* 2003;21(2):73-4. doi: 10.1080/0264041031000071038 [published Online First: 2003/03/13]

29. Akenhead R, Nassis GP. Training load and player monitoring in high-level football: current practice and perceptions. *Int J Sports Physiol Perform* 2016;11(5):587-93. doi: 10.1123/ijspp.2015-0331 [published Online First: 2015/10/13]

30. Al Haddad H, Simpson BM, Buchheit M, et al. Peak match speed and maximal sprinting speed in young soccer players: effect of age and playing position. *Int J Sports Physiol Perform* 2015;10(7):888-96. doi: 10.1123/ijspp.2014-0539 [published Online First: 2015/02/25]

31. Mendez-Villanueva A, Buchheit M, Simpson B, et al. Does on-field sprinting performance in young soccer players depend on how fast they can run or how fast they do run? *J Strength Cond Res* 2011;25(9):2634-8. doi: 10.1519/JSC.0b013e318201c281 [published Online First: 2011/07/20]