This is an Accepted Manuscript of an article published in Journal of Sports Sciences – accepted on 23 January 2021, available online: https://doi.org/10.1080/02640414.2021.1882726

Version: Accepted for publication

Publisher: © Taylor & Francis

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.

Journal of Sports Sciences

Inter-unit reliability of IMU Step metrics using IMeasureU Blue Trident inertial measurement units for running-based team sport tasks

M. Armitage, M. Beato, and S.A. McErlain-Naylor.

School of Health and Sports Sciences, University of Suffolk, Ipswich, IP4 1QJ, UK

ABSTRACT

The aim of this study was to determine the inter-unit reliability of IMU Step biomechanical load monitoring metrics using IMeasureU Blue Trident inertial measurement units in tasks common to running-based team sports. Knowledge of variability between units is required before researchers and practitioners can make informed decisions on 'true' differences between limbs. Sixteen male college soccer players performed five running-based tasks, generating 224 trials and 17012 steps. Data were analysed for each task and for the whole session, investigating six IMU Step metrics: step count; impact load; bone stimulus; and low, medium and high intensity steps. Inter-unit reliability was *excellent* (ICC ≥ 0.90) for 21 out of 26 metrics, and *good* (0.83 ≤ ICC ≤ 0.86) for all other metrics except for Yo-Yo impact load (ICC = 0.79) which was *acceptable*. These findings confirm the inter-unit reliability of IMU Step metrics using IMeasureU Blue Trident inertial measurement units for running-based team sports. Now that inter-unit variability has been quantified, researchers and practitioners can use this information when interpreting inter-limb differences for monitoring external biomechanical training load.

Keywords: accelerometer; bone stimulus; tibial acceleration; impact load; training load

INTRODUCTION

The term training load is common in both research and applied sport settings and is categorised as internal or external load (Impellizzeri et al., 2019). Internal load describes the body's response to the external activities performed (Cardinale & Varley, 2017). Traditionally, adaptations to training load have been quantified in relation to physiological stress (Vanrenterghem et al., 2017). However, mechanical stress also contributes to load-adaptation pathways and so training load should be considered from a physiological and biomechanical perspective (Vanrenterghem et al., 2017). To infer decisions from different forms of loading (*e.g.* internal/external, physiological/biomechanical) practitioners typically use a combined approach (Delaney et al., 2018). Global position systems (GPS) have become extremely popular tools to monitor external physiological load (*e.g.* distance covered and speed thresholds) in running-based team sports (Burgess, 2017). Many GPS providers also integrate tri-axial accelerometers into their units creating acceleration derived metrics (*e.g.* PlayerLoad^{TM} and Dynamic Stress Load) to estimate external biomechanical load (Beato et al., 2019; Verheul et al., 2020). The ability of tri-axial accelerometers within scapulae worn GPS units to capture accurate whole-body accelerations (*i.e.* external biomechanical load) has been questioned (Delaney et al., 2019). Recent evidence suggests a need to measure segmental accelerations closest to the position of interest (Greig et al., 2018; Nedergaard et al., 2017; Sheerin et al., 2019), with shank mounted accelerometry increasing in popularity for field-based tibial loading measures (Rice et al., 2018; Verheul et al., 2020; Willy, 2018).

The relationship between measured segmental accelerations and whole-body biomechanical loading is influenced by factors including the kinematics of the lower limbs at initial foot-ground-contact and acceleration attenuation between body segments (Nedergaard et al., 2017). Scapulae worn accelerometers may be oversensitive to upper body kinematics (Barrett et al., 2016), and could be distorted by the typical positioning within an elasticated harness (Edwards et al., 2019). Skin mounted tibial accelerometers are commonly used as a proxy for the impact experienced at the tibia (Sheerin et al., 2019) and are sensitive to changes in running speed (Sheerin et al., 2017), technique (Crowell & Davis, 2011), and ground reaction force loading rate (Tenforde et al., 2019). Tibial accelerometry is presently limited to surface acceleration (Vigotsky et al., 2019) and will remain a measure of external, rather than internal, load unless muscle forces are considered (Matijevich et al., 2019). Nonetheless, tibial accelerations have been used to aid clinical assessments of field-based rehabilitation amongst soccer players (Greig et al., 2018), modify running technique post-injury (Creaby & Franettovich Smith, 2016), and predict bone-stress injury in runners (Milner et al., 2006). Despite a large body of evidence using shank mounted accelerometry for field-based tibial loading measurement (Rice et al., 2018; Verheul et al., 2020; Willy, 2018), there is limited evidence regarding the reliability of such devices (Sheerin et al., 2019). While laboratory-grade accelerometers are attractive for data-driven insights, automatically generated metrics are required to meet the rapid data processing and output needs of clinicians and coaches (Davis & Gruber, 2019).

IMU Step combines tri-axial tibial accelerometer units (IMeasureU Blue Trident) with associated data processing (IMU Step dashboard) to provide automatically generated external biomechanical load metrics of step count, impact load, bone stimulus, and number of low, medium and high intensity steps. Bone stimulus is an exponentially weighted metric to model tibial response to cyclic mechanical loading. Based on previous research (Ahola et al., 2010; Beaupre & Orr, 1990) it incorporates both the number of cycles and load magnitude, being more sensitive to the latter (Besier, 2019). Impact load is the sum of the peak resultant acceleration in g from each step and is therefore directly proportional to the number and intensity of impacts.

Research using a previous IMeasureU unit model (Blue Thunder) demonstrated reliability of step peak resultant acceleration during treadmill running at different speeds at one week (90% CI: 0.90 – 0.96 ICC, *excellent*) and six month (0.89 – 0.95 ICC, *excellent*) (Sheerin et al., 2017) intervals. Recently, Burland et al. (2020) added to this using newer Blue Trident units and reported inter-session reliability for impact load (95% CI; 0.58 – 0.89 ICC, *fair* to *excellent*) and bone stimulus (0.90 – 0.97 ICC, *excellent*) metrics across three repeated sessions of sport-specific tasks. Furthermore, they analysed unilateral step counts reporting reliability outputs for acceleration-deceleration (0.73 – 0.84 ICC, *good* to *excellent*), change of direction (0.73 – 0.96 ICC, *good* to *excellent*) and cutting (0.70 – 0.87 ICC, *good* to *excellent*) tasks. Reliability values were lowest for the kicking task (0.59 – 0.68 ICC, *fair* to *good*), attributed to the inherent variability associated with this task.

Whilst these findings offer researchers and practitioners information regarding the reliability of IMU Step metrics across repeated sessions, differences in sensitivity between each capacitive based microelectromechanical systems unit may lead to inter-unit differences in measured accelerations and automatically generated metrics. Before inter-limb, and thus inter-unit, comparisons can be made, agreement between units must first be ascertained. For differences in inter-limb variation to be confidently interpreted as ‘real’ they must be greater than the known inter-unit coefficient of variation for that metric (Bishop, 2020). Furthermore, the reliability of low, medium and high intensity steps are yet to be investigated, as is reliability of any of these metrics when using the manufacturer's provided straps. This is especially important given the effect of attachment method on measured tibial accelerations (Sheerin et al., 2019) and the likelihood of practitioners using the provided and recommended attachments.

The aim of this study was therefore to determine the inter-unit reliability of IMU Step metrics (step count; impact load; bone stimulus; and low, medium and high intensity steps) during tasks common to running-based team sports. It was hypothesised that all metrics would demonstrate *good* or better inter-unit reliability.

MATERIALS AND METHODS

Participants

Sixteen male full-time college soccer academy players participated in this study (age 17 ± 1 years; mass 68.5 ± 10.4 kg; height 1.78 ± 0.06 m). Signed informed consent was given by each participant independently (age ≥ 18 years) or via ascent with parent / guardian support (age < 18 years). The study was performed in accordance with the Declaration of Helsinki for the study of human subjects and was approved by the institutional ethics board of the University of Suffolk (UK).

Data Collection

Data were collected using IMeasureU Blue Trident inertial measurement units (Vicon Motion Systems Ltd, Oxford, UK). Each unit (42 x 27 x 11 mm, 9.5 grams) incorporates two tri-axial accelerometers: one with a range of ± 16 g (1125 Hz; 16 bit resolution) to provide resolution at lower accelerations; and one with a range of ± 200 g (1600 Hz; 13 bit resolution) which is used when the first accelerometer's range is exceeded. Two IMeasureU Blue Trident units were affixed to the right distal anteromedial shank of each participant using the provided manufacturer's straps, ensuring a tight but comfortable fit (Rice et al., 2018). The first unit was positioned 20 mm proximal to the superior aspect of the medial malleolus, the mean of two previously reported positions (Rice et al., 2018; Sheerin et al., 2018). The second unit was placed superior to the first unit (Figure 1), positioned as close as possible without causing inter-unit contact during the tasks. Units were randomly allocated.

Figure 1 - positioning of IMeasureU Blue Trident sensors on the right shank.

Data were collected in an indoor hall to standardise environmental conditions. Participants completed five tasks (Figure 2) designed to replicate actions common to running-based team sports. The testing session was repeated back-to-back with three different groups (n = 6, 5, 5). Each session began with the same standardised warm-up, led by an accredited Strength and Conditioning Coach (UKSCA; >10 years of experience). Units were worn throughout the warm-up for familiarisation, but the warm-up data were not analysed.

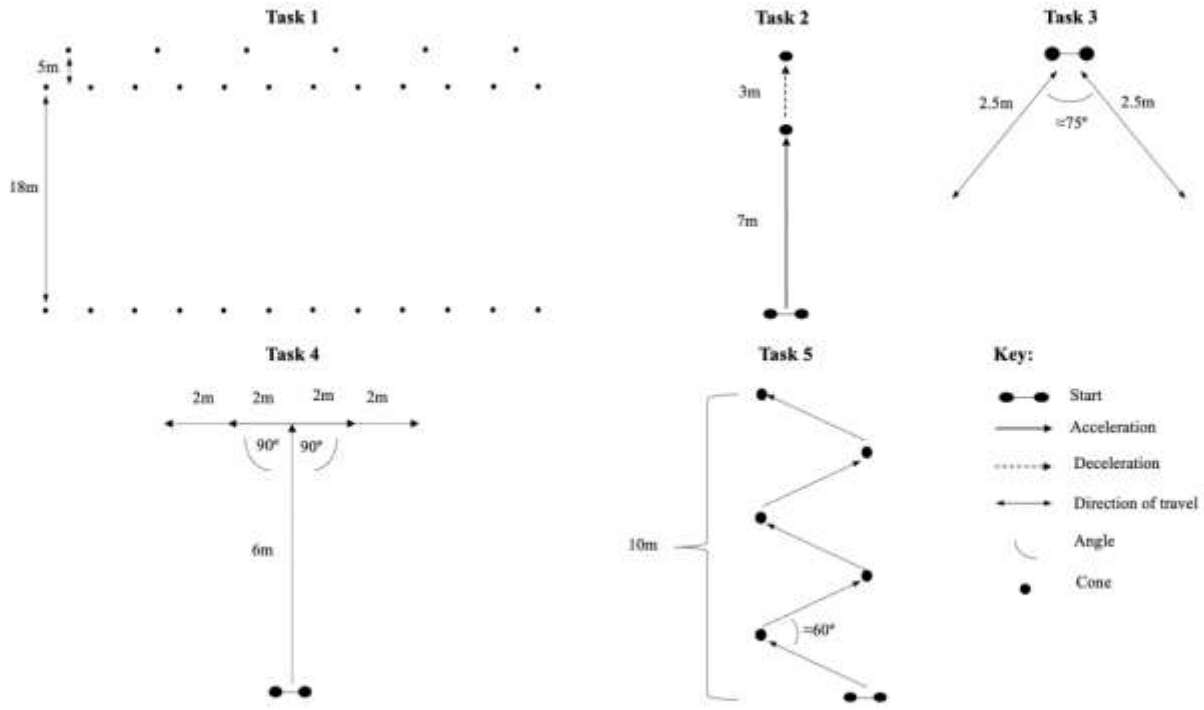

Figure 2 - diagram of the tasks.

Sport-specific tasks

Submaximal intermittent running was achieved through a modified Yo-Yo Intermittent Recovery Test Level *2* (Task 1). Participants were instructed to run back and forth between two cones 18 m apart (modified from the typical 20 m to ensure a submaximal nature) and then walk around a cone 5 m away in time with an audio ‘bleep’. The activity started at Level 13 and was terminated after 4 min (Veugelers et al., 2016). Sport-specific tasks were adapted from previous work which investigated other wearable technologies for running-based team sport tasks (Luteberget et al., 2018; Roell et al., 2019). Participants were asked to perform each sport-specific task maximally and rested for 1 min between trials and 3 min between tasks (Figure 3). Before and after each trial of each task, participants stood stationary for ~ 5 s, which in addition to the required rest periods facilitated extraction of data (*i.e.* a clear start and end point).

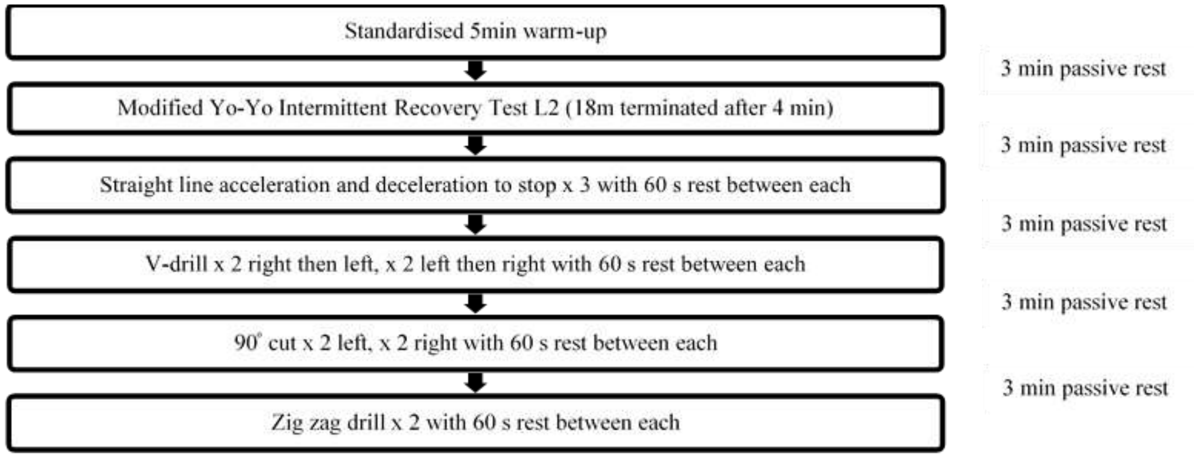

Figure 3 - details and order of tasks and trials.

Task 2 involved three trials of straight-line sprinting with a 7 m linear acceleration and 3 m deceleration zone (Figure 2). Task 3 (V-Drill) required participants to run 2.5 m from the start position at an angle of 37.5° from the forward direction on their right-hand-side and then backwards to the start, before immediately repeating on the left. Participants completed two trials on each side. Task 4 was achieved by a 6 m straight-line sprint, a 90° cut, 2 m acceleration to a cone and 2 m deceleration to the next cone. Participants completed two trials to the left and two to the right. Task 5 was a Zig-Zag running circuit consisting of two 60° cuts alternatively to left and right before arriving at the stop gate. Participants completed two trials.

Data Processing

All data were captured in real-time using the manufacturer's IOS application (app version 2.7.523). All acceleration data were downloaded after data collection using IMU Step software version 2.7.1, with footnotes added retrospectively to identify each drill. Metrics were output for all individual tasks (including inter-trial rest periods) and for the entire session (including all rest periods). The IMU Step software outputs the automatically generated metrics of step count, impact load, bone stimulus, and number of low (LIS: default threshold of peak resultant tibial acceleration ≤ 6 g), medium (6 g < MIS ≤ 21.5 g) and high (HIS > 21.5 g) intensity steps. Data from the two units per participant were randomly allocated as either unit one or unit two for subsequent statistical analysis.

Statistical Analysis

All statistical analysis was performed using JASP (Amsterdam, Netherlands) software version 0.9.2. All descriptive data were presented as mean ± standard deviation (SD). Normality of distributions were assessed by Shapiro-Wilk test ($0.074 \leq p \leq 0.998$). Inter-unit reliability was assessed for all task-metric combinations containing an average of ≥ 20 steps per participant. Inter-unit reliability was calculated by two-way mixed model intra-class correlation coefficient (ICC), interpreted as: *excellent* ≥ 0.9; 0.9 > *good* ≥ 0.8; 0.8 > *acceptable* ≥ 0.7; 0.7 > *questionable* ≥ 0.6; 0.6 > *poor* ≥ 0.5; *unacceptable* < 0.5 (Atkinson & Nevill, 1998). Technical error of measurement (TE) was calculated as $SD \cdot \sqrt{(1-ICC)}$ (Hopkins et al., 2001). Confidence intervals (CI) at 95% were reported. TE was reported as coefficient of variation (CV), considered as: *good* < 10%; 10% ≤ *questionable* ≤ 15%; *poor* > 15% (Cormack et al., 2008).

RESULTS

On average participants performed 530 steps of which 56 ± 5, 19 ± 4 and 24 ± 5% were LIS, MIS and HIS respectively. The Yo-Yo contributed the most steps across all bands with the remaining four tasks being relatively comparable (Table I). Inter-unit reliability was *excellent* (0.90 ≤ ICC ≤ 0.98) for most metrics (21 out of 26), including all step count, LIS, HIS and bone stimulus metrics (Table II). Inter-unit reliability was *good* (0.83 ≤ ICC ≤ 0.86) for all other metrics except for Yo-Yo impact load (ICC = 0.79; CI: 0.40, 0.93) which was *acceptable*. TE (CV%) was *good* (0.7% ≤ TE ≤ 9.7%) for all metrics assessed except for impact load during the overall session, Yo-Yo, sprint and Zig-Zag tasks which were *questionable* (10.8 – 14.5 %) (Table III).

DISCUSSION

The aim of this study was to determine the inter-unit reliability of IMU Step metrics (step count; impact load; bone stimulus; and low, medium and high intensity steps) during tasks common to running-based team sports. In accordance with the hypothesis, all task-metric combinations displayed *good* or *excellent* ICC except for Yo-Yo impact load which was *acceptable*. Most metrics (22 out of 26) displayed *good* CV, although impact load was *questionable* for the whole session, Yo-Yo, sprint and Zig-Zag tasks.

The present findings are comparable to previous research which reported reliability (0.89 – 0.96 ICC, *excellent*) for step peak resultant acceleration during treadmill running in a laboratory using earlier model IMeasreU Blue Thunder units (Sheerin et al., 2017). This study however, adds new IMU Step metrics, utilises updated IMeasureU Blue Trident units and involves tasks more common to team-based running sports in agreement with Burland et al. (2020). Combining the results of this study (inter-unit reliability) with those of Burland et al. (2020) (inter-session reliability) researchers and practitioners can have greater confidence when assessing step frequency, magnitude and symmetry to evaluate training load. The ICC values for running-based tasks were comparable or greater in the current study compared to Burland *et al.* (2020) for impact load (0.79 – 0.96 *vs* 0.75 – 0.89) and step count (0.91 - 0.98 *vs* 0.70 – 0.96). As mentioned by Burland *et al.* (2020), the reliability of each measure is a function of hardware reliability and movement consistency. Consistency of movement will be greater in a single trial compared to repeated session designs, perhaps providing a better measure of hardware reliability. Furthermore, whilst Burland *et al.* (2020) analysed unilateral steps by placing a unit on each tibia (in accordance with manufacturer's recommendations), any differences between units were not known. Thus, in addition to inter-session reliability, researchers and practitioners can now be confident that there is little difference between units (inter-unit reliability) in metrics derived from IMU Step software. This finding could have large potential implications for inferring differences in limb loading when evaluating training load.

Similarly to previous inter-session measures (Burland et al., 2020) bone stimulus reported the greatest inter-unit reliability of all IMU Step metrics. Due to its cumulative nature this metric considers all preceding impacts and so represents the entire session up to that time point. It is unable to differentiate between separate tasks within a session because individual tasks are dependent upon earlier loading cycles. Based on bone mechanobiology (Ahola et al., 2010; Beaupre & Orr, 1990; Besier, 2019), bone stimulus is intended to predict the mechanical stimulus responsible for bone remodelling which plateaus with repeated cycles (Besier, 2019). This results in a large increase during the first activity and continued rise with additional tasks, resulting in an overall value which is matched by the last task. The linear impact load metric provides greater indicative

Table I. Mean ± SD IMU Step metric values for steps performed throughout the data collection session (overall) and during sport-specific tasks (n = 16 players, 224 trials).

Tasks	**SC_1 and SC_2**	**LIS_1 and LIS_2**	**MIS_1 and MIS_2**	**HIS_1 and HIS_2**	**IL_1 and IL_2**	**BS_1 and BS_2**
Overall	534 ± 52	300 ± 38	102 ± 21	131 ± 35	7265 ± 2020	235 ± 9
	529 ± 44	301 ± 33	97 ± 18	130 ± 29	7086 ± 1668	235 ± 8
Yo-Yo	235 ± 15	74 ± 13	75 ± 22	86 ± 26	4487 ± 1419	N/A
	235 ± 14	76 ± 11	74 ± 21	84 ± 22	4280 ± 941	
Sprint	58 ± 15	49 ± 15	3 ± 2	6 ± 1	392 ± 138	N/A
	57 ± 13	49 ± 13	3 ± 2	5 ± 2	424 ± 171	
V-Drill	46 ± 7	29 ± 6	6 ± 3	11 ± 3	536 ± 125	N/A
	44 ± 7	27 ± 6	7 ± 4	9 ± 3	550 ± 168	
90L	30 ± 5	23 ± 4	2 ± 2	5 ± 1	324 ± 128	N/A
	31 ± 5	24 ± 4	2 ± 2	5 ± 2	332 ± 125	
90R	33 ± 8	26 ± 8	3 ± 2	5 ± 2	283 ± 137	N/A
	33 ± 8	26 ± 7	5 ± 2	4 ± 1	287 ± 125	
Zig-Zag	42 ± 9	25 ± 7	6 ± 4	11 ± 4	592 ± 179	N/A
	40 ± 7	24 ± 6	5 ± 3	11 ± 3	574 ± 155	

SC = step count, LIS = low intensity steps, MIS = medium intensity steps, HIS = high intensity steps, IL = impact load, BS = bone stimulus, SD = standard deviation, $_1$ & $_2$ = the randomly allocated unit 1 and unit 2. Note: BS is a metric for assessing entire sessions only.

Table II. IMU Step inter-unit (IMeasureU Blue Trident) reliability as calculated by intra-class coefficient (ICC) for steps performed throughout the data collection session (overall) and during sport-specific tasks (n = 16 players, 224 trials).

Tasks	**SC** **ICC (95% CI)** ***interpretation***	**LIS** **ICC (95% CI)** ***interpretation***	**MIS** **ICC (95% CI)** ***interpretation***	**HIS** **ICC (95% CI)** ***interpretation***	**IL** **ICC (95% CI)** ***Interpretation***	**BS** **ICC (95% CI)** ***interpretation***
Overall	0.96 (0.90, 0.99) *excellent*	0.95 (0.86, 0.98) *excellent*	0.86 (0.60, 0.95) *good*	0.96 (0.88, 0.98) *excellent*	0.85 (0.57, 0.95) *good*	0.97 (0.92, 0.99) *excellent*
Yo-Yo	0.91 (0.74, 0.96) *excellent*	0.95 (0.87, 0.98) *excellent*	0.94 (0.83, 0.98) *excellent*	0.96 (0.89, 0.99) *excellent*	0.79 (0.40, 0.93) *acceptable*	N/A
Sprint	0.98 (0.93, 0.99) *excellent*	0.97 (0.92, 0.99) *excellent*	Not calculated	Not calculated	0.90 (0.72, 0.97) *excellent*	N/A
V-Drill	0.98 (0.94, 0.99) *excellent*	0.98 (0.95, 0.99) *excellent*	Not calculated	Not calculated	0.83 (0.51, 0.94) *good*	N/A
90L	0.94 (0.83, 0.98) *excellent*	0.94 (0.82, 0.97) *excellent*	Not calculated	Not calculated	0.96 (0.91, 0.99) *excellent*	N/A
90R	0.98 (0.95, 0.99) *excellent*	0.98 (0.96, 0.99) *excellent*	Not calculated	Not calculated	0.96 (0.90, 0.98) *excellent*	N/A
Zig-Zag	0.95 (0.88, 0.98) *excellent*	0.97 (0.92, 0.99) *excellent*	Not calculated	Not calculated	0.84 (0.55, 0.94) *good*	N/A

SC = step count, LIS = low intensity steps, MIS = medium intensity steps, HIS = high intensity steps, IL = impact load, BS = bone stimulus, CI = confidence interval. Note: BS is a metric for assessing entire sessions only. Inter-unit reliability was assessed for all task-metric combinations containing an average of ≥ 20 steps per participant (Table 1).

Table III. IMU Step inter-unit (IMeasureU Blue Trident) reliability as calculated by technical error of measurement (TE) and coefficient of variation (CV) for steps performed throughout the data collection session (overall) and during sport-specific tasks (n = 16 players, 224 trials).

Variables	**SC** **TE (CV%)** ***interpretation***	**LIS** **TE (CV%)** ***interpretation***	**MIS** **TE (CV%)** ***interpretation***	**HIS** **TE (CV%)** ***interpretation***	**IL** **TE (CV%)** ***interpretation***	**BS** **TE (CV%)** ***interpretation***
Overall	10.4 (1.9%) *good*	8.5 (2.8%) *good*	7.9 (7.7%) *good*	7.0 (5.3%) *good*	782 (10.8%) *questionable*	1.6 (0.7%) *good*
Yo-yo	4.5 (1.9%) *good*	2.9 (3.9%) *good*	5.4 (7.2%) *good*	5.2 (6.0%) *good*	650 (14.5%) *questionable*	N/A
Sprint	2.1 (3.7%) *good*	2.6 (5.3%) *good*	Not calculated	Not calculated	43.6 (11.1%) *questionable*	N/A
Vdrill	1.0 (2.2%) *good*	0.8 (2.9%) *good*	Not calculated	Not calculated	51 (9.6%) *good*	N/A
90L	1.2 (4.1%) *good*	1.0 (4.3%) *good*	Not calculated	Not calculated	25 (7.9%) *good*	N/A
90R	1.1 (3.4%) *good*	1.1 (4.4%) *good*	Not calculated	Not calculated	27 (9.7%) *good*	N/A
Zig-Zag	2.0 (4.8%) *good*	1.2 (4.8%) *good*	Not calculated	Not calculated	71 (12.1%) *questionable*	N/A

SC = step count, LIS = low intensity steps, MIS = medium intensity steps, HIS = high intensity steps, IL = impact load, BS = bone stimulus.

Note: BS is a metric for assessing entire sessions only. Inter-unit reliability was assessed for all task-metric combinations containing an average of ≥ 20 steps per participant (Table 1).

insights within sessions because it is calculated by summing the peak acceleration of each step (*e.g.* number of steps x 1 g) + number of steps x 2 g + . . . number of steps x *n* g). It is therefore unaffected by loading earlier in the session and so can be split to enable task level analysis. Greater impact loads are caused by either higher magnitude impacts and/or a greater number of impacts. In this study, impact load demonstrated *acceptable* to *excellent* (0.79 – 0.96) ICC and *questionable* to *good* (7.9 – 14.5%) CV which was lower than other metrics. To investigate between-device agreement units were positioned as close as possible without causing contact, to limit a known attenuation effect along the tibia (Lucas-Cuevas et al., 2017). Any attenuation of acceleration signals between the two units would logically have the greatest effect on impact load metrics. Whilst other metrics count impacts (*e.g.* step count), categorise impacts into large ordinal 'bins' (*e.g.* LIS, MIS and HIS), or plateau with increasing load (*e.g.* bone stimulus), impact load is sensitive to small differences in peak resultant accelerations which are summed each step and thus more prone to error. It was not possible to place both units in exactly the same position on the tibia, although such a true measure of inter-unit reliability would likely result in greater ICC and lower CV values than those reported in the present study due to the removal of signal attenuation artefacts.

This study is the first to report reliability data for the automatically 'binned' IMU Step metrics describing step intensity. Reliability was *excellent* for LIS (0.95 ICC), MIS (0.94 ICC) and HIS (0.96 ICC) during the Yo-Yo task, and *good* to *excellent* for the session overall (LIS 0.95, *excellent*; MIS 0.86, *good*; HIS 0.96, *excellent*). The lower ICC value for MIS overall may be partly explained by the selected tasks facilitating more LIS and HIS. Nevertheless, these findings suggest that as well as low magnitude accelerations, IMU Step is reliable for measuring medium and higher magnitude (> 6 g) intermittent acceleration and deceleration activities. Future research should confirm this finding during discrete high acceleration tasks (*e.g.* sprinting, cutting and changing direction) for which the MIS and HIS step counts in the present study were insufficient to enable task level analysis other than for the Yo-Yo.

Researchers and practitioners can be confident that there is little variation between IMeasureU Blue Trident units in metrics derived from IMU Step software. As such inter-limb comparisons, using automatically generated metrics as arbitrary measures (Hughes et al., 2019), can now be considered. Researchers and practitioners can make decisions regarding inter-limb asymmetry in direct relation to the presently reported magnitudes of inter-unit reliability. Specifically, inter-limb variation in IMU Step metrics should only be considered indicative of asymmetry if they are greater than the reported inter-unit CV for that metric (Bishop, 2020). Future research should establish what magnitude of asymmetry, beyond the now known inter-unit variation, could be deemed clinically meaningful (Harrison et al., 2020).

The reliability found in this study is similar to those reported for back-worn GPS embedded accelerometers using a similar protocol (Roell et al., 2019). However, GPS units worn at the torso only provide an indirect measure of the mechanical loads experienced at the lower limbs (Glassbrook et al., 2020). *Poor* to *questionable* reliability and high variability has been reported when comparing trunk worn GPS accelerometers to laboratory methods (Edwards et al., 2019). Differences between systems should be expected due to variations in unit location and specification such as capture frequency, sensitivity, or resolution (Edwards et al., 2019; Glassbrook et al., 2020). GPS-integrated tri-axial accelerometers typically capture data at 100 Hz (Malone et al., 2017) with laboratory-grade accelerometers and IMeasureU units (1125 to 1600 Hz) possessing higher sampling frequencies (Sheerin et al., 2019). The combined use of both technologies could give greater insights into training load management (*e.g.* asymmetry in impact load reported within specific ranges of running speeds in representative sporting environments), compared to using each technology independently (Glassbrook et al., 2020).

In this study, data were automatically processed within the manufacturer's IMU Step software to investigate the entire biomechanical load monitoring system (hardware + software) and enhance applicability to researchers and practitioners using automated outputs. The calculation of metrics based on peak resultant acceleration per step were explained previously, whereas processing of raw acceleration signals prior to extraction of peak values (*e.g.* the filtering method used) are unknown and may be explored as part of future validation research. IMU Step enables the user to export raw acceleration data, which might further enhance reliability through manual processing and selection of filters or 'intensity' thresholds (Malone et al., 2017). Any effects of high-frequency noise or filter selection will be included within the present inter-unit reliability analysis. Likewise, whilst damping effects of footwear are unlikely to have affected the within-limb comparisons, standardised footwear may be considered within future research designs. Now that favourable inter-unit reliability has been reported for automated metrics derived by IMU Step software using IMeasureU Blue Trident inertial measurement units, research establishing the validity of these metrics is necessary. If validated, they could provide researchers and practitioners with useful insights into external biomechanical training load. Whilst the reliable bone stimulus metric is based upon the mechanobiology of bone response to loading (Ahola et al., 2010; Beaupre & Orr, 1990), information regarding muscle activation will be necessary to model the adaptation of muscle and tendon to their mechanical environment (Young et al., 2016).

CONCLUSION

IMU Step is a biomechanical load monitoring system that uses tri-axial tibial accelerometer units on each leg to support in the quantification of lower limb loading in the field through automatically generated metrics (step count; impact load; bone stimulus; and low, medium and high intensity steps). Knowledge of agreement between units was required to enable researchers and practitioners to make informed decisions on differences between limbs. This study is the first to report such data. All task-metric combinations displayed *good* or *excellent* intra-class correlation coefficient, except for Yo-Yo impact load which was *acceptable*. Most metrics (22 out of 26) displayed *good* coefficient of variation, although impact load was *questionable* for the whole session, Yo-Yo, sprint and Zig-Zag tasks. These findings confirm the inter-unit reliability of IMU Step metrics for running-based team sports. Inter-unit and hence inter-limb comparisons can now be made with reference to known levels of inter-unit reliability.

REFERENCES

Ahola, R., Korpelainen, R., Vainionpää, A., & Jämsä, T. (2010). Daily impact score in long-term acceleration measurements of exercise. *Journal of Biomechanics*, *43*(10), 1960–1964. https://doi.org/10.1016/j.jbiomech.2010.03.021

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine (Auckland, N.Z.)*, *26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Barrett, S., Midgley, A. W., Towlson, C., Garrett, A., Portas, M., & Lovell, R. (2016). Within-match playerload^{TM} patterns during a simulated soccer match: Potential implications for unit positioning and fatigue management. *International Journal of*

Sports Physiology and Performance, *11*(1), 135–140. https://doi.org/10.1123/ijspp.2014-0582

Beato, M., De Keijzer, K. L., Carty, B., & Connor, M. (2019). Monitoring fatigue during intermittent exercise with accelerometer-derived metrics. *Frontiers in Physiology*, *10*(JUN). https://doi.org/10.3389/fphys.2019.00780

Beaupre, G. S., & Orr, T. E. (1990). An approach for time-dependent bone modeling and remodeling—application: A preliminary remodeling simulation - Beaupré - 2005 - Journal of Orthopaedic Research - Wiley Online Library. *Journal of Orthopaedic …*, 662–670. https://doi.org/10.1167/iovs.13-11708

Besier, T. (2019). *Bone Stimulus*. https://imeasureu.com/wp-content/uploads/2019/10/IMU_Bone_Stimulus.pdf

Bishop, C. (2020). Interlimb Asymmetries. *Strength and Conditioning Journal*, 1. https://doi.org/10.1519/SSC.0000000000000554

Burgess, D. J. (2017). The research doesn't always apply: Practical solutions to evidence-based training-load monitoring in elite team sports. *International Journal of Sports Physiology and Performance*, *12*, 136–141. https://doi.org/10.1123/ijspp.2016-0608

Burland, J. P., Outerleys, J. B., Lattermann, C., & Davis, I. S. (2020). Reliability of wearable sensors to assess impact metrics during sport-specific tasks. *Journal of Sports Sciences*, *00*(00), 1–6. https://doi.org/10.1080/02640414.2020.1823131

Cardinale, M., & Varley, M. C. (2017). Wearable training-monitoring technology: Applications, challenges, and opportunities. *International Journal of Sports Physiology and Performance*, *12*, 55–62. https://doi.org/10.1123/ijspp.2016-0423

Cormack, S. J., Newton, R. U., McGulgan, M. R., & Doyle, T. L. A. (2008). Reliability of measures obtained during single and repeated countermovement jumps. *International Journal of Sports Physiology and Performance*, *3*(2), 131–144. https://doi.org/10.1123/ijspp.3.2.131

Creaby, M. W., & Franettovich Smith, M. M. (2016). Retraining running gait to reduce tibial loads with clinician or accelerometry guided feedback. *Journal of Science and Medicine in Sport*, *19*(4), 288–292. https://doi.org/10.1016/j.jsams.2015.05.003

Crowell, H. P., & Davis, I. S. (2011). Gait retraining to reduce lower extremity loading in runners. *Clinical Biomechanics*, *26*(1), 78–83. https://doi.org/10.1016/j.clinbiomech.2010.09.003

Davis, J. J., & Gruber, A. H. (2019). Quantifying exposure to running for meaningful insights into running-related injuries. *BMJ Open Sport and Exercise Medicine*, *5*(1), 1–4. https://doi.org/10.1136/bmjsem-2019-000613

Delaney, J. A., Duthie, G. M., Thornton, H. R., & Pyne, D. B. (2018). Quantifying the relationship between internal and external work in team sports: development of a novel training efficiency index. *Science and Medicine in Football*, *2*(2), 149–156. https://doi.org/10.1080/24733938.2018.1432885

Delaney, J. A., Wileman, T. M., Perry, N. J., Thornton, H. R., Moresi, M. P., & Duthie, G. M. (2019). The validity of a global navigation satellite system for quantifying small-area team-sport movements. *Journal of Strength and Conditioning Research*, *33*(6), 1463–1466. https://doi.org/10.1519/jsc.0000000000003157

Edwards, S., White, S., Humphreys, S., Robergs, R., & O'Dwyer, N. (2019). Caution using data from triaxial accelerometers housed in player tracking units during running. *Journal of Sports Sciences*, *37*(7), 810–818. https://doi.org/10.1080/02640414.2018.1527675

Glassbrook, D. J., Fuller, J. T., Alderson, J. A., & Doyle, T. L. A. (2020). Foot accelerations are larger than tibia accelerations during sprinting when measured with inertial measurement units. *Journal of Sports Sciences*, *38*(3), 248–255. https://doi.org/10.1080/02640414.2019.1692997

Greig, M., Emmerson, H., & McCreadie, J. (2018). Is There a Role for GPS in Determining Functional Ankle Rehabilitation Progression Criteria? A Preliminary Study. *J Sport Rehabil*, 1–19.

Harrison, A. J., McErlain-Naylor, S. A., Bradshaw, E. J., Dai, B., Nunome, H., Hughes, G. T. G., Kong, P. W., Vanwanseele, B., Vilas-Boas, J. P., & Fong, D. T. P. (2020). Recommendations for statistical analysis involving null hypothesis significance testing. *Sports Biomechanics*, *00*(00), 1–7. https://doi.org/10.1080/14763141.2020.1782555

Hopkins, W. G., Schabort, E. J., & Hawley, J. A. (2001). Reliability of power in physical performance tests. *Sports Medicine (Auckland, N.Z.)*, *31*(3), 211–234. https://doi.org/10.2165/00007256-200131030-00005

Hughes, T., Jones, R. K., Starbuck, C., Sergeant, J. C., & Callaghan, M. J. (2019). The value of tibial mounted inertial measurement units to quantify running kinetics in elite football (soccer) players. A reliability and agreement study using a research orientated and a clinically orientated system. *Journal of Electromyography and Kinesiology*, *44*(October 2018), 156–164. https://doi.org/10.1016/j.jelekin.2019.01.001

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance*, *14*(2), 270–273. https://doi.org/10.1123/ijspp.2018-0935

Lucas-Cuevas, A. G., Encarnación-Martínez, A., Camacho-García, A., Llana-Belloch, S., & Pérez-Soriano, P. (2017). The location of the tibial accelerometer does influence impact acceleration parameters during running. *Journal of Sports Sciences*, *35*(17), 1734–1738. https://doi.org/10.1080/02640414.2016.1235792

Luteberget, L. S., Spencer, M., & Gilgien, M. (2018). Validity of the Catapult ClearSky T6 local positioning system for team sports specific drills, in indoor conditions. *Frontiers in Physiology*, *9*(APR), 1–10. https://doi.org/10.3389/fphys.2018.00115

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the black box: Applications and considerations for using gps devices in sport. *International Journal of Sports Physiology and Performance*, *12*, 18–26. https://doi.org/10.1123/ijspp.2016-0236

Matijevich, E. S., Branscombe, L. M., Scott, L. R., & Zelik, K. E. (2019). Ground reaction force metrics are not strongly correlated with tibial bone load when running across speeds and slopes: Implications for science, sport and wearable tech. *PLoS ONE*, *14*(1), 1–19. https://doi.org/10.1371/journal.pone.0210000

Milner, C. E., Ferber, R., Pollard, C. D., Hamill, J., & Davis, I. S. (2006). Biomechanical factors associated with tibial stress fracture in female runners. *Medicine and Science in Sports and Exercise*, *38*(2), 323–328. https://doi.org/10.1249/01.mss.0000183477.75808.92

Nedergaard, N. J., Robinson, M. A., Eusterwiemann, E., Drust, B., Lisboa, P. J., & Vanrenterghem, J. (2017). The relationship between whole-body external loading and body-worn accelerometry during team-sport movements. *International Journal of Sports Physiology and Performance*, *12*(1), 18–26. https://doi.org/10.1123/ijspp.2015-0712

Rice, H. M., Saunders, S. C., McGuire, S. J., O'Leary, T. J., & Izard, R. M. (2018). Estimates of Tibial Shock Magnitude in Men and Women at the Start and End of a Military Drill Training Program. *Military Medicine*, *183*(9–10), e392–e398. https://doi.org/10.1093/milmed/usy037

Roell, M., Mahler, H., Lienhard, J., Gehring, D., Gollhofer, A., & Roecker, K. (2019). Validation of wearable sensors during team sport-specific movements in indoor environments. *Sensors (Switzerland)*, *19*(16), 1–13. https://doi.org/10.3390/s19163458

Sheerin, K. R., Besier, T. F., & Reid, D. (2018). The influence of running velocity on resultant tibial acceleration in runners. *Sports Biomechanics*, *December*, 1–11. https://doi.org/10.1080/14763141.2018.1546890

Sheerin, K. R., Besier, T. F., Reid, D., & Hume, P. A. (2017). The one-week and six-month reliability and variability of three-dimensional tibial acceleration in runners. *Sports Biomechanics*, *17*(4), 1–10. https://doi.org/10.1080/14763141.2017.1371214

Sheerin, K. R., Reid, D., & Besier, T. F. (2019). The measurement of tibial acceleration in runners—A review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use. *Gait and Posture*, *67*, 12–24. https://doi.org/10.1016/j.gaitpost.2018.09.017

Tenforde, A. S., Hayano, T., Jamison, S., Outerleys, J., & Davis, I. S. (2019). Tibial acceleration measured from wearable sensors is associated with loading rates in injured runners. *Pm&R*, 1–6. https://doi.org/10.1002/pmrj.12275

Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A., & Drust, B. (2017). Training Load Monitoring in Team Sports: A Novel Framework Separating Physiological and Biomechanical Load-Adaptation Pathways. *Sports Medicine*, *47*(11), 2135–2142. https://doi.org/10.1007/s40279-017-0714-2

Verheul, J., Nedergaard, N. J., Vanrenterghem, J., & Robinson, M. A. (2020). Measuring biomechanical loads in team sports – from lab to field. *Science and Medicine in Football*, *00*(00), 1–7. https://doi.org/10.1080/24733938.2019.1709654

Veugelers, K. R., Naughton, G. A., Duncan, C. S., Burgess, D. J., & Graham, S. R. (2016). Validity and Reliability of a Submaximal Intermittent Running Test in Elite Australian Football Players. *Journal of Strength and Conditioning Research*, *30*(12), 3347–3353. https://doi.org/10.1519/JSC.0000000000001441

Vigotsky, A. D., Zelik, K. E., Lake, J., & Hinrichs, R. N. (2019). *Mechanical Misconceptions*. *2018*. https://doi.org/10.31236/osf.io/5648m

Willy, R. W. (2018). Innovations and pitfalls in the use of wearable devices in the prevention and rehabilitation of running related injuries. *Physical Therapy in Sport*, *29*, 26–33. https://doi.org/10.1016/j.ptsp.2017.10.003

Young, S. R., Gardiner, B., Mehdizadeh, A., Rubenson, J., Umberger, B., & Smith, D. W. (2016). Adaptive Remodeling of Achilles Tendon: A Multi-scale Computational Model. *PLoS Computational Biology*, *12*(9), 1–30. https://doi.org/10.1371/journal.pcbi.1005106