Validity and reliability of smartphone orientation measurement to quantify dynamic balance function

By: Nikita A. Kuznetsov, Rebecca K. Robbins, Benjamin Long, Jason T. Jakiela, F. Jay Haran, Scott E. Ross, W. Geoffrey Wright, and Christopher K. Rhea

Kuznetsov, N.A., Robins, R.K., Long, B., Jakiela, J.T., Haran, F.J., Ross, S.E., Wright, G.W., & Rhea, C.K. (2018). Validity and reliability of smartphone orientation measurement to quantify dynamic balance function. *Physiological Measurement, 39*(2), 02NT01

Made available courtesy of IOP Publishing: https://doi.org/10.1088/1361-6579/aaa3c2

Abstract:

Objective: Postural control is frequently compromised after sub-concussive and concussive head trauma, and balance testing is an integral part of neuromotor assessment and management. The main objective of this paper is to develop a novel smartphone-based neuromotor assessment protocol for screening of dynamic balance decrements stemming from head trauma. *Approach*: Experiments 1 and 2 compared Android smartphone orientation detection algorithms to a biomechanics laboratory motion capture system using a pendulum (i.e. non-biological movement) and a human stepping task (i.e. biological movement). Experiment 3 examined the test-retest reliability of a stepping-in-place protocol in three different sensory conditions (eyes open, no-vision, head shake) using temporal and spatial variability metrics extracted from thigh orientation signal in a sample of healthy young adults. *Main results*: Smartphone sensors provided valid measurements of movement timing and amplitude variables. However, sensor firmware version and Android OS version significantly affected quality of measurement. High test-retest reliability was shown for the temporal and spatial variables of interest during the stepping-in-place task. *Significance*: Collectively, these experiments show that our smartphone application is a valid and reliable way to measure leg movement characteristics (mean stride time and its variability (CV), Peak Thigh SD, Thigh ROM, and Peak Return Velocity) during dynamic balance activity, which could provide an objective way to assess neuromotor function after head trauma and in other populations with balance dysfunction.

Keywords: smartphone sensors | measurement | gait | variability | reliability | validity | intraclass correlation

Article:

Introduction

Recent work has focused on the development of portable sensor-based balance assessment protocols that could be used by clinicians in the field to screen for neuromotor symptoms of sport-related concussion to allow making evidence-based return-to-play decisions and to track recovery of neuromotor function (McCrory *et al* 2013). Rapid and objective screening in the field is important because only a small number of concussive incidents are clearly identifiable based on visual observation (e.g. loss of consciousness) and a large number of sport concussions have subtle effects that are difficult to identify objectively (Parker *et al* 2008). Moreover, sub-concussive head trauma has received more attention in recent years because it is more prevalent than concussive head trauma (McKee *et al* 2009, Gysland *et al* 2012, Talavage *et al* 2014, Abbas *et al* 2015, Poole *et al* 2015). For example, male collegiate football players receive approximately 1000 head impacts throughout a season and only a very small number of them lead to a concussion, classifying the majority of head trauma as sub-concussive (Gysland *et al* 2012). Both the short-term (Rhea *et al* 2017) and long-term (Gavett *et al* 2011) negative consequences of sub-concussive and concussive trauma on health-related behavior have been identified, justifying the need to develop better ways to assess and track the effects of head trauma.

Objective screening tools that can be implemented in the field would be beneficial to help with screening for neuromotor dysfunction from head trauma. Objective assessments offer the potential for more reliable assessment relative to subjective assessments, which are commonly used due to their ease of administration. To this end, there has been an increase in the number of sensor-based balance assessments available to the research and clinical community, fueled by widespread access to portable technology such as smartphones and tablets, along with an improvement in the quality of sensors available in these devices (Chen *et al* 2012). As an example of an instrumented portable balance test, SWAY Medical (Cleveland, OH) developed a comprehensive neurocognitive assessment that includes a static balance test with an iPhone (Apple Inc., CA) positioned around the thoracic region to record center of gravity fluctuations during quiet stance in different stance positions, such as feet together, single leg, and tandem stance (Amick *et al* 2015). BTracks (San Diego, CA) has developed a portable force plate (O'Connor *et al* 2016) at a fraction of the cost of research-grade systems that allows to quantify center of pressure variability in the field setting (Goble *et al* 2016). Several research groups have also developed sensor-instrumented versions of the BESS (Alberts *et al* 2015, Alsalaheen *et al*2015).

One limitation of these portable balance assessment protocols is the over-reliance on testing of static postural control—maintenance of a fixed posture as still as possible in the absence of other movements. While static balance assessment is valuable, dynamic balance during activities such as walking or crossing over obstacles may be more sensitive to neuromotor symptoms of concussion (Basford *et al* 2003, Chou *et al* 2004) and is more functionally relevant, as most sport-related concussive injuries happen during dynamic activities performed by the athletes. Thus, there is a need to develop head trauma screening tools that focus on dynamic balance activities (Johnston *et al*2016).

In this study, we developed a smartphone app to track thigh and trunk motion during a dynamic balance test that consists of stepping-in-place in three sensory-probing conditions (eyes open, no-vision, and head shake) to evaluate neuromotor control of dynamic balance in the field setting using minimal space. The rationale for the sensory manipulations was to emphasize proprioceptive and vestibular perturbations to the postural control system, as previous studies have indicated presence of visual and vestibular sensory deficits in individuals after a concussion (Alsalaheen *et al* 2010, Gottshall and Hoffer 2010, Ellis *et al* 2015, Wright *et al* 2017). The task is similar to the Fukuda test, except we do not focus on trunk rotation and positional displacement, as these variables have been shown to be invalid for the assessment of peripheral vestibular dysfunction (Honaker *et al* 2009). Our focus is on characterizing temporal and kinematic variability of the leg and trunk movement during the dynamic balance task because motor variability (both magnitude and structure) is a commonly-used marker of neurological dysfunction (Newell and James 2008). To measure movement variability during this dynamic balance task, we developed a smartphone application (AccWalker) that quantifies thigh and trunk orientation.

The aim of this paper is to validate the AccWalker in comparison to laboratory motion capture equipment (Experiments 1 and 2) and to report test-retest reliability of the testing protocol in a sample of healthy young adults (Experiment 3). Specifically, experiment 1 evaluated AccWalker's performance to detect phone orientation in comparison to a research-grade motion capture system using pendulum movement, similar to previously used methods (Godfrey *et al* 2007).Two different versions of Android OS (4.4.4 and 5.1) were tested to check for any alterations in orientation estimation as a function of OS. We hypothesized that the AccWalker would produce valid measurement of the pendulum angle with respect to motion capture and that the two versions of Android OS would provide equivalent measurements. Experiment 2 examined validity of the AccWalker to measure temporal and spatial variables of thigh and trunk motion in the context of the stepping-in-place task. We hypothesized that AccWalker and 3D motion capture would provide similar estimates of thigh and trunk motion and used the Bland–Altman limits of agreement test (Bland and Altman 1999) and ICC(3,k) to test this hypothesis. In addition, we evaluated the effects of placing the phone away from its ideal reference position on the thigh on the measurements provided by the app. We also examined the ability of the phone to detect stride time variability during treadmill walking because temporal metrics are of primary interest for our future neuromotor assessments of individuals with concussion. Experiment 3 evaluated test-retest reliability of AccWalker metrics, with the hypothesis that the stepping-in-place protocol would show minimal practice effects with high test-retest reliability within each condition.

Method

Experiment 1: Android orientation sensor validation

A physical pendulum was constructed from a square poplar wood plank ($L = 95$ cm, $m = 57$ g) attached to a wheel bearing at the pivot point (figure 1(A)). Motion capture markers (5 g) were placed on the pendulum's arm and the pivot point. The phone (Motorola Moto X2 XT1095, 144 g) was placed at the end of the pendulum arm. The pendulum was released from an angle of 30° and the resulting oscillation was recorded for 60 s. Motion capture data were used to calculate pendulum angle, θ, with respect to the vertical as

$$\theta = acos\left(\frac{\boldsymbol{A} \cdot \boldsymbol{B}}{\|\boldsymbol{A}\| \, \|\boldsymbol{B}\|}\right), \tag{1}$$

where **A** is the vertical 2D vector starting at the pivot and pointing straight down and **B** is the vector pointing from the pivot to the marker on the pendulum's arm.

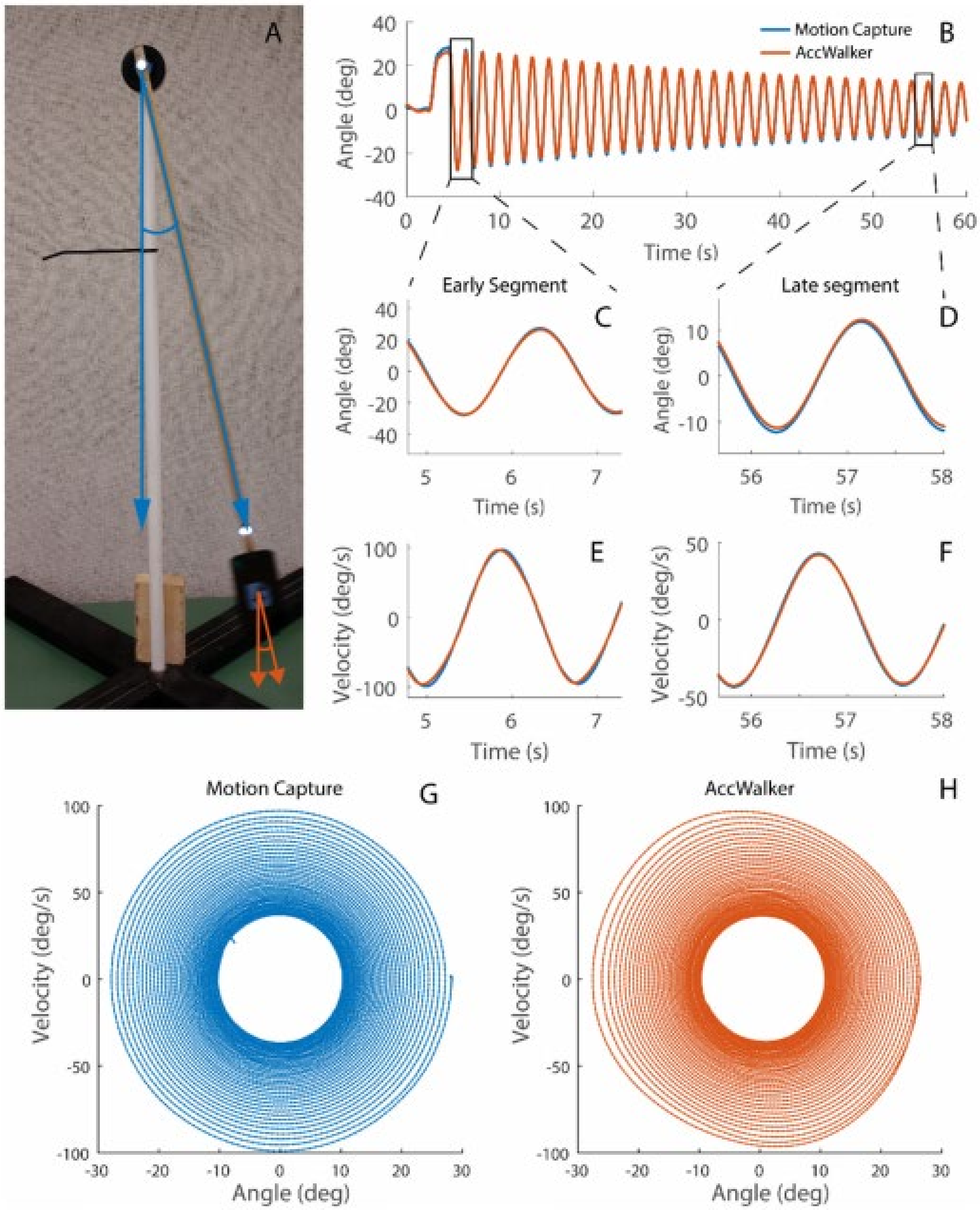

Figure 1. Testing of the AccWalker to detect pendulum angle in comparison to Qualisys motion capture system.

Phone orientation was estimated using the Rotation Vector function from Android SDK 4.4 W.2 API 19. The AccWalker app was installed on Motorola X2 XT1095 (Schaumburg, IL) because it was relatively inexpensive and has sensors required by the Rotation Vector function: a 3-axis accelerometer, a 3-axis gyroscope (InvenSense Inc., MPU-6515 MEMS, San Jose, CA), and a 3-axis magnetometer (Asahi Kasei Corp, AK8963, Tokyo, Japan). Our primary consideration for selecting the phone was that the accelerometer had a dynamic range of ±8 g and gyroscope ±1000 ° s^{-1}. Details of orientation estimation and the JAVA code to implement it are provided in supplementary material (stacks.iop.org/PM/39/02NT01/mmedia).

3D motion capture system (Qualisys, Gothenburg, Sweden) data were sampled at 100 Hz, while the AccWalker data were sampled at approximately 100.86 Hz due to asynchronies inherent to the Android sensor framework. AccWalker recordings were interpolated and resampled at 100 Hz using cubic spline interp1.m in Matlab 2016b (Mathworks, Natick, MA). Both signals were filtered using the 4th-order 5 Hz low-pass Butterworth filter and angular velocity was calculated using the 3-point formula. Motion capture and phone recordings were time-synchronized using velocity spike resulting from a finger tap on the phone prior to trial onset.

Three trials of pendulum oscillation were used to compare the performance of the AccWalker app running on Android 4.4.4 and 3D motion capture. Stock KitKat 4.4.4 OS was downloaded from the XDA Developers forum and installed on the phone using TWRP software. Three additional trials were performed using the same phone after updating the OS and sensor framework to Android 5.1.

Experiment 2: concurrent validity of AccWalker and 3D motion capture during stepping-in-place and treadmill walking

Participants

A convenience sample of nine healthy young adults (mean age 25.12 ± 2.86 yrs; eight men) took part in the study after signing an IRB-approved consent form at the University of North Carolina at Greensboro.

Materials

Two identical smartphones (Motorola Moto X2; Android OS 4.4.4) were used to measure orientation of the right thigh and upper trunk (figure 3). The leg phone measured absolute thigh segment angle with respect to the vertical in the anterior–posterior (AP) plane. The trunk phone measured orientation of the upper trunk with respect to the vertical in the medial-lateral (ML) plane. The leg phone was secured using a phone strap (Belkin, Playa Vista, CA) and the trunk phone was secured using a chest mount (Velocity Clip, Richmond, CA). Phone orientation methods and data processing were the same as in Experiment 1.

Motion capture markers were placed on the skin over greater trochanter, knee, lateral malleolus, L4, and T12. Absolute thigh segment angle was calculated using equation (1), where **A** was a 2D vector in the sagittal plane starting at the greater trochanter marker and ending at a point straight down from the trochanter marker (this point was determined by offsetting the z-coordinate of the trochanter marker by 0.1 m) and **B** was a vector starting at the greater trochanter marker and ending at the knee marker. The ML trunk angle was defined as the angle between the 2D vector in the coronal plane connecting L4 to T12 and the vertical vector staring at L4 and pointing up. Simultaneous recordings from the motion capture and phone were time-synchronized based on the first thigh flexion peak during the trial.

Procedures

Each participant performed two stepping-in-place trials with the instruction to synchronize each step to an auditory metronome (period = 0.575 s) for the first 10 s and continue stepping at the same pace for 60 s. Participants were asked to use comfortable range of motion (ROM) at the hip and knee, to lift the foot fully off the ground, and to maintain visual fixation on the target located 1.5 m in front of them at the eye level. An additional stepping-in-place trial was performed with the phone shifted anteriorly on the thigh (~4–5 cm) to simulate the effects of improper phone placement on the calculation of temporal and kinematic variables of the leg movement (figure 3(C)). Participants were also recorded walking on a treadmill at 1.34 m s^{-1}. Each trial was performed immediately after the previous one.

Dependent measures

While there are a number of metrics that could be derived from the time series recorded by the smartphone, we purposefully chose a set of metrics based on previous postural control and gait research. Due to the recognition that neuromotor dysfunction can resonate in both the temporal and spatial domains (Grabiner *et al* 2001), we included both types of metrics in our approach. Temporal metrics have a vast literature that suggests gait timing can be used as an indicator of neuromotor dysfunction (Hausdorff 2007, Verghese *et al* 2009, Stergiou and Decker 2011, Rhea and Kiefer 2014). Spatial metrics that are commonly examined are step length, step width, and center-of-mass movement. Due to the constraint of having a single phone on the thigh and the structure of the dynamic balance task, we were not able to derive these commonly used spatial metrics. We did, however, include a number of other potentially useful spatial metrics, such as the ROM, which decreases with age (Kang and Dingwell 2008), and standard deviation of the ROM, which increased after traumatic brain injury (TBI) (Buster *et al* 2013).

Temporal metrics

Stride time was identified based on maximal thigh flexion. Average stride time was used to characterize how well the participants maintained metronome pace throughout the trial (target stride time was 1.15 s). Drift in the stride time throughout the trial (Pace Drift) was quantified as the absolute difference of the average stride time during the first and last 5 s of the trial. Coefficient of variation (CV) and autocorrelation at lag 1 (ACF1) were used to characterize the magnitude and structure of stride time variability, respectively.

Spatial metrics

Spatial variability was characterized using the standard deviation of (1) phone angle at the peak thigh flexion (Peak Thigh SD), (2) peak velocity during leg lift (Lift Velocity SD), and (3) peak velocity during leg return (Return Velocity SD). Thigh range of motion (Thigh ROM) was quantified as the difference between the average phone angle during stance and the average phone angle at maximum thigh flexion.

Trunk movement

Variability of trunk orientation in the ML plane was quantified using standard deviation of the phone angle and velocity.

Statistical analysis

Bland–Altman LOA was used to provide a descriptive estimate of the magnitude of estimates provided by the AccWalker and 3D motion capture for each of the dependent variables. Standard deviation of LOA (SD LOA) and 95% LOA were estimated using the algorithm for repeated measures designs implemented in *R* using the *BA.est* function from the *MethComp* library (Carstensen *et al*2015). Difference scores between the methods for each subject were evaluated for normality and heteroscedasity. Bias of the AccWalker measurements was evaluated using repeated-measures *t*-test, with an alpha level 0.05. Effect sizes of the bias were calculated using pooled variance Cohen's *d*.

The criterion for validating the AccWalker against motion capture was based on ICC(3,k). This version of ICC was selected because the results are expected to generalize only to the measurement methods used in this study, and they were based on $k = 2$ trials per subject. Relevant terms for calculating ICC were extracted from one-way repeated measures ANOVA with Method (Mocap versus AccWalker) and Subjects as factors. ICC values greater than 0.9 were considered adequate to substitute motion capture for AccWalker measurement to allow individual decision-making based on AccWalker metrics (Portney and Watkins 2000).

Experiment 3: test-retest reliability

Participants

Thirty-two healthy young adults (14 men and 18 women; mean age 24.66 ± 4.73 yrs) from the Department of Kinesiology at the University of North Carolina at Greensboro and the Department of Physical Therapy at Temple University took part in the study. All study procedures were approved by the IRBs at both universities.

Procedures

Eyes open (EO), no-vision (NV), and head shake (HS) stepping-in-place trials were performed in the order listed, with three trials per condition (figure 4(A)). The procedures and the instruction for the EO condition were the same as in Experiment 2. In the NV condition, participants wore a taped-over ski mask and did not remove it between trials. This served to ensure that they were not aware of any change in their heading or position that occurred during the NV trials and did not attempt to deliberately correct for it in the subsequent trial. In the HS condition, participants were instructed to move their head side-to-side (about 20°) while mainintaining visual fixation on the target in front of them. They were instructed to couple their head movement to the leg movement and to keep moving their head continuously throughout the trial. Each trial was performed immediately after the previous one in the EO and EC conditions. In the HS condition, participants took short brakes (up to 30–45 s) in between trials to minimize dizziness. The participants completed all of the EO, EC, and HS trials in each of two sessions separated by approximately a week (7.31 ± 1.2 d on average).

Materials

Phone specifications and placement were identical to Experiment 2. Head yaw angle was measured using the XSens inertial measurement unit (MTw Development Kit, Enschede, Netherlands) during the HS condition.

Statistical analysis

One-way repeated-measures analysis of variance (ANOVA) with Session (1 versus 2) was performed separately for each experimental condition (EO, EC, HS) and dependent variable. Performance on the three trials per condition were averaged for the analysis. ANOVA results were used to calculate the intra-class correlation coefficient ICC(2,k) and the standard error of the mean (Shrout and Fleiss 1979, Weir 2005). ICC(2,k) was utilized to estimate test-retest reliability because this ICC type incorporates both systematic and random error (Weir 2005). SEM was included as a measure of absolute reliability. It was calculated as a square root of mean square of the error term in the ANOVA in order to minimize the effect of systematic variability on this metric and to minimize its dependence on the exact version of ICC (Weir 2005). Based on the suggestion from Koo and Li (2016), ICC values can be interpreted as the following: less than 0.50 (poor reliability), 0.50–0.75 (moderate reliability), 0.75–0.90 (good reliability), and greater than 0.90 (excellent reliability). Such an interpretation was adopted for this manuscript.

To determine if there were performance differences between the conditions and to determine if there were practice effects, we used a Condition×Session repeated-measures ANOVA, followed up with post-hoc *t*-tests. Head ROM and velocity in the horizontal plane were also evaluated for practice effects. The alpha level for the main effect of session was set at 0.1 to provide a more liberal detection of practice effects (Fleiss 1999).

Results

Experiment 1: Android orientation sensor validation

Pendulum movement pattern recorded by the motion capture and the AccWalker were visually similar (figure 1(B)). The mean absolute difference between the maxima of the two recordings was 0.35° (SD = 0.14°). The average timing difference between the maxima was 0.009 s. Mean absolute difference was 1.0° (SD = 0.17°), and the timing difference was 0.01 s for the minima.

The angular trajectory of AccWalker differed slightly from the motion capture in the initial 5–6 oscillations (figure 1(C)), primarily due to an asymmetric velocity profile near peak velocities (figure 1(E)). However, AccWalker-estimated velocity became more symmetrical and similar to motion capture over time (figure 1(D)) when the maximum velocities were around 50 °

s^{-1} (figure 1(F)). In addition, both signals were plotted in phase space to simultaneously visualize the angle and angular velocity of the pendulum illustrating this observation (figures 1(G) and (H)).

Performance of the AccWalker significantly degraded after upgrading to Android OS 5.1 (figure 2). Figure 2 shows AccWalker recordings of the pendulum oscillation (orange) and corresponding phase space (blue) when running Android 4.4.4 and Android 5.1. The AccWalker showed substantial drift in the estimated pendulum angle after OS upgrade.

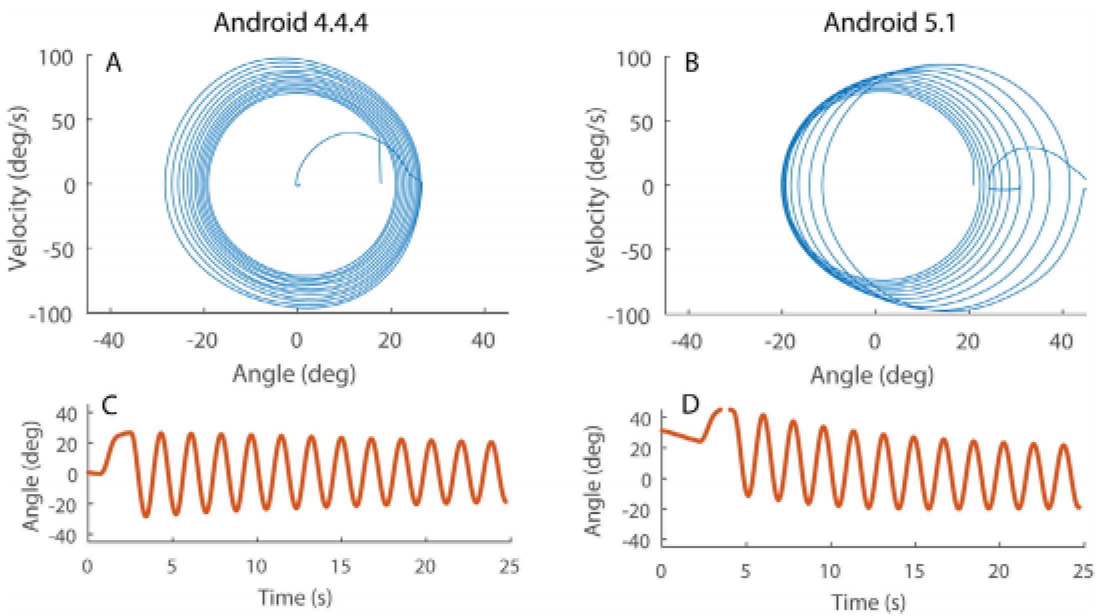

Figure 2. Performance of the AccWalker when running on Android OS 4.4.4 versus Android 5.1. Panels (C) and (D) show the time series of the pendulum angle measured by the same phone while running these two different operating systems. Drift is clearly evident in the angle measurement provided by Android 5.1. Corresponding position-velocity phase spaces are shown in Panels (A) and (B)—an ideal pendulum oscillation measurement should show circular trajectories.

Experiment 2: concurrent validity of AccWalker and 3D motion capture during stepping-in-place and treadmill walking

Stepping-in-place

Time series of thigh angle and angular velocity closely corresponded between the measurement systems when the phone was properly placed on the thigh (i.e. screen perpendicular to thigh motion in AP plane; figure 3(A)). Placing the phone anteriorly biased AccWalker's ROM measurement (figure 3(C)), but did not affect peak thigh flexion timing. ML trunk velocity corresponded to 3D motion capture measurement more closely than the ML trunk angle (figure 3(D)).

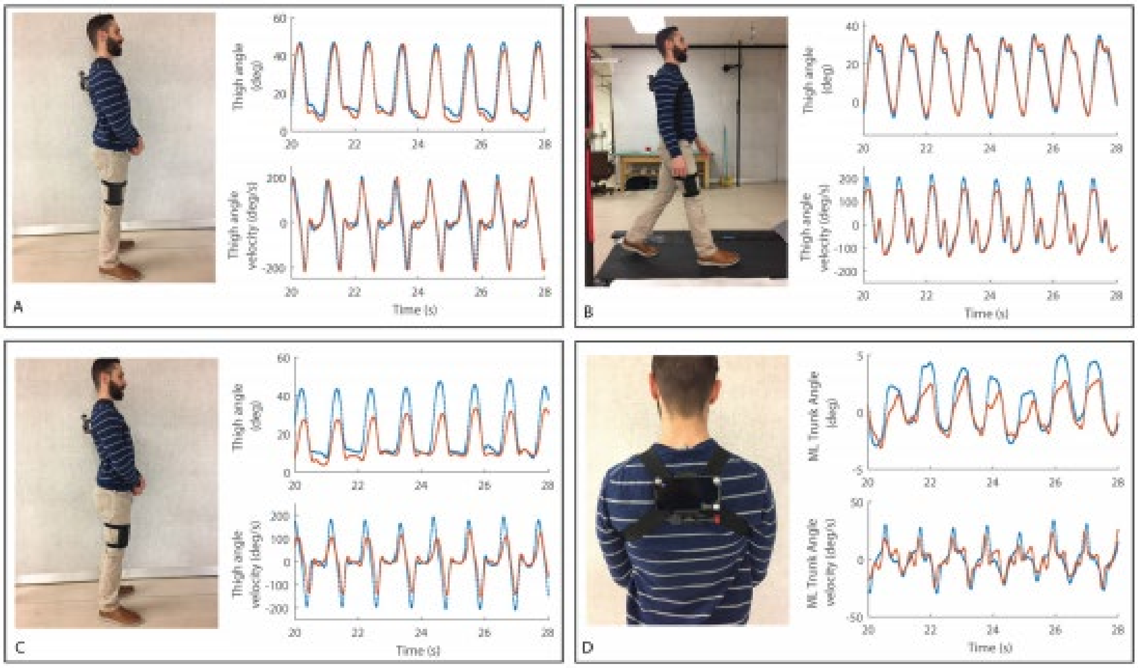

Figure 3. Examples of thigh angle and velocity time series recorded from AccWalker (orange lines) and motion capture (blue lines). Panels (A) and (B) depict the conditions with proper phone placement of the phone on the thigh during the stepping-in-place task and treadmill walking, respectively. Panel (C) indicates an anterior shift in the phone's placement on the thigh and the corresponding thigh angle and velocity time series during stepping in place to the right of the panel. Panel (D) shows the phone placed on the trunk and the corresponding ML angle and velocity time series during stepping-in-place.

Table 1 presents the LOA and ICC results for each dependent variable. Results showed that all temporal metrics were reliably detected by AccWalker (ICC values $>$ 0.9), without measurement bias. Spatial metrics calculated based on AccWalker were also reliable (ICC values $>$ 0.9), but showed bias in the estimates compared to motion capture. For example, ROM was about 2.65° larger according to AccWalker as compared to 3D motion capture. However, the effect sizes of these biases were small to medium (Cohen's ranging from 0.15 to 0.39). Measurements of trunk ML variability were not reliably detected by the AccWalker compared to motion capture as indicated by ICC values $<$ 0.9.

Effect of anterior shift of phone placement

Shifting the phone anteriorly on the thigh worsened AccWalker thigh movement amplitude measurement (figure 3(C)). However, Mean Stride Time, CV Stride Time, Pace Drift, Peak thigh SD, and Peak Return Velocity SD remained reliable as indicated by ICC values greater than 0.9 (table 2). On the other hand, ACF1, Thigh ROM and Peak Lift Velocity SD became unreliable and showed greater measurement bias. Most clearly, shifting the phone affected thigh ROM estimation by the AccWalker, which was 11.47° smaller than the ROM detected by the motion capture, making this measurement not robust to phone misplacement.

Table 1. Thigh and trunk metrics calculated from 3D motion capture and AccWalker during stepping in place when the phone was properly placed on the thigh (see figure 3(A)).

	Unit	Motion capture Mean	Motion capture SD	AccWalker Mean	AccWalker SD	Bland–Altman LOA Bias	Bias *p*-value	Bias Effect Size	SD	Lower 95% CI	Upper 95% CI	ICC (3,k)
Temporal metrics												
Mean stride time	s	1.14	0.04	1.14	0.04	0.000	0.18	0.00	0.001	−0.001	0.001	1.00
CV stride time	%	2.06	0.32	2.07	0.32	0.01	0.31	−0.04	0.07	−0.11	0.15	1.00
ACF1	a.u.	0.31	0.16	0.30	0.15	−0.01	0.57	0.04	0.06	−0.13	0.11	0.98
Pace drift	s	0.04	0.02	0.04	0.02	0.00	0.48	0.01	0.00	0.00	0.00	1.00
Spatial metrics												
Peak thigh SD	deg	1.99	0.57	2.11	0.55	0.12	0.00	−0.15	0.08	−0.05	0.28	1.00
Thigh ROM	deg	44.03	8.54	46.63	8.57	2.65	<.001	−0.22	1.36	−0.07	5.37	0.99
Peak lift vel SD	deg/s	11.90	2.95	10.91	2.30	−0.99	0.02	0.27	1.21	−3.41	1.43	0.96
Peak return vel SD	deg/s	12.23	2.69	13.98	3.63	1.76	0.00	−0.39	1.32	−0.88	4.39	0.96
Trunk												
ML SD	deg	1.41	0.43	1.61	0.50	0.20	0.14	−0.30	0.39	−0.57	0.97	0.83
ML velocity SD	deg/s	8.24	2.53	9.88	2.57	1.64	0.03	−0.46	1.93	−2.20	5.48	0.85

Note. SD—standard deviation, bias—average difference between the 3D motion capture and AccWalker, SD LOA—standard deviation of the difference between the 3D motion capture and AccWalker, bias *p*-value—*t*-test comparison of the AccWalker to motion capture. Effect size was calculated using Cohen's pooled variance formula.

Table 2. Average thigh and trunk metrics calculated from 3D motion capture and AccWalker, and the results of Bland–Altman LOA test when the phone was placed more anteriorly on the leg (see figure 3(C)).

	Unit	Motion capture Mean	Motion capture SD	AccWalker Mean	AccWalker SD	Bland–Altman LOA Bias	Bias *p*-value	Bias effect size	SD	Lower 95% CI	Upper 95% CI	ICC (3,1)
Temporal metrics												
Mean stride time	s	1.13	0.04	1.13	0.04	0.000	0.56	−0.002	0.001	−0.001	0.001	1.00
CV stride time	%	1.88	0.30	1.98	0.33	0.10	0.05	−0.23	0.13	−0.16	0.37	0.91
ACF1	a.u.	0.18	0.14	0.11	0.11	−0.07	0.08	0.37	0.10	−0.26	0.13	0.71
Pace drift	s	0.02	0.01	0.02	0.01	0.00	0.34	0.02	0.00	0.00	0.00	1.00
Spatial metrics												
Peak thigh SD	deg	1.98	0.57	1.91	0.60	−0.08	0.31	0.09	0.21	−0.49	0.33	0.94
Thigh ROM	deg	40.56	6.85	29.09	7.15	−11.47	0.00	1.16	4.68	−20.64	−2.30	0.78
Peak lift vel SD	deg/s	12.52	3.19	10.78	2.53	−1.74	0.01	0.43	1.66	−5.00	1.52	0.83
Peak return vel SD	deg/s	11.31	3.49	11.27	4.48	−0.04	0.93	0.01	1.44	−2.86	2.78	0.94

Note. SD—standard deviation, bias—average difference between the 3D motion capture and AccWalker, SD LOA—standard deviation of the difference between the 3D motion capture and AccWalker, SD LOA/SD—ratio of the SD LOA to SD of the 3D motion capture (expressed as percentage).

Treadmill walking

AccWalker provided valid measurements of stride time and its CV, ACF1, and drift during treadmill walking, with some additional statistically significant biases in the AccWalker measurements. The effect sizes of the biases were small, apart from ACF1 (see table I in supplementary materials).

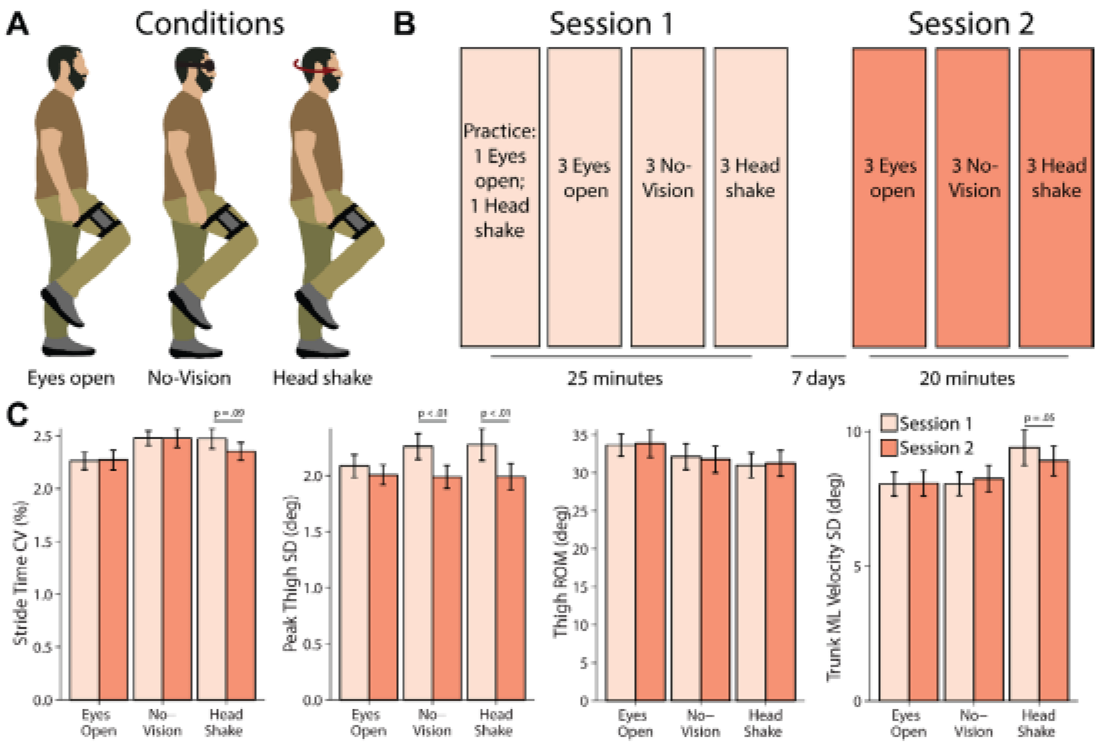

Figure 4. (A) Experimental conditions, (B) study design, (C) changes in the dependent measures for each sensory condition and session.

Table 3. ICC(2,k) and SEM values for each variable and condition.

		ICC(2,k)			SEM		
	Unit	Eyes open	No-vision	Head shake	Eyes open	No-vision	Head shake
Temporal metrics							
Mean stride time	s	0.80[a]	0.80	0.81	0.02	0.02	0.03
CV stride time	%	0.77	0.82	0.81[a]	0.34	0.27	0.31
ACF1	a.u.	0.49[a]	0.78	0.44	0.16	0.10	0.15
Pace drift	s	0.23	0.47	0.54	0.02	0.02	0.02
Spatial metrics							
Peak thigh SD	deg	0.82	0.73[a]	0.89[a]	0.32	0.46	0.36
Thigh ROM	deg	0.90	0.92	0.94	4.10	3.82	3.23
Peak lift vel SD	deg/s	0.82	0.77[a]	0.88[a]	1.39	1.97	1.78
Peak return vel SD	deg/s	0.90	0.85[a]	0.89[a]	1.64	2.27	2.23
Trunk: spatial metrics							
ML SD	deg	0.59	0.89	0.94[a]	0.33	0.17	0.16
ML velocity SD	deg/s	0.64	0.90	0.96[a]	1.74	0.94	0.86

[a]Signifies presence of a practice effect.

Experiment 3: test-retest reliability

The ICC(2,k) and SEM values for all dependent measures are presented in table 3. Mean stride time, stride time CV, peak thigh flexion SD, thigh ROM, and thigh velocity maxima had good test-retest reliability (ICC $>$ 0.75) in each of the sensory conditions and Trunk ML SD and velocity were reliable in the no-vision and head shake conditions. However, there were practice effects for stride time CV in the head shake condition (Cohen's d = 0.24), for peak thigh SD in the no-vision ($d = 0.45$) and head shake ($d = 0.39$), and for trunk ML Velocity SD in the head shake condition ($d = 0.17$) as illustrated in figure 4(C). The ICC values for ACF1 and pacing drift were generally lower than 0.75, indicating low test-retest reliability of these variables.

Stride time CV was greater in the no-vision and head shake conditions compared to the eye open condition (both p's $< .01$; figure 4(C)). Thigh ROM was lower in the no-vision and head shake conditions compared to eyes open condition ($p = .01$ and $p < .001$, respectively). Trunk ML Velocity SD was greater in the head shake condition compared to no-vision and eyes open conditions (both p's $< .01$).

Horizontal range of head motion adopted by participants in the head shake condition was 62.69° (SD = 14.03°) and 61.18° (SD = 11.81°) in session 1 and 2, respectively, $p = .23$. Peak head velocity was 179.50 $°s^{-1}$ (43.31) and 177.78 $° s^{-1}$ (35.58) in session 1 and 2, $p = .30$. However, variability of peak head velocity decreased from session 1 ($M = 22.43$, SD = 5.97) to session 2 (19.82, SD = 4.66), $p < 0.01$.

Discussion

Experiment 1: Android orientation sensor validation

Experiment 1 showed that the smartphone orientation sensor provides an accurate measurement of pendulum angle kinematics in comparison to research-grade motion capture system when running on Android OS 4.4.4, but not on 5.1. When running 4.4.4, the angle measurements provided by the phone differed from motion capture by only 0.35° to 1.0°, which is consistent with previously reported values for inertial measurement units (Umek and Kos 2016). However, using the same app on Android 5.1 led to a substantial drift in the estimated angle in early segment of the trial.

This result suggests that the sensor fusion algorithm implemented by InvenSense (San Jose, CA) on a Motorola Moto X2 running on Android 4.4.4 is of sufficient quality for orientation measurement of human motion. The phone may have performed sub-optimally on Android 5.1 because the MPU-6515 sensor (the accelerometer and gyroscope unit) has internal fusion algorithms that were specifically optimized for inertial orientation tracking in smartphones running Android 4.4.4 as described in manufacturer's specifications for the sensor (six-axis MEMS MotionTracking Device, 2017). The degree to which other smartphones would be susceptible to the same issues with OS upgrades needs to be tested for each phone independently prior to using them for human motion analysis applications. Similar issues were identified in different versions of iPad for reaction time measurements (Schatz *et al* 2015).

Two other issues became apparent. First, the phone must be oriented parallel to the plane of motion to detect pendulum angle accurately (i.e. phone's screen should be perpendicular to the plane of pendulum oscillation). As we describe in Experiment 2, tilting the phone with respect to the dominant plane of motion reduces accuracy of angle amplitude estimation. Second, magnetic field sources affect orientation measurements and prevent the Rotation Vector sensor from initiating. In our experience, the sensor stopped working when the strength of the field was greater than 130 μT. The solution is to remove the source of magnetic field and re-calibrate the phone by performing figure-8 calibration.

Information provided in this experiment is relevant for understanding the limitations of internal Android functions for orientation estimation prior to using smartphones to quantify human movement kinematics in the field setting. Change in the Android OS versions influences the quality of measurement provided by the sensors. It is impossible to know *a priori* how changes in the sensor components and/or software may affect the quality of orientation detection in commercial smartphones. The simple pendulum setup can be used in future validation studies of smartphone orientation detection algorithms. Such validation is crucial when adopting commercial smartphones for clinical measurement of balance function.

Experiment 2: concurrent validity of AccWalker and 3D motion capture during stepping-in-place and treadmill walking

Experiment 2 showed that thigh orientation detection implemented in the AccWalker app running on Android 4.4.4 provides valid measures of temporal and spatial characteristics of thigh movement during stepping-in-place compared to laboratory-grade motion capture when the phone is positioned perpendicularly to the plane of hip flexion. Mean stride time and its variability (CV), Pace Drift, Peak Thigh SD, and Peak Return Velocity SD during leg return to stance were robust to shifts of phone position on the thigh. Stride time and its variability were also reliably detected by the AccWalker during treadmill walking. Measurements of ML trunk orientation however were not reliably detected by the phone compared to motion capture, suggesting that they measure slightly different aspects of trunk movement. The AccWalker measured the ML tilt of the upper trunk because the phone was placed on that segment of the trunk. For the motion capture, the definition of ML trunk tilt was based on the angle between the vector connecting lumbar and cervical markers and the vertical—such measure is less sensitive to upper trunk movement than the AccWalker, which may explain lower convergence between the for characterizing trunk movement.

Experiment 3: test-retest reliability

Results of Experiment 3 indicate that all AccWalker measures other than lag 1 autocorrelation (ACF1) and drift of stride time pacing had good test-retest reliability on the dynamic balance test. Good reliability of stride time CV and kinematic variability metrics (Peak Thigh SD, thigh ROM, SD of thigh velocity) was observed for group comparisons as indicated by ICC values greater than 0.75 in all sensory conditions of the stepping-in-place protocol. However, these metrics (except thigh ROM) are also subject to practice effects primarily in the NV and HS conditions, second session being generally less variable.

Practice effects are problematic for screening neuromotor sequelae of head trauma because they add an additional factor affecting motor variability above and beyond any changes in the neuromotor status due to the trauma, making it difficult to interpret minimum detectable change scores from the baseline. However, other balance tests such as the SOT and BESS have also been reported to have practice effects as well (McLeod *et al* 2004, Wrisley *et al* 2007). The effect sizes of the practice effects in the in eyes closed and head shake conditions of stepping in place test were smaller compared to the previously reported practice-related changes of the SOT composite scores (SOT Cohen's $d = 1$ compared to a maximum $d = 0.44$ in our study) (Wrisley *et al* 2007). The ICC values were higher for our protocol than for the SOT scores in the individual conditions and composite scores Wrisley *et al*(2007). Total BESS scores also decrease over repeated administrations with an effect size $d = 0.74$, but show high inter-test reliability (McLeod *et al* 2004).

Despite the presence of practice effects, our dynamic balance task may be more taxing for individuals after head trauma than static balance testing and their performance may still show deterioration regardless of any practice effects. The next step would be to perform discriminant validity study to test the hypothesis that this dynamic balance protocol successfully detects variability alterations after head trauma. In our previous work using repeated neuromotor assessments with a similar smartphone app and dynamic balance protocol, we could successfully identify stride time CV changes after sub-concussive head trauma due to low-level blast exposure (Rhea *et al* 2017).

Differences between the sensory conditions followed the expected pattern, as the visual and vestibular perturbations generally increased movement variability (Wuehr *et al* 2013). The reduction of the ROM in the no-vision and head shake conditions is consistent with previous studies documenting decreased gait velocity while walking without visual input (Hallemans *et al* 2009) and may be related to an attempt to reduce the risk of falling by lifting the foot less. Trunk velocity was highest in the head shake condition, which may be related to the increased trunk movement and to the destabilizing vestibular effect on the function of the horizontal semi-circular canals. The VOR was activated by the head movements because peak head velocity was greater than the minimally required peak velocity to activate the vestibular-ocular reflex ($40\ ^\circ\ s^{-1}$) (Peterka *et al* 1990).

Conclusion

This study introduces a new portable field-based protocol to test dynamic balance function based on a stepping-in-place task. We performed validity and reliability testing of the orientation sensor on an Android smartphone placed on the thigh and upper trunk compared to motion capture. Experiment 1 identified differences in orientation detection between different versions of Android OS and established accuracy of pendulum angle measurement using the phone. Experiment 2 showed that all temporal and spatial variability are reliable when the phone is placed perpendicular to the plane of thigh flexion and identified variables that are robust to phone misplacement. Experiment 3 additionally identified several variables that are reliable over time. Overall, the results indicate that mean stride time and its variability (CV), Peak Thigh SD, Thigh ROM, and Peak Return Velocity calculated based on the smartphone orientation sensor could be reliably used to assess characteristics of thigh movement during stepping-in-place. A limitation of the data presented here is that the experiment was laboratory-based in a controlled environment. However, the motivation for developing an objective dynamic balance test with a smartphone app is the potential for portable assessment in many different environments, such as sideline assessment. Prior to adoption to help with clinical care, a series of 'real-world' experiments with our smartphone app and dynamic balance protocol should be carried out, including examining how different surfaces, footwear, and worn sports equipment influences the variables presented in this paper. Nevertheless, this manuscript is the necessary first step toward more ecologically valid assessment by describing the validity and reliability of our app and dynamic balance protocol in a controlled setting. Once normative data for a variety of populations/environments are developed, the app could be modified to provide the administrator with a simple red/yellow/green light indicating whether the athlete is outside/borderline/within normative performance on the dynamic balance test, providing a fast screening method to assist in clinical care. Such a device would be useful to quantify dynamic balance in a variety of populations with neurological dysfunction, such as chronic ankle instability, older adults, and populations with sub-concussive or concussive head trauma.

Acknowledgments

The US Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick MD 21702-5014 is the awarding and administering acquisition office. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. (W81XWH-15-1-0094) to Christopher K. Rhea. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense. Researchers and/or clinicians interested in obtaining our app and user manual for testing purposes are encouraged to contact the corresponding author.

References

Abbas K, Shenk T E, Poole V N, Breedlove E L, Leverenz L J, Nauman E A, Talavage T M and Robinson M E 2015 Alteration of default mode network in high school football athletes due to repetitive subconcussive mild traumatic brain injury: a resting-state functional magnetic resonance imaging study Brain Connect. **5** 91–101

Alberts J L, Thota A, Hirsch J, Ozinga S, Dey T, Schindler D D, Koop M M, Burke D and Linder S M 2015 Quantification of the balance error scoring system with mobile technology Med. Sci. Sports Exerc. **47**2233–40

Alsalaheen B A, Haines J, Yorke A, Stockdale K and Broglio S P 2015 Reliability and concurrent validity of instrumented balance error scoring system using a portable force plate system Phys. Sportsmed. **43** 221–6

Alsalaheen B A, Mucha A, Morris L O, Whitney S L, Furman J M, Camiolo-Reddy C E, Collins M W, Lovell M R and Sparto P J 2010 Vestibular rehabilitation for dizziness and balance disorders after concussion J. Neurol. Phys. Ther. **34** 87–93

Amick R Z, Chaparro A, Patterson J A and Jorgensen M J 2015 Test-retest reliability of the sway balance mobile application J. Mob. Technol. Med. **4** 40–7

Basford J R, Chou L S, Kaufman K R, Brey R H, Walker A, Malec J F, Moessner A M and Brown A W 2003 An assessment of gait and balance deficits after traumatic brain injury Arch. Phys. Med. Rehabil. **84** 343–9

Bland J M and Altman D G 1999 Measuring agreement in method comparison studies Stat. Methods Med. Res. **8** 135–60

Buster T, Burnfield J, Taylor A P and Stergiou N 2013 Lower extremity kinematics during walking and elliptical training in individuals with and without traumatic brain injury J. Neurol. Phys. Ther. **37** 176–86

Carstensen B, Gurrin L C, Ekstrom C and Figurski M 2015 MethComp: functions for analysis of agreement in method comparison studies R package version 1.22.2(http://bendixcarstensen.com/methcomp/)

Chen K Y, Janz K F, Zhu W and Brychta R J 2012 Re-defining the roles of sensors in objective physical activity monitoring Med. Sci. Sports Exerc. **44** S13

Chou L S, Kaufman K R, Walker-Rabatin A E, Brey R H and Basford J R 2004 Dynamic instability during obstacle crossing following traumatic brain injury Gait Posture **20** 245–54

Ellis M J, Cordingley D, Vis S, Reimer K, Leiter J and Russell K 2015 Vestibulo-ocular dysfunction in pediatric sports-related concussion J. Neurosurg. **16** 248–55

Fleiss J L 1999 The design and analysis of clinical experiments vol 73 (New York: Wiley)

Gavett B E, Stern R A and Mckee A C 2011 Chronic traumatic encephalopathy: a potential late effect of sport-related concussive and subconcussive head trauma Clin. Sports Med. **30** 179–88

Goble D J, Manyak K A, Abdenour T E, Rauh M J and Baweja H S 2016 An initial evaluation of the BTrackS balance plate and sports balance software for concussion diagnosis Int. J. Sports Phys. Ther.**11** 149–55

Godfrey A, Hourigan T and Ólaighin G M 2007 Pendulum analysis of an integrated accelerometer to assess its suitability to measure dynamic acceleration for gait applications Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE 2007 pp 4891–4

Gottshall K R and Hoffer M E 2010 Tracking recovery of vestibular function in individuals with blast-induced head trauma using vestibular-visual-cognitive interaction tests J. Neurol. Phys. Ther. **34** 94–7

Grabiner P C, Biswas S T and Grabiner M D 2001 Age-related changes in spatial and temporal gait variables Arch. Phys. Med. Rehabil. **82** 31–5

Gysland S M, Mihalik J P, Register-Mihalik J K, Trulock S C, Shields E W and Guskiewicz K M 2012 The relationship between subconcussive impacts and concussion history on clinical measures of neurologic function in collegiate football players Ann. Biomed. Eng. **40** 14–22

Hallemans A, Beccu S, Van Loock K, Ortibus E, Truijen S and Aerts P 2009 Visual deprivation leads to gait adaptations that are age and context-specific: I. Step-time parameters Gait Posture **30** 55–9

Hausdorff J M 2007 Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking Hum. Mov. Sci. **26** 555–89

Honaker J A, Boismier T E, Shepard N P and Shepard N T 2009 Fukuda stepping test: sensitivity and specificity J. Am. Acad. Audiol. **20** 311–4

Johnston W, Coughlan G F and Caulfield B 2016 Challenging concussed athletes: the future of balance assessment in concussion QJM **110** 779–83

Kang H G and Dingwell J B 2008 Effects of walking speed, strength and range of motion on gait stability in healthy older adults J. Biomech. **41** 2899–905

Koo T K and Li M Y 2016 A guideline of selecting and reporting intraclass correlation coefficients for reliability research J. Chiropr. Med. **15** 155–63

Mccrory P, Meeuwisse W H, Aubry M, Cantu B, Dvořák J, Echemendia R J, Engebretsen L, Johnston K, Kutcher J S and Raftery M 2013 Consensus statement on concussion in sport: the 4th international conference on concussion in sport held in Zurich, November 2012 Br. J. Sports Med. **47** 250–8

Mckee A C, Cantu R C, Nowinski C J, Hedley-Whyte E T, Gavett B E, Budson A E, Santini V E, Lee H-S, Kubilus C A and Stern R A 2009 Chronic traumatic encephalopathy in athletes: progressive tauopathy after repetitive head injury J. Neuropathol. Exp. Neurol. **68** 709–35

Mcleod T C V, Perrin D H, Guskiewicz K M, Shultz S J, Diamond R and Gansneder B M 2004 Serial administration of clinical concussion assessments and learning effects in healthy young athletes Clin. J. Sport Med. **14** 287–95

Newell K M and James E G 2008 The amount and structure of human movement variability Routledge Handbook of Biomechanics and Human Movement Science (New York: Routledge) pp 93–104

O'Connor S M, Baweja H S and Goble D J 2016 Validating the BTrackS balance plate as a low cost alternative for the measurement of sway-induced center of pressure J. Biomech. **49** 4142–5

Parker T M, Osternig L R, Van Donkelaar P and Chou L S 2008 Balance control during gait in athletes and non-athletes following concussion Med. Eng. Phys. **30** 959–67

Peterka R J, Black F O and Schoenhoff M 1990 Age-related changes in human vestibulo-ocular reflexes: sinusoidal rotation and caloric tests Journal of Vestibular Research **1** 49–59

Poole V N, Breedlove E L, Shenk T E, Abbas K, Robinson M E, Leverenz L J, Nauman E A, Dydak U and Talavage T M 2015 Sub-concussive hit characteristics predict deviant brain metabolism in football athletes Dev. Neuropsychol. **40** 12–7

Portney L G and Watkins M P 2000 Foundations of Clinical Research: Applications to Practice(Englewood Cliffs, NJ: Prentice Hall)

Rhea C K and Kiefer A W 2014 Patterned variability in gait behavior: how can it be measured and what does it mean? Gait Biometrics: Basic Patterns, Role of Neurological Disorders and Effects of Physical Activity ed L Li and M Holmes (Hauppauge, NY: Nova Science Publishers)

Rhea C K, Kuznetsov N A, Ross S E, Long B, Jakiela J T, Bailie J M, Yanagi M A, Haran F J, Wright W G and Robins R K 2017 Development of a portable tool for screening neuromotor sequelae from repetitive low-level blast exposure Mil. Med. **182** 147–54

Schatz P, Ybarra V and Leitner D 2015 Validating the accuracy of reaction time assessment on computer-based tablet devices Assessment **22** 405–10

Shrout P E and Fleiss J L 1979 Intraclass correlations: uses in assessing rater reliability Psychol. Bull.**86** 420–8

Stergiou N and Decker L M 2011 Human movement variability, nonlinear dynamics, and pathology: Is there a connection? Hum. Mov. Sci. **30** 869–88

Talavage T M, Nauman E A, Breedlove E L, Yoruk U, Dye A E, Morigaki K E, Feuer H and Leverenz L J 2014 Functionally-detected cognitive impairment in high school football players without clinically-diagnosed concussion J. Neurotrauma **31** 327–38

Umek A and Kos A 2016 Validation of smartphone gyroscopes for mobile biofeedback applications Pers. Ubiquitous Comput. **20** 657–66

Verghese J, Holtzer R, Lipton R B and Wang C 2009 Quantitative gait markers and incident fall risk in older adults J. Gerontol. A **64** 896–901

Weir J P 2005 Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM J. Strength Cond. Res. **19** 231–40

Wright W, Tierney R and Mcdevitt J 2017 Visual-vestibular processing deficits in mild traumatic brain injury J. Vestib. Res. **27** 27–37

Wrisley D M, Stephens M J, Mosley S, Wojnowski A, Duffy J and Burkard R 2007 Learning effects of repetitive administrations of the sensory organization test in healthy young adults Arch. Phys. Med. Rehabil. **88** 1049–54

Wuehr M, Schniepp R, Pradhan C, Ilmberger J, Strupp M, Brandt T and Jahn K 2013 Differential effects of absent visual feedback control on gait variability during different locomotion speeds Exp. Brain Res. **224** 287–94