Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology 15 (2019) 29–34

Contents lists available at ScienceDirect

Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology

journal homepage: www.ap-smart.com

Original Article

Validity and reliability of the Nintendo Wii Fit Stillness score for assessment of standing balance

Jonathan J. Negus [a, b, c, *], Donald Cawthorne [c], Ross Clark [d], Oliver Negus [c], Joshua Xu [b], Prof Lyn March [a, b], David Parker [a]

[a] *Sydney Orthopaedic Research Institute, Chatswood, NSW, Australia*
[b] *Faculty of Medicine, University of Sydney, NSW, Australia*
[c] *Jointworks, Orthopaedic Research, Sydney, NSW, Australia*
[d] *Australian Catholic University, Melbourne, VIC, Australia*

A R T I C L E I N F O

Article history:
Received 9 May 2018
Received in revised form
13 August 2018
Accepted 3 September 2018
Available online 14 September 2018

Keywords:
Wii-fit
Standing
Balance
Stillness score
Wii balance board
CoP

A B S T R A C T

Background/objective: Standing balance has become an important clinical measure in patient populations who are at risk of falls or have osteoarthritis. With custom-written software, the Wii Balance Board (WBB) has been shown to be a valid and reliable force platform that can be used to assess standing balance. However, no studies to date have assessed the use of the more readily available Wii Stillness Score (WSS) as a measure of balance.
Methods: Twenty-four individuals without lower limb pathology performed a combination of unilateral and double leg standing balance tests with eyes open or closed on two separate occasions. At each session, data from the WBB were acquired on a laptop computer running custom software and then by Wii-Fit software on a Wii console. The reliability of the WSS was determined by assessing reproducibility, while the validity of the WSS was determined by comparing the results of the WSS to that of the custom-written software.
Results: We found that the WSS exhibited excellent intra and inter device reliability in three out of four stances tested. The Bland-Altman plots also showed good concurrent validity for the three analysed stances. However, there remain significant limitations with the use of the WSS such as its rigid thirty-second time parameter and single score result.
Conclusion: The readily available WBB may be a used as a portable and inexpensive device to assess standing balance with custom written software. However, with the current limitations of the WSS, we would discourage its use as a clinical measure of balance.

Introduction

Balance is the equal distribution of weight that allows an individual to remain upright and steady. It plays a crucial role in human biomechanical function during standing, walking and in response to unpredictable perturbations during sports and activities of daily living. Poor postural control is a known risk factor for falls, which can cause injuries such as hip fractures.[1] Furthermore, the psychological effects of a fall can have detrimental and lasting impacts on a person's self-confidence.[2] The ability to assess standing balance can provide important information that ranges from predicting falls in the elderly to assessing post-operative recovery after orthopaedic surgery.[3–5] Qualitative tests such as the Berg Balance Scale have been commonly used to assess standing balance, but can lack precision.[6,7] Accelerometers have also been used as a quantitative measure of balance, although the accuracy of this method is still not fully elucidated.[8] Laboratory-based force platforms are one of the most commonly used tools to accurately assess balance, but can be expensive and difficult to transport.[9,10] Thus, there has been significant interest in using the Wii Balance Board (WBB) as a device to assess standing balance due to it being portable, affordable and accessible.[11]

Introduced in 2007, the Nintendo WBB (Nintendo, Kyoto, Japan)

---

* Corresponding author. Sydney Orthopaedic Research Institute, Level 1 The Gallery, 445 Victoria Avenue, Chatswood, NSW, 2067, Australia.
E-mail address: jjn@jointworks.com.au (J.J. Negus).

https://doi.org/10.1016/j.asmart.2018.09.001

was a commercially available force platform that was designed to work with the Wii console and Wii-Fit software. By measuring centre of pressure (CoP) changes during a balance test or training game within the Wii-Fit software, a score is produced as a measure of balance ability. Wii-Fit uses the Wii Stillness score (WSS) which gives the subject a score out of 100, based upon how much their CoP changes during a 30 s period. CoP measurements have been shown to be sensitive to postural instability in older adult patients and those with Parkinson's disease.[12–15] Clark et al.[10] used custom-written software for the WBB and compared its accuracy to an industry-grade force platform. From this they found the WBB to be a reliable and valid tool for assessing standing balance.[10] Other studies using the WBB also corroborate this finding.[16,17] However, no studies have been performed to determine the validity and reliability of the more readily available Wii-Fit software in the assessment of standing balance.

By assessing reproducibility of results, we aimed to investigate the reliability of the WSS. In this study, the aim was also to assess the validity of the WSS by comparing it to the results produced when connected to custom-written software.

Methods

Participants

Twenty-four injury free individuals were tested on two occasions within 4 weeks of each other and at least 24 h apart. There were 8 males and 16 females with a mean age of 28.4 years (SD 9.1) and mean BMI of 23.3 (SD 4.4). No participant reported any major symptoms or history of pathology or recent injury in the back or lower limb. There was also no history of medication use or neurological disease history that could have influenced standing balance. The local Human Research Ethics Committee approved the study and all participants provided informed consent.

Reliability procedure

The reliability of the Wii-Fit software was determined via a test-retest method. On each of the two assessment dates, participants performed standing balance tasks on a WBB placed on the floor 1.5 m from a wall. A cross was marked on the wall at eye level to provide a consistent focal point. The subjects received no visual or auditory feedback during testing.

At each assessment, the subjects were instructed to take up each of the four stances and attempt to complete three trials for each. These stances were chosen based on the paper by Clark et al.[10] The WBB was connected either to the Wii-Fit measuring the WSS or to a laptop running custom software to measure CoP coordinates. The WSS is a measure of CoP movement over 30 s that results in a percentage score.

Data was collected from the four stances with the WBB connected to each data collection device in turn. The order of tasks and testing devices was randomly assigned for each participant, but remained consistent between testing sessions.

The four stances:

1. Double-leg Stance with Eyes Open and feet a comfortable distance apart (DSEO).
2. Double-leg Stance with Eyes Closed and feet together (DSEC).
3. Unilateral-leg Stance with Eyes Open (USEO).
4. Unilateral-leg Stance with Eyes Closed (USEC).

For DSEO, the distance between heels was measured and kept consistent between tests and sessions for that subject. Unilateral stances were performed on the dominant leg. This was determined by questioning the patient about ball kicking.

The participants were instructed to keep hands by their sides, remaining as still as possible for the trial duration, which lasted 30 s. If their support foot shifted position or their non-support foot touched the floor then the trial was ceased and recorded as unsuccessful. Three successful trials were attempted for each stance and testing was ceased if there were 3 unsuccessful attempts at any particular stance. There was a 15 s rest between trials and 60 s between stances and devices.

Validity procedure

Each individual WBB only allows its output signal to be received by one Bluetooth-linked device. For the data of each separate trial to be collected simultaneously by two different devices, two WBBs were stacked on top of each other, transmitting a separate but simultaneous Bluetooth signal for each trial. One was linked to the Wii console and data analysed by the Wii-Fit software. The other was linked to a laptop running the custom software, producing CoP coordinate data. The data were then analysed for correlation.

Testing configuration of WBB

Stacking of the WBB on a force platform has been shown to produce accurate results.[16] When testing our stacked WBB configuration, two WBBs were placed on top of each other and positioned on the centre of the Biodex SD force platform (BIODEX, USA).

First, the uppermost WBB was paired via Bluetooth with the custom software. A single subject then performed multiple standing trials. The data output was collected simultaneously on the laptop and by the Biodex SD force platform. The lower of the two WBBs was then paired via Bluetooth with the laptop and the subject repeated multiple standing trials with repeat simultaneous Biodex SD data collection. The output of the WBBs in each position was correlated with the Biodex SD data.

Deriving stillness formula

Using two WBBs stacked on each other, a single subject performed 50 trials creating simultaneous data sets. Each trial produced a % ‘Stillness’ score from the WBB paired to the Wii and a set of CoP coordinates from the WBB paired to the laptop.

The subject changed the magnitude, direction and velocity of their CoP position in a methodical manner to replicate as many potential standing balance results as possible. These variations are summarised in Table 1.

The laptop data sets were analysed using the formulae of known valid measures of standing balance, which were then correlated against the WSS for the corresponding trial.

Data analysis

The data from the WBBs were transmitted via Bluetooth connections. The Wii console, analysed the WBB data using the 30-s ‘Stillness’ test in its Wii-Fit software (Nintendo). This produced a percentage score, with a higher percentage suggesting less

Table 1
The variations in the magnitude, direction and velocity of the subject's centre of pressure standing position for different trial numbers.

Trial no.	Variation
0–10	Low magnitude – all directions – low velocity
11–20	High magnitude - medial lateral (ML) – low velocity
21–30	High magnitude - anterior posterior (AP) – low velocity
31–40	High magnitude - all directions – low velocity
41–50	High magnitude - all directions – high velocity

movement of the CoP.

The laptop computer used custom-written software (Labview 8.5 National Instruments, Austin, TX, U.S.A.). The Labview software was calibrated as detailed in Clark's paper.[10] The Labview software produces a series of X,Y coordinates, at a specified frequency, representing the CoP measured. Our data was sampled at 40 Hz and filtered using an eighth order Butterworth filter with a low-pass cut-off frequency of 12 Hz as per Clark et al. The authors do not know any data filtering and cut-off levels used by the Wii-Fit software. The Biodex force platform also produces CoP coordinates. This data was stored in the integrated console and exported into a '.csv' file. The median value of each set of three trials was used.

Statistical analysis

To examine agreement between the two software analyses of the data, Bland-Altman plots were used. To create these plots, the data from each device had to be in the same units (%). The maximum CoP (X-Y) amplitude for each trial of the laptop data was converted into a percentage score using the equation of the line of best fit from Fig. 1.

The result was a Stillness % score that was derived from the laptop data, enabling direct comparison to the Wii Stillness % score. The difference in percentage score between the two software packages was plotted against the mean results – Bland Altman plots. A two-way, random-effects, single measure (median of the three trials) intraclass correlation coefficients ($ICC_{(2,1)}$) model was used to assess reliability. In conjunction with the ICC values, standard error of measurement (SEM) and minimum detectable change (MDC) values were calculated to assess the concurrent validity between the laptop derived Stillness score and the WSS, as well as the within-device test-retest reliability and measurement error over the two testing sessions. Point estimates of the ICCs were interpreted as follows: excellent (0.75–1), Modest (0.4–0.74), or poor (0–0.39).[18]

All statistical analyses were conducted using IBM SPSS, version 22 (Chicago, IL, USA). The MDC, which is based on the reliable change index score, was calculated using the equations reported previously by Jacobson and Truax.[19] It is expressed as the percentage test-retest change in score required, to find a significant difference at an alpha level of 0.05 based on the Day 1 mean value.

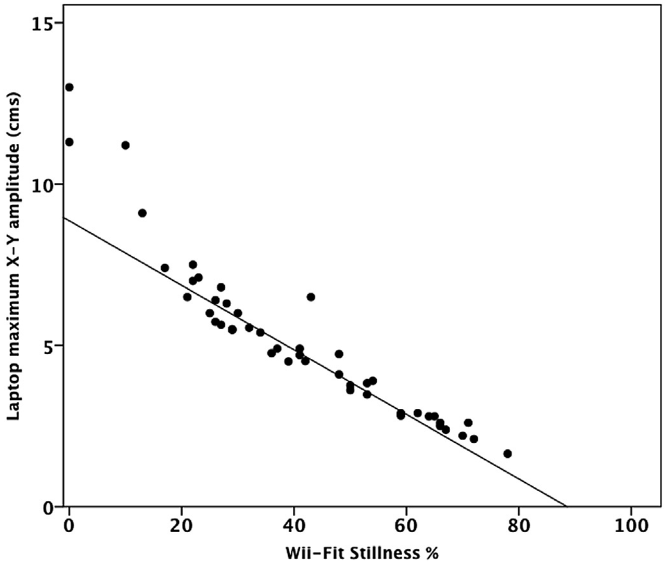

Fig. 1. Correlation of Wii-Fit Stillness score against Labview data converted into an equivalent measure of amplitude.

Results

The full results table is shown in Table 2.

Reliability

Test-retest analysis was performed on the data for twenty-four participants in the three analysed stances (DSEO, DSEC, USEO). The Stillness score as measured on the Wii showed excellent test-retest reliability for these three stances (ICC = 0.85–0.89). The USEC was not analysed as most patients in this stance could not maintain their support foot position or touched their non-support foot to the floor. Seven patients in USEC failed the stance 3 times and thus the trial was ceased for these patients and recorded as unsuccessful.

Validity

Validity of stacked WBB configuration

The analysis of the CoP coordinate data from the Biodex showed an excellent correlation to each of the stacked WBBs paired with the laptop (Fig. 2). The position of the WBB being tested had no effect.

Calculating the stillness measure

The Stillness score did not correlate well with path length or sum of distance from origin but it correlated well with the sum of the ML and AP ranges (i.e. the maximum amplitudes in the x and y axes).

The formula for the sum of the individual maximum X and Y ranges of the raw CoP data resulted in an excellent correlation ($R^2 = 0.919$).

Maximum X-Y amplitude (cms) = ABS (Max X − Min X) + ABS (Max Y − Min Y)

X = (Y - 8.86)/- 0.1) Line of best fit, excluding 3 outliers at very low Wii Stillness % scores. In this equation, X is the WSS and Y is the maximum X-Y amplitude in cm.

Concurrent validity

Looking at inter-device ICCs between the laptop custom software and the WSS, the DSEC validity was excellent on both testing sessions. (0.91 & 0.89). The DSEO and USEO validity were modest to excellent. (0.70–0.77). For the USEC stance, fourteen participants were unable to complete at least one trial on both testing days for both machines. Due to this paucity of test-retest data, meaningful analyses were not possible for this stance.

The MDC values were reasonably high (16–96%) for both the Wii console Stillness score and the laptop measured maximum X-Y amplitude, as expected from previous studies.[10]

The Bland-Altman plots for the Wii Stillness % score and the derived % score from the laptop amplitude data are shown in Fig. 3. There was no relationship evident for any of the three analysed stances.

Discussion

This study had two major objectives. The first of these was to investigate, by means of assessing the reproducibility of results, the reliability of the WSS. The second objective was to assess the validity of the WSS by comparing it to the results produced when a WBB had been connected to custom-written software. Although

Table 2
Reliability and concurrent validity analysis of Stillness % and Labview sum of range converted to % scores during each of the three analysed standing balance trials.

USEO	Laptop sum of range converted to %		Wii Stillness %		Mean diff	(95% CI)	ICC	(95% CI)
	%	SD	%	SD				
Day 1	30.3	13.6	30.5	8.8	0.2	−16.4 to 17.2	0.72	0.45 to 0.87
Day 2	29.5	16.3	33	11	3.5	−16.0 to 23.1	0.7	0.42 to 0.86
Mean diff (95% CI)	0.8	−24.53 to 21.85	2.52	−1.1 to 0.67				
ICC (95% CI)	0.65	0.34 to 0.83	0.85	0.68 to 0.93				
SEM	7.74		3.44					
MDC %	71		31					
DSEO	%							
Day 1	57	14.3	65.3	10.8	8.3	−9.6 to 26.3	0.74	0.49 to 0.89
Day 2	59.1	8.35	64.3	10.9	5.1	−8.1 to 18.4	0.77	0.54 to 0.89
Mean diff (95% CI)	2.18	−17.53 to 21.89	1.04	−14.61 to 18.24				
ICC (95% CI)	0.64	0.31 to 0.82	0.88	0.74 to 0.95				
SEM	8.68		3.8					
MDC %	43%		16					
DSEC								
Day 1	35.7	18.9	42.6	20.2	7.3	−10.4 to 25.6	0.91	0.80 to 0.96
Day 2	38.9	18.2	43.3	17.7	4.8	−12.0 to 21.7	0.89	0.77 to 0.95
Mean diff (95% CI)	3.18	−17.50 to 23.86	0.69	−14.57 to 19.88				
ICC (95% CI)	0.85	0.68 to 0.93	0.89	0.75 to 0.95				
SEM	7.42		6.86					
MDC %	58%		45					

USEO: unilateral-leg stance, eyes open; DSEO: Double-leg stance, eyes open; DSEC: Double-leg stance, eyes closed; SD: standard deviation; diff: difference; ICC: intraclass correlation coefficient; CI: confidence interval; SEM: standard error of the measurement; MDC: minimum detectable change, expressed as a percentage of the Day 1 mean value.

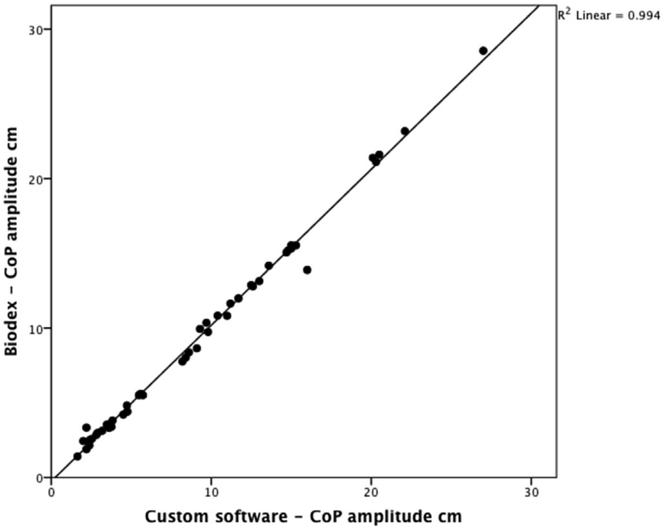

Fig. 2. Correlation of centre of pressure data from stacked WBBs placed on a Biodex force platform. The trials were performed simultaneously.

there is evidence of the WSS being both reliable and valid as a test of CoP range, further studies are required to assess its use as a valid measure of standing balance.

For reliability, the intra-device ICC was lower for the USEC (0.69) stance when compared to the other stances. For all other stances, the ICCs for the WSS were much higher (0.85–0.89). The inter-device ICCs were modest to excellent for all stances (0.70–0.91) except USEC, which was very poor (0.56 & 0.09). The reliability measurements in the USEC stance were complicated by the lack of completed trials. Clark et al. customized their protocols so the USEC stance was only tested for 10 s, leading to a higher number of completed trials.[10] Springer et al. showed that the ability to stand on one leg diminishes rapidly from the age of 50 years old with eyes open or closed.[20] The Wii-Fit software tests for 30 s cannot be customized, which means that there will always be a significant lack of completed trials, particularly in the older population or those with pathology. To further investigate the reliability of the WSS in the USEC stance, a study with a larger sample size would have to be performed to achieve the number of completed trials necessary to produce a significant result.

The validity of Nintendo's Stillness measure is more complicated. Whilst the WSS can provide information about the maximum amplitude of the subject's CoP, it is prone to underscoring a subject's balance ability through its reliance on extreme data points. For example, a subject with very good balance could complete a trial with the majority of their CoP points in a tight circle around the origin, save for one excursion. The score would only reflect the maximum limits of that one excursion. Furthermore, in elderly patients with balance impairment there can be a mismatch in the maximum sway amplitude and Stillness score making the WSS a poor measure in this population.

When the laptop's data was converted to a maximum X-Y amplitude, any value > 8.86 cm resulted in a negative derived Stillness score (DSS). Due to the fact that the WSS cannot produce a negative score, all negative DSS were converted to zero percent, replicating the limits of the WSS. This demonstrates the limitation inherent in the WSS with respect to poor scores. The responsiveness of the WSS is not known and is clearly a factor in the USEC stance. For this stance, there were only six participants who scored more than zero % for the DSS and only one of those managed it on both testing days.

The graph comparing the WSS to the maximum amplitude of the laptop CoP data (Fig. 1), had three obvious outliers in the low percentage range on the X-axis which were excluded from the line of best-fit calculation. This was done to maximize the correlation with most of the data and determine the measurement Nintendo were using for the WSS. One explanation for the outliers could be a lower sampling frequency of the Wii compared to the Labview

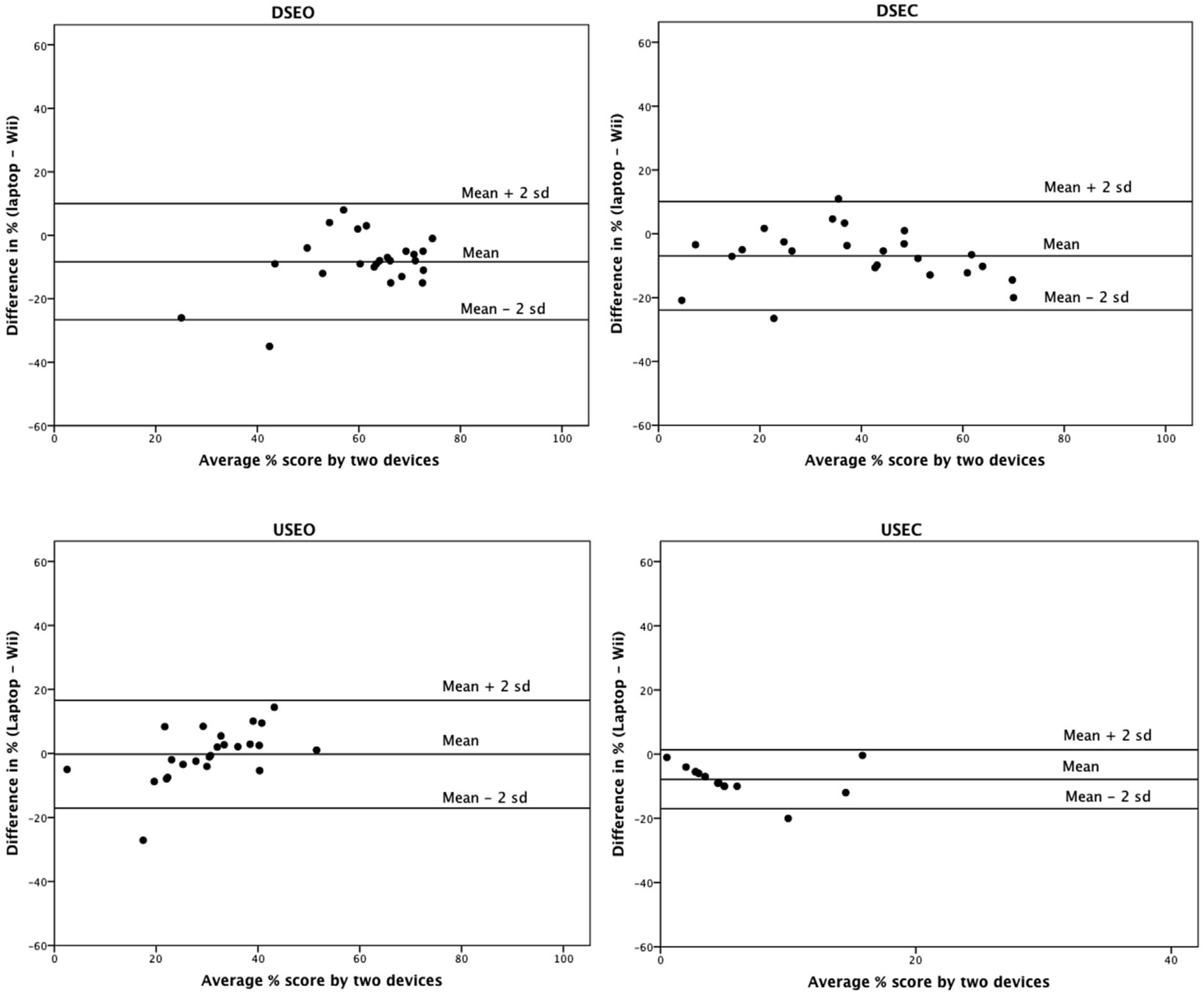

Fig. 3. Bland-Altman plots representing comparisons between a WBB linked to the Wii-fit and a WBB linked to a laptop (amplitude converted to a %) under the four testing conditions: (DSEO) double-limb, eyes open; (DSEC) double-limb, eyes closed; (USEO) single-limb, eyes open; (USEC) single-limb, eyes closed. The mean line represents the mean difference between the devices, with the upper and lower lines representing the limits of agreement (2SD).

software. During a trial that scored a low percentage, the changes in amplitude and velocity of the CoP are likely to be greater. During a large and fast excursion, the Wii-Fit may miss the extreme CoP co-ordinate, between samples whereas the Labview will record it leading to a larger amplitude measurement. It must be noted though that Nintendo have not published the sampling frequency of the WBB so this only remains a theory. It is partly for this reason that we would recommend that the median score be analysed as the subjects often had one outlying trial scoring much lower than the other two.

The Bland-Altman plots showed little difference between the two devices for three stances across the range of scores, suggesting good concurrent validity. (Fig. 3). However, in the USEC stance, the difference between the two machines is always negative, meaning that the WSS is higher than the DSS. This is likely to be due to the frequency mismatch as already mentioned. The Wii also appears to have a maximum CoP amplitude measurement threshold, which is well inside the CoP amplitudes of the trials. The laptop is able to measure beyond this threshold leading to the laptop registering a larger excursion and therefore a lower DSS than the WSS. Also, the method of conversion from the maximum COP into WSS may have ensured that the Bland-Altman plots indicated good agreement between the software analyses. However this conversion into the same units was necessary to compare the two measures. If the devices measured in the same units and did not require conversion, we would be comparing the measuring ability the devices and not the software.

The MDC results revealed that large increases in WSS would be required in this population to detect a significant improvement in performance. The MDC for the DSEO stance was 16%. Applying the MDC of the DSEO stance to the median of 65%, the minimum value needed to detect a clinical change would be a difference in WSS of 11% (16%*65). The younger population without pathology are likely to make small changes in their scores, suggesting that the WSS may not be so clinically relevant.

In conclusion, this study showed significant reproducibility of results in all but the USEC stance, suggesting that the WSS score can be a reliable measure of CoP in certain stances. As for the validity of the WSS when compared to custom-written software, there remain significant limitations with its use. Larger sample sizes are required to further analyse the reliability, particularly in the USEC stance. Further assessment of patients with pathology and undergoing rehabilitation are needed to determine if the WSS is a valid tool to evaluate improvements in CoP measurements. Ultimately, due to the current limitations of the WSS, we believe it would not be an ideal clinical measure of balance.

Authorship and conflicts of interest statement

There are no conflicts of interest to declare from any of the authors.

References

1. Nguyen ND, Pongchaiyakul C, Center JR, Eisman JA, Nguyen TV. Identification of high-risk individuals for hip fracture: a 14-year prospective study. *J Bone Miner Res*. 2005;20:1921–1928.
2. O'Loughlin JL, Robitaille Y, Boivin JF, Suissa S. Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly. *Am J Epidemiol*. 1993;137:342–354.
3. Bascuas I, Tejero M, Monleon S, Boza R, Muniesa JM, Belmonte R. Balance 1 year after TKA: correlation with clinical variables. *Orthopedics*. 2013;36:e6–e12.
4. Levinger P, Menz HB, Wee E, Feller JA, Bartlett JR, Bergman NR. Physiological risk factors for falls in people with knee osteoarthritis before and early after knee replacement surgery. *Knee Surg Sports Traumatol Arthrosc*. 2011;19: 1082–1089.
5. Schwartz I, Kandel L, Sajina A, Litinezki D, Herman A, Mattan Y. Balance is an important predictive factor for quality of life and function after primary total knee replacement. *J Bone Joint Surg Br*. 2012;94:782–786.
6. La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? A revisited Rasch analysis study. *Arch Phys Med Rehabil*. 2012;93:1209–1216.
7. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Phys Ther*. 2008;88:559–566.
8. Seimetz C, Tan D, Katayama R, Lockhart T. A comparison between methods of measuring postrual stability: force plates versus accelerometers. *Biomed Sci Instrum*. 2012;48:386–392.
9. Haas BM, Burden AM. Validity of weight distribution and sway measurements of the Balance Performance Monitor. *Physiother Res Int*. 2000;5:19–32.
10. Clark RA, Bryant AL, Pua Y, McCrory P, Bennell K, Hunt M. Validity and reliability of the Nintendo Wii balance board for assessment of standing balance. *Gait Posture*. 2010;31:307–310.
11. Bartlett HL, Ting LH, Bingham JT. Accuracy of force and center of pressure measures of the Wii Balance Board. *Gait Posture*. 2014;39:224–228.
12. Beuter A, Hernandez R, Rigal R, Modolo J, Blanchet PJ. Postural sway and effect of levodopa in early Parkinson's disease. *Can J Neurol Sci*. 2008;35:65–68.
13. Swanenburg J, de Bruin ED, Uebelhart D, Mulder T. Falls prediction in elderly people: a 1-year prospective study. *Gait Posture*. 2010;31:317–321.
14. Maki BE, Holliday PJ, Fernie GR. Aging and postural control. A comparison of spontaneous- and induced-sway balance tests. *J Am Geriatr Soc*. 1990;38:1–9.
15. Li Z, Liang YY, Wang L, Sheng J, Ma SJ. Reliability and validity of center of pressure measures for balance assessment in older adults. *J Phys Ther Sci*. 2016;28:1364–1367.
16. Huurnink A, Fransz DP, Kingma I, van Dieen JH. Comparison of a laboratory grade force platform with a Nintendo Wii Balance Board on measurement of postural control in single-leg stance balance tasks. *J Biomech*. 2013;46: 1392–1395.
17. Holmes JD, Jenkins ME, Johnson AM, Hunt MA, Clark RA. Validity of the Nintendo Wii(R) balance board for the assessment of standing balance in Parkinson's disease. *Clin Rehabil*. 2013;27:361–366.
18. Fleiss J. *The Design and Analysis of Clinical Experiments*. New York, USA: John Wiley & Sons; 1986.
19. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59: 12–19.
20. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for unipedal stance. *J Geriatr Phys Ther*. 2007;30:8.