Movement fluency metrics during multi-phase sit-to-walk and reach-to-grasp: test-retest reliability and agreement between laboratory-based and portable 3D motion analysis systems

Jacob Wells[a]. School of Health Sciences, University of East Anglia, UK. E-mail: Jacob.Wells@uea.ac.uk

Elizabeth Chandler[a]. School of Health Sciences, University of East Anglia, UK. E-mail: E.Chandler@uea.ac.uk

Allan Clark[b]. Norwich Medical School, University of East Anglia, UK. E-mail: Allan.Clark@uea.ac.uk

Canan Yüksel[a.] School of Health Sciences, University of East Anglia, UK. E-mail: C.Yuksel@uea.ac.uk

Merve Kizilay[a]. School of Health Sciences, University of East Anglia, UK. E-mail: M.Kizilay@uea.ac.uk

David Payne[a]. School of Health Sciences, University of East Anglia, UK. E-mail: David.Payne@uea.ac.uk

Nicola Hancock[a]. School of Health Sciences, University of East Anglia, UK. E-mail: N.Hancock@uea.ac.uk

Louise Gilbert[c.] Norwich Community Health and Care NHS Trust. UK. E-mail: Louise.Gilbert@nchc.nhs.uk

Valerie M Pomeroy[a]. School of Health Sciences, University of East Anglia, UK. E-mail: v.pomeroy@uea.ac.uk

Corresponding author
Professor Valerie M Pomeroy. School of Health Sciences, Queen's Building, University of East Anglia, Norwich Research Park, NR4 7TJ, UK. E-mail: v.pomeroy@uea.ac.uk

Keywords: reliability; reproducibility; validity; kinematics; smoothness; movement.

Manuscript word Count = 4432

Abstract word count: 250

Figures and tables combined = 6

Competing interests: The authors do not have any competing interests.

Highlights: provided in separate file

Abstract

Background: International consensus recommends use of kinematic metrics of movement during standardized functional tasks after stroke to ascertain whether rehabilitation is driving behavioral restitution or compensation. Quality of human movement can be characterized by fluency metrics including smoothness and hesitation. Before using these metrics in stroke rehabilitation it is important to find whether 'reference values', from healthy adults, are repeatable.

Research question: Do kinematic metrics of smoothness and hesitation have test-retest reliability during standardized functional tasks performed by healthy adults?

Methods: a correlational agreement study. Testing sessions separated by 1-4 weeks. Participants, 75 adults reporting no neurological or musculoskeletal diagnosis, performed standardized multi-phase sit-to-walk and reach-to-grasp with kinematic data collected simultaneously by optokinetic and video-based 3D motion analysis systems. Smoothness was derived using the Spectral Arc Length method. Hesitation was the maximum decrease in thorax forward velocity as percentage of peak value. Analysis used the intra-class correlation coefficient (ICC) with 95% confidence intervals [95% CI]. The clinically acceptable level of test-retest reliability was set as, ICC 0.75 [lower 95% CI 0.70 or above].

Findings: of the 75 participants, 72 completed both sessions. Neither smoothness nor hesitation kinematic metrics reached clinically acceptable test-retest reliability for either motion analysis system. Intra-participant variability was observed within sessions, e.g., mean coefficients of variation for sit-to-walk hesitation ranged from 13%- 22%.

Significance: Test-retest reliability of smoothness and hesitation kinematic metrics was clinically unacceptable for multi-phase tasks. Intra-participant within-session variation was observed. This intra-participant variation could hamper establishment of reference values for use in stroke rehabilitation.

1. Introduction

Kinematic metrics are recommended to identify the mechanism of action of post-stroke recovery of multi-phase functional abilities e.g., drinking from a cup and standing up from a chair [1]. This is because motor recovery of multi-phase functional activities can occur through: behavioral restitution, defined as restoration of movement patterns to those typical of people without motor impairment (normal) [2]; behavioral adaptation (use of atypical movement patterns in the impaired limb); and/or behavioral compensation (atypical use of other body parts to achieve the task) [2, 3]. Behavioral restitution is more desirable not least because adaptation/compensation early after stroke may hamper later motor recovery [4]. Hence the need for adequately powered studies [5] to produce kinematic metrics of stroke survivors' quality of movement during standardized multi-phase functional tasks with reference to the movement patterns of age-matched individuals without movement impairment (healthy adults [1, 2].

Quality of movement is characterized by smoothness /non-intermittency [6, 7] with the term fluency used to encompass kinematic metrics [8]. Fluency metrics of smoothness and hesitation have: distinguished between younger adults, older adults and older adults with risk of falling [8], and differ between sit-to-walk and sit-to-stand tasks performed by people after stroke [9]. However, test-retest reliability and smallest detectable change of movement fluency metrics applied to multi-phase functional activities is uncertain. Specifically in healthy adults (reference population for people with stroke) for the recommended smoothness metric, spectral arc length (SPARC) [7, 10] and a reported hesitation metric [8, 9].

Also important is consideration of how fluency metrics can be made in the clinical stroke rehabilitation settings of peoples' homes (during early supported discharge and community

rehabilitation) which are unsuitable for the recommended high-resolution digital optoelectronic systems [1]. Over the last decade there has been an increasing interest in use of inertial measurement units (IMUs) to undertake movement analysis outside of a specialized laboratory. A systematic review identified two studies that has used an IMU to provide a fluency metric, smoothness [11]. However, another systematic review reports that metrics derived from IMUs may only be valid during simple single-plane tasks [12]. As measurement of quality of movement during functional tasks is important in stroke rehabilitation [1] then potentially, video-based motion analysis employing pose estimation could be used [13] such as the Biokido system (Biokido, Ankara, Turkey). The system, depth camera, portable computer, and BioPose software, can be deployed in environments where people are located at least 1.5 meters from a single camera with clear space in-between. If there is agreement between metrics derived from the Biokido and optoelectronic systems, then research will not be restricted to specialist laboratories but opens potential for use in clinical settings.

The research questions were:

1. Do kinematic metrics of movement fluency, smoothness-SPARC and hesitation, have test-retest reliability when made with an optoelectronic system and the Biokido motion analysis system?
2. What is smallest detectable change (SDC) in smoothness-SPARC and hesitation?
3. Is there concurrent validity between fluency metrics derived from data collected from an optoelectronic system and the Biokido system?

2. Methods

2.1 Design and ethics

This was a correlational study. Data were collected simultaneously by an optoelectronic system (Vicon, Oxford Metrics, UK) and the Biokido system on two occasions separated by one to four weeks. The University Faculty of Medicine and Health Ethics Committee provided approval (Ref: 2019/20-044). All participants provided written informed consent.

2.2 Participants and sample size

Participants were aged 18 years or above and reported that they did not have a clinical diagnosis of a neurological or musculoskeletal pathology (healthy adults).

A sample size was calculated in accordance with the COSMIN initiative https://www.cosmin.nl/finding-right-tool/conducting-study-measurement-properties/. The sample size estimation of 54 complete data sets assumed an intra-class correlation coefficient (ICC) of 0.85 with a 95% confidence interval (95% CI) of no more than 0.15. Allowing for the possibility of a 20% participant attrition rate and a 15% poor data quality rate the sample size was set at 75 participants.

2.3. Data collection

The study setting was a university movement laboratory. Fluency metrics were made during two standardized everyday functional tasks: reach-to-grasp (RTG) and sit-to-walk (STW) (Table 1). A "go" signal consisting of a buzzer informed participants when to start and provided a voltage signal that stamped the Vicon data.

The rationale for choice of multi-phase RTG was that it is essential for everyday activity involving the arm and hand for grasping, using and replacing objects and STW because it requires fluent transition between sit-to-stand and gait initiation that is challenging for stroke survivors. Lack of fluent transition results in performance of sit-to-stand and then gait initiation.

All participants wore a tight-fitting sleeveless top and shorts. Loose material was not permitted to reduce the possibility of movement artefact for the optoelectronic system and obscuration of shape detected by the marker-less system.

Participants undertook practice trials to ensure correct task performance then were asked to complete 3-5 trials of each RTG condition and 5-10 trials of the STW condition (Table 1). In-between trials, the researchers undertook quality checks (section 2.6). When a recording error was suspected that trial was discarded, and a participant was asked to undertake an additional trial.

Vicon data were collected at 100Hz using ten T-series infrared cameras. Eight cameras were wall-mounted to provide comprehensive coverage of the laboratory space, ensuring marker visibility during movement tasks. Two additional cameras were mounted on tripods and positioned in front of the participant, angled to maintain visibility of the reflective markers during sit-to-walk tasks, where marker occlusion was more likely.

In addition, movement data were collected using the marker-less Biokido system. A Biopose Intel RealSense D435 depth camera was placed directly in front of the participant, capturing a depth point-cloud at 90 Hz and saving it as an image.

Experienced researchers (two physiotherapists and two clinical movement analysts) applied 39 reflective markers to participants in accordance with the Vicon Plug-In-Gait model [14, 15].

2.6. Biomechanical data quality checks

Data quality was a priority at each stage of the study. During live trials, two members of the research team observed the participant. Trials were repeated if a movement was perceived as unnatural, such as body weight shifts, loss of focus, or clear deviations from previous attempts. Due to the dynamic nature of the trials and the high possibility of obscuring the pelvic markers while rising, Vicon data was reviewed after each trial to monitor marker visibility. Trials with high levels of dropout of any marker (>25 frames) were repeated. This protocol was implemented for the final 45 participants due to poorer-than-expected data quality from the first 30 participants. It was not possible to review the marker-less data from the Biokido system between each trial. Post-processing, two members of the research team graphically reviewed each trial for both the Biokido and Vicon systems. Anomalies were discussed, and decisions were made on whether to include each trial. Each metric required only a single input: the thorax marker for the sit-to-walk trials and the midway point between wrist markers for the reach-to-grasp trials. Due to the constraints of the task, erroneous data were identifiable through irregularities in trajectory position and velocity.

2.7 Data processing

Researchers processed Vicon system data using Vicon Nexus 2.7.0 with the Plug-In-Gait model, which automatically calculates joint segments as part of its standard processing. However, for this study, only the thorax marker for STW trials and the wrist midpoint (midway between the medial and lateral wrist markers) for RTG trials were used to calculate

velocity profiles. If there were missing frames then the Thorax markers were filled using a rigid body fill, the wrist markers were filled using a pattern fill from the opposing wrist marker. These are built in in Vicon. Only gaps less than 20 frames were filled. Raw trajectory Vicon data were not filtered before being exported to MATLAB 2023a.

A member of the Biokido manufacturing team wrote and ran the software code used to process Biokido data independently of the researchers. Each frame of the point cloud was passed through the Biokido proprietary deep learning model to obtain the location of the thorax and wrist landmarks. For STW, the depth image data was used to obtain the location of body landmarks with respect to the camera position and orientation. For RTG, the left infrared image was obtained since the depth image was inadequate for detecting wrist landmarks. Body landmarks detected in the infrared image were then mapped onto the depth image to obtain their position with respect to the camera position and orientation. Then any gaps in the trajectory position data were filled using linear interpolation. If a trial contained more than 10% missing frames, then the Biokido team discarded it. A 5^{th} order Butterworth lowpass filter was then applied by the Biokido team. Numerical differentiation to the trajectory position data was used to calculate velocity using the timestamps of the individual image frames.

2.8. Metrics extraction

For STW trials, the data were cropped from the timepoint at which the forward trajectory of the thorax exceeded 5% of its maximum velocity until the second foot strike. For RTG trials the baseline was the five seconds immediately before the buzzer. Start of the movement was defined as the timepoint when forward velocity of the wrist midpoint exceeded the baseline mean plus two standard deviations for a period of at least 30ms.

Movement end, hand back to start position after replacing the telephone in its original position (Table 1), was determined using the same onset algorithm after mirroring the data. As there was not a baseline period for the Biokido system, movement onset was when the wrist midpoint velocity exceeded 5% of its peak velocity. RTG trials were cropped from movement start to end.

Smoothness was derived using the Spectral Arc Length method (SPARC) [7, 16]. SPARC is recommended to measure smoothness of reach-to-grasp tasks after stroke and any movement type [7]. It is reported as able to distinguish between gait smoothness in people with Parkinson's disease and healthy adults and elderly people who do and do not fall [17, 18]. The smoothness-SPARC calculation was made using open-source MATLAB scripts. Vicon data were not filtered before running the MATLAB script. Biokido data were filtered by the manufacturer using their proprietary algorithms. A higher smoothness-SPARC value indicates greater movement fluency.

RTG smoothness-SPARC was derived from forward velocity of the wrist mid-point for the Vicon system and pose estimation of wrist for the Biokido system. STW smoothness-SPARC was derived from thorax forward velocity and pose estimation of thorax for the Vicon and Biokido systems respectively.

STW hesitation was defined as: the maximum decrease in forward velocity of the thorax, expressed as a percentage of the peak value. Other studies have used forward velocity of the center of mass [8] and clavicle [9] but the focus here was on the thorax as that is tracked by the Biokido system and the Vicon Plug-in-Gait model. Purpose-written MATLAB scripts

were used. A lower value indicates greater movement fluency. Hesitation is not a characteristic of RTG so was not calculated.

2.9. Statistical analysis

The statistical analysis was undertaken using Stata 18.0 SE (Texas). A participant's inclusion in the test-retest reliability analysis required the presence of at least one trial from each visit for a system. For concurrent validity only trials where a metric was extracted from both Vicon and Biokido data were used to ensure use of identical trials.

For test-retest reliability and concurrent validity the ICC (95% CI) was estimated using a single measurement two-way random effect analysis of variance using all the data available on each set of trials with Vicon and Biokido data. Values for ICC were interpreted as indicating good (acceptable) reliability when ranging from 0.75 to 0.90 and excellent when greater than 0.90 [19]. Interpretation of clinically acceptable reliability also considered that the lower 95% CI needed to be 0.70 or greater for eventual implantation into clinical practice. Additionally, for test-retest reliability the Bland-Altman limits of agreement (LOA) were estimated [20]. We used the equations for replicated data collected in pairs, where the underlying value of quantity being measured is changing. The concurrent validity of the Vicon and Biokido were estimated using the Bland-Altman limits of agreement (LOA).

The metrics' smallest detectable change (SDC) was calculated using: $1.96\ x\ \sqrt{2}\ x\ SEM$. Where SEM is the standard error of measurement and is calculated as $SEM = SD$ x $\sqrt{1-r}$, where SD is the standard deviation and r is the correlation between the two time-points.

3. Results

The target sample size of 75 participants was achieved. Participants mean (SD) age was 47.8 (17.4) years (Table 2). Only three participants withdrew (4%) before session two.

Not all trials were of sufficient quality to provide metrics for statistical analysis. The combined reasons for the two systems were: irregularities in expected trajectories due to noisy Vicon signals, the 'go' signal did not mark the Vicon data, frames in Biokido data 'dropped', reflective markers were not visible to the Vicon system, and a recording fault made data 'unrecoverable'. Between 184 and 363 quality trials per kinematic metric were available for statistical analysis (supplementary data Tables 1-6). Mean values for each kinematic metric, sessions one and two, are provided in Table 3. The number of matched trials per participant for each statistical analysis are provided in Tables 4, 5 and 6. The trajectories of an exemplar participant are provided in the supplementary data (pages S19-S21).

3.1. Test-retest reliability (question 1)

The ICCs [95% Cis] for the metrics derived from the Biokido system on visits 1 and 2 ranged from 0.40 [0.27,0.49] for RTG-MD smoothness-SPARC to 0.59 [0.49,0.69] for STW smoothness-SPARC. The ICCs [95% Cis] for metrics derived from the Vicon system ranged from 0.44 [0.31,0.51] for RTG-MD smoothness-SPARC to 0.63 [0.53,0.71] for STW smoothness-SPARC. No metric derived from either system reached acceptable test-retest reliability (Table 4).

The LOAs were wide (Table 5). For metrics derived from the Biokido system, LOA ranged from -0.36 (-2.51 to 1.80) for RTG-MD smoothness-SPARC to 2.13 (-54.82 to 59.09)

for STW hesitation. Vicon system-derived metrics LOA ranged from -0.23 (-1.95 to 1.50) for RTG-MND smoothness-SPARC to 4.26 (-34.63 to 43.14) for STW hesitation.

3.2. Smallest detectable change (question 2)

Findings for SDC are provided in Table 4. For smoothness-SPARC derived from the Biokido system the SDCs ranged from 0.41 units for STW to 2.29 units for RTG-MD. Those for the Vicon system were smaller ranging from 0.25 units for STW to 1.71 units for RTG-MD. For STW hesitation the SDC for the Biokido system was 62.83% and for the Vicon system was 41.20%.

3.3 Concurrent validity (question 3)

The ICCs [95% CI] for concurrent validity of the Biokido and Vicon systems ranged from 0.39 [0.15,0.56] for RTG-MND smoothness-SPARC to 0.58 [0.52,0.65] for STW smoothness-SPARC. Acceptable concurrent validity was not attained by any of the metrics (Table 6).

4. Discussion

Test-retest reliability did not reach the clinically acceptable value for metrics of smoothness-SPARC and hesitation during standardized multi-phase RTG and STW performed by healthy adults (question 1). And the LOAs were wide. The SDCs were large in respect of the range of possible values (question 2). Concurrent validity between metrics derived from the Vicon and Biokido systems did not reach the clinically acceptable value (question 3). Observed intra-participant, within session, variation could hamper establishment of reference values for stroke rehabilitation practice and clinical research.

The SPARC values found in this study were lower than previously reported for healthy adults as ranging from mean (standard deviation) of -1.48 (0.05) to 1.60 (0.12) [21-23]. The currently reported SPARC values were also lower than previously reported for people with stroke ranging from a median (Quartile 1, quartile 3) of -1.82 (-2.14, -9.83) to a mean (standard deviation of -1.480 (0.059) [21, 22]. The difference in reported values is expected due to differences in experimental tasks. The earlier studies derived SPARC values from single phase point-to-point reaching tasks whereas the current study used multi-phase tasks with more inherent intermittencies: STW and reaching (Table 1). Tasks with more intermittencies are less smooth and therefore have lower SPARC values [7].

The findings of clinically unacceptable concurrent validity between an optoelectronic system and a video system with BioPose software differ from conclusions of previous reports. For example, a report of concurrent validity during treadmill walking of 30 healthy adults [24] despite all the lower 95% CIs being below 0.70 and therefore not constituting clinically acceptable agreement [19]. This different interpretation of acceptability of ICC values is illustrated by another study of camera-based and optoelectronic systems [25]. Conversely, some temporal spatial gait parameters [25, 26] and sagittal range of motion of lower limb segments [27] have clinically acceptable ICC values for concurrent validity. But then other studies reporting agreement between video-based systems and optoelectronic systems for some joint angle metrics do not provide ICC values [28-30]. So, the present findings of lack of clinically acceptable concurrent validity concur with some but not all earlier reports irrespective of the different metrics.

It is conceivable that clinically unacceptable concurrent validity was found in this study because of differences in data processing techniques for the two motion analysis systems. Compared with other smoothness metrics SPARC is less affected by aspects like movement

duration and measurement noise [31, 32]. That said, it's not immune to the effects of data processing. Like any metric that relies on velocity profiles, SPARC is sensitive to how the signal is defined and processed. Changes in filtering or how derivatives are calculated can alter SPARC values [33]. So, differences in processing, reported in the methods section may have contributed to the finding of clinically unacceptable concurrent validity. Key differences between Biokido and Vicon processing, respectively, that could have impacted SPARC values are: linear interpolation fill for missing frames versus rigid body fill for thorax marker and pattern fill for wrist marker; only gaps less than 10% of the whole trial filled compared with 25 frames or less; pose estimation versus forward velocity; data filtered before SPARC calculated compared with no prior filtering; no baseline period compared to a baseline period; and RTG onset when wrist midpoint velocity exceed 5% of its peak compared to timepoint when forward velocity of wrist midpoint exceeded baseline mean plus 2SD for at least 30ms. There are likely unknown impacts due to the proprietary nature of the Biokido data processing techniques.

It is also conceivable that differences in processing, reported in the methods section, contributed to the finding of clinically unacceptable concurrent validity. Key differences between Biokido and Vicon processing, respectively, that could have impacted SPARC values are: linear interpolation fill for missing frames versus rigid body fill for thorax marker and pattern fill for wrist marker; only gaps less than 10% of the whole trial filled compared to less than 20 frames; pose estimation versus forward velocity; data filtered before SPARC calculated compared with no prior filtering; no baseline period compared to a baseline period; and RTG onset when wrist midpoint velocity exceeds 5% of its peak compared to timepoint when forward velocity of wrist midpoint exceeded baseline mean plus 2SD for at least 30ms. There are likely unknown impacts due to the proprietary nature of the Biokido data processing techniques.

For test-retest reliability of smoothness-SPARC and hesitation the authors are only aware of one previous study that investigated test-retest reliability [23]. Smoothness-SPARC was found to have a test-retest, ICC of 0.75 or above for six functional reaching tasks but the 95% CIs were not provided so exact comparison with the present findings cannot be made [23]. Direct comparison is also limited by the ages of the participants and sample sizes. First, ages of participants were mean of 22±4 years [23] compared with 47.8 (17.4, 20.5–85.4) years in this study. Consequently, this could have impacted the apparent greater variability as it is a known characteristic of older people (for example [34]). Additionally, reference values from an exclusively younger portion of the adult population are limited in clinical application including for stroke rehabilitation. Second, the earlier study recruited 19 participants whereas our a priori sample size estimation was that 54 complete data sets (sessions 1 and 2) were required. So, the earlier study only recruited 35% of the estimated sample size. It is possible therefore that replication of the earlier study with an adequate number of participants would provide a more precise estimate of the ICC especially if the 95% CIs are provided.

There are two other important differences between the two studies. The earlier study filtered marker position data before calculating SPARC whilst the present study did not. The earlier study divided the upper limb tasks into reaching and object manipulation whilst the present study calculated the SPARC metric on the entire movement task. The inflection points in the task used in the present study certainly afford a greater number of intermittencies that likely influenced the SPARC values. They also present opportunities for greater variation in movement strategies than just the reaching component. Although SPARC is affected by movement speed and amplitude [31] neither study controlled these parameters.

Potential reasons for the present finding of clinically unacceptable test-retest reliability are: kinematic metrics might not reach the clinically acceptable level of reliability; intra-

participant variability in functional task performance is characteristic of human movement; and/or smoothness-SPARC might not be the appropriate metric during STW and RTG.

First, kinemetric metrics derived from optoelectronic and video-based systems, other than smoothness-SPARC and hesitation, might not reach the clinically acceptable level of reliability. For example, a systematic review of kinematic assessment of upper limb movement after stroke found that only one of the 30 different metrics investigated for test-retest reliability had adequate clinical test-retest reliability from a study of at least moderate methodological quality [35]. Since publication of this review, studies using optoelectronic and/or video-based systems have reported on the test-retest reliability in healthy adults of kinematics during everyday functional tasks relevant to stroke rehabilitation (not isolated limb movements such as hand-to-mouth). For example, an investigation of test-retest reliability of lower limb angles during functional activity [24] but acceptable test-retest reliability is unsupported as none of the ICCs reached 0.75 with a lower 95% CI of at least 0.70 [19]. Other studies of healthy adults report: (a) repeatability of joint angles during walking but did not calculate ICCs [36]; (b) ICC values above 0.75 with lower 95% CIs of 0.70 or greater for 4 of 15 joint angles during gait, 1 of 15 joint angles during sit-to-stand [37], and 18 of 60 kinematic metrics during running on a treadmill [38]; (c) acceptable ICC values for four of seven metrics (phase durations and peak velocity) across two reaching tasks [39]; and (d) no ICCs reached the acceptable level for angles of trunk, hip and knee whilst walking overground and on a treadmill [40]. Therefore, not all kinematic metrics during functional tasks have been found to possess clinically acceptable test-retest reliability.

The second potential reason for clinically unacceptable test-retest reliability for fluency metrics here is that intra-participant variability is a characteristic of human movement: “bliss

of motor abundance" [41]. The intra-participant variability, mean CVs (range), for the sets (six conditions in two sessions) calculated from Biokido data are generally higher than for those calculated from the optoelectronic system (Table 3). Likely due to less precision in pose estimation than in 3D tracking of markers. So, we focus on Vicon data only. Although variability was lower for metrics derived from the optoelectronic system it was relatively high. For example, 23% of the mean CV of an individual's trials for RTG-MND smoothness-SPARC (Vicon session one, Table 3) and 47% for STW hesitation (Vicon session one, Table 3). Possibly there was an insufficient number of trials to reach data stability [42]. However, the recommendation of 2-5 trials of a drinking task [43] suggests otherwise. For STW the number of trials required to consider intra-participant variability is less clear although for gait initiation 1-10 trials were required to reach an ICC of 0.75 depending on the metric [44]. In the absence of specific data in respect of STW no conclusion can be drawn.

Third, the possibility that smoothness-SPARC might be inappropriate for multi-phase STW and RTG. A comparison of use of SPARC and the log dimensionless jerk (LDJ) found the latter to have greater discriminative ability and responsiveness to change in stroke survivors [45]. Furthermore, the LDJ could be the most appropriate metric for multi-phase movement tasks [46]. Of note, the recommendation for use of smoothness-SPARC above other metrics came from a simulation rather than human participant study [10].

Considering that variability is a characteristic of human movement it has been said that "only nonlinear measures that distinguish random from controlled variability would identify that variability increase is desirable" [47]. Non-linear analyses, (such as Lyapunov exponent $(\lambda_1)_1$, entropy and fractal analysis) regard movement variability as the metric of interest and

not noise [48]. Sample entropy, in particular, has been used to analyze postural stability and might be useful for postural transition tasks [47]. These observations are pertinent particularly to STW as it is a task of transition i.e., sit, transition, walk and people with impaired movement control tend to have a pattern of sit, transition, stand, transition, walk.

4.1. Strengths and limitations

A strength of this study is that the sample size, based on a power calculation, was achieved. Another potential strength is the use of more than one operator placing markers across the two sessions to reflect multi-center clinical trials when there are different assessors in different centers, and clinical practice when the same therapist conducting assessments on two or more separate occasions cannot be guaranteed. Thereby these findings are relevant to the metrics' intended use. Conversely, that more than one operator placed markers across the two sessions could have contributed to metrics' variability [49]. But, recognizing that experience mitigates differences in marker placement precision [49] all operators in this study were physiotherapists/clinical movement analysts with experience and training in marker placement. Furthermore, the potential for variability was reduced for the metrics of smoothness-SPARC and hesitation because they were not calculated from either the center of mass or joint angles. Indeed, none of the lower limb markers associated with variability in metrics [14, 50] were involved in the calculations.

5. Conclusion

Inadequate test-retest clinical reliability was found for smoothness-SPARC and hesitation in multi-phase RTG and STW. Observed intra-participant variation within sessions impacted reliability. Yet, variability is a characteristic of human movement. Consequently, the calculation of metrics of movement fluency may need to consider identifying the number of

trials needed to reach data stability for STW and RTG. This study also found inadequate clinical concurrent validity of the two metrics derived from different movement analysis systems.

Acknowledgements

The research was supported financially by the National Institute for Health and Care Research (NIHR) Brain Injury MedTech Co-operative based at Cambridge University Hospitals NHS Foundation Trust and University of Cambridge. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Canan Yüksel and Merve Kizilay were supported financially by the Ministry of National Education of the Republic of Türkiye. This support, provided under the Turkish state's Law No. 1416. The funding body was not involved in the design, execution, interpretation, or writing of the study.

We thank our industrial collaborators, Biokido. The industrial team were not co-applicants for the funding application to the NIHR Brain Injury MedTech Co-operative. Neither were they contributors to the scientific protocol, data collection or statistical analysis. They provided the Biokido motion analysis system used in the research for no cost. We specifically thank, Ozgur Reyhanoglu and Dr Sarper Alkan of Biokido. The former for the in-kind resources that enabled Dr Sarper Alkan to write the software code he used to process the Biokido data and extract values for statistical analysis.

The research team also thank the people who gave their personal time to participate in this study. Their generosity enabled this project.

Data statement

The data for the study reported in this manuscript are available from the corresponding author upon reasonable request. They will also be deposited in an open data repository.

References

[1] G. Kwakkel, N.A. Lannin, K. Borschmann, C. English, M. Ali, L. Churilov, et al., Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable, International Journal of Stroke 12(5) (2017) 451-461.
[2] G. Kwakkel, C. Stinear, B. Essers, M. Munoz-Novoa, M. Branscheidt, R. Cabanas-Valdes, et al., Motor rehabilitation after stroke: European Stroke Organisation (ESO) consensus-based definition and guiding framework, Eur Stroke J 8(4) (2023) 880-894. https://www.ncbi.nlm.nih.gov/pubmed/37548025.
[3] M.F. Levin, J.A. Kleim, S.L. Wolf, What do motor recovery and compensation mean in patients following stroke?, Neurorehabil Neural Repair 23(4) (2009) 313-319.
[4] R.J. Nudo, Recovery after brain injury: mechanisms and principles, Frontiers in Human Neuroscience 7 (2013).
[5] S.H. Scott, C.R. Lowrey, I.E. Brown, S.P. Dukelow, Assessment of neurological impairment and recovery using statistical models of neurologically healthy behavior, Neurorehabil Neural Repair 37(6) (2023) 394-408. https://www.ncbi.nlm.nih.gov/pubmed/35932111.
[6] T.J. Sejnowski, Making smooth moves, Nature 394 (1998) 725-726.
[7] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, E. Burdet, On the analysis of movement smoothness, Journal of Neuroengineering and Rehabilitation 12 (2015) 112.
[8] A. Kerr, V.P. Pomeroy, P.J. Rowe, P. Dall, D. Rafferty, Measuring movement fluency during the sit-to-walk task, Gait & Posture 37(4) (2013) 598-602.
[9] E.A. Chandler, T. Stone, V.M. Pomeroy, A.B. Clark, A. Kerr, P. Rowe, et al., Investigating the relationships between three important functional tasks early after stroke: movement characteristics of sit-to-stand, sit-to-walk, and walking, Front Neurol 12 (2021) 660383. https://www.ncbi.nlm.nih.gov/pubmed/34054703.

[10] M.I. Mohamed Refai, M. Saes, B.L. Scheltinga, J. van Kordelaar, J.B.J. Bussmann, P.H. Veltink, et al., Smoothness metrics for reaching performance after stroke. Part 1: which one to choose?, J Neuroeng Rehabil 18(1) (2021) 154. https://www.ncbi.nlm.nih.gov/pubmed/34702281.
[11] A. Martino Cinnera, P. Picerno, A. Bisirri, G. Koch, G. Morone, G. Vannozzi, Upper limb assessment with inertial measurement units according to the international classification of functioning in stroke: a systematic review and correlation meta-analysis, Top Stroke Rehabil 31(1) (2024) 66-85. https://www.ncbi.nlm.nih.gov/pubmed/37083139.
[12] J. Ghattas, D.N. Jarvis, Validity of inertial measurement units for tracking human motion: a systematic review, Sports Biomech (2021) 1990383. https://www.ncbi.nlm.nih.gov/pubmed/34698600.
[13] J. Stenum, K.M. Cherry-Allen, C.O. Pyles, R.D. Reetzke, M.F. Vignos, R.T. Roemmich, Applications of Pose Estimation in human health and performance across the lifespan, Sensors 21(21) (2021) 7315. https://www.ncbi.nlm.nih.gov/pubmed/34770620.
[14] R. Davis, S. Ounpuu, D. Tyburski, J.R. Gage, A gait analysis collection and reduction technique, Human Movement Science 10 (1991) 575-87.
[15] M.P. Kadaba, H.K. Ramakrishnan, M.E. Wootten, Measurement of lower extremity kinematics during level walking, J Orthop Res 8(3) (1990) 383-92. https://www.ncbi.nlm.nih.gov/pubmed/2324857.
[16] A. Melendez-Calderon, C. Shirota, S. Balasubramanian, Estimating movement smoothness from inertial measurement units, Front Bioeng Biotechnol 8 (2020) 558771. https://www.ncbi.nlm.nih.gov/pubmed/33520949.
[17] Y. Beck, T. Herman, M. Brozgol, N. Giladi, A. Mirelman, J.M. Hausdorff, SPARC: a new approach to quantifying gait smoothness in patients with Parkinson's disease, J Neuroeng Rehabil 15(1) (2018) 49. https://www.ncbi.nlm.nih.gov/pubmed/29914518.
[18] A.I. Figueiredo, G. Balbinot, F.O. Brauner, A. Schiavo, R.R. Baptista, A.S. Pagnussat, et al., SPARC metrics provide mobility smoothness assessment in oldest-old with and without a Hhstory of falls: a case control study, Front Physiol 11 (2020) 540. https://www.ncbi.nlm.nih.gov/pubmed/32587523.
[19] T.K. Koo, M.Y. Li, A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research, J Chiropr Med 15(2) (2016) 155-63. https://www.ncbi.nlm.nih.gov/pubmed/27330520.
[20] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Statistical Methods in Medical Research 8 (1999) 135-160.
[21] N. Bayle, M. Lempereur, E. Hutin, D. Motavasseli, O. Remy-Neris, J.M. Gracies, et al., Comparison of various smoothness metrics for upper limb movements in middle-aged healthy subjects, Sensors 23(3) (2023) 1158. https://www.ncbi.nlm.nih.gov/pubmed/36772197.
[22] M. Saes, M.I. Mohamed Refai, J. van Kordelaar, B.L. Scheltinga, B.F. van Beijnum, J.B.J. Bussmann, et al., Smoothness metric during reach-to-grasp after stroke: part 2. longitudinal association with motor impairment, J Neuroeng Rehabil 18(1) (2021) 144. https://www.ncbi.nlm.nih.gov/pubmed/34560898.
[23] S.M. Engdahl, D.H. Gates, Reliability of upper limb movement quality metrics during everyday tasks, Gait & Posture 71 (2019) 253-260. https://www.ncbi.nlm.nih.gov/pubmed/31096132.

[24] S. Liang, Y. Zhang, Y. Diao, G. Li, G. Zhao, The reliability and validity of gait analysis system using 3D markerless pose estimation algorithms, Front Bioeng Biotechnol 10 (2022) 857975. https://www.ncbi.nlm.nih.gov/pubmed/36032709.
[25] M. Yamamoto, K. Shimatani, M. Hasegawa, Y. Kurita, Y. Ishige, H. Takemura, Accuracy of temporo-spatial and lower limb joint kinematics parameters using OpenPose for various gait patterns with orthosis, IEEE Trans Neural Syst Rehabil Eng 29 (2021) 2666-2675. https://www.ncbi.nlm.nih.gov/pubmed/34914592.
[26] A.C. Aberg, F. Olsson, H.B. Ahman, O. Tarassova, A. Arndt, V. Giedraitis, et al., Extraction of gait parameters from marker-free video recordings of Timed Up-and-Go tests: Validity, inter- and intra-rater reliability, Gait Posture 90 (2021) 489-495. https://www.ncbi.nlm.nih.gov/pubmed/34628196.
[27] Z. Ripic, M. Nienhuis, J.F. Signorile, T.M. Best, K.A. Jacobs, M. Eltoukhy, A comparison of three-dimensional kinematics between markerless and marker-based motion capture in overground gait, J Biomech 159 (2023) 111793. https://www.ncbi.nlm.nih.gov/pubmed/37725886.
[28] S. Chakraborty, A. Nandy, T. Yamaguchi, V. Bonnet, G. Venture, Accuracy of image data stream of a markerless motion capture system in determining the local dynamic stability and joint kinematics of human gait, J Biomech 104 (2020) 109718. https://www.ncbi.nlm.nih.gov/pubmed/32151378.
[29] K. Song, T.J. Hullfish, R. Scattone Silva, K.G. Silbernagel, J.R. Baxter, Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements, J Biomech 157 (2023) 111751. https://www.ncbi.nlm.nih.gov/pubmed/37552921.
[30] R.M. Kanko, E.K. Laende, E.M. Davis, W.S. Selbie, K.J. Deluzio, Concurrent assessment of gait kinematics using marker-based and markerless motion capture, J Biomech 127 (2021) 110665. https://www.ncbi.nlm.nih.gov/pubmed/34380101.
[31] A. Caronni, P. Arcuri, I. Carpinella, A. Marzegan, T. Lencioni, M. Ramella, et al., Smoothness of movement in idiopathic cervical dystonia, Sci Rep 12(1) (2022) 5090. https://www.ncbi.nlm.nih.gov/pubmed/35332258.
[32] N. Hogan, D. Sternad, Sensitivity of smoothness measures to movement duration, amplitude, and arrests, J Mot Behav 41(6) (2009) 529-34. https://www.ncbi.nlm.nih.gov/pubmed/19892658.
[33] S. Singh, J. Bible, Z. Liu, Z. Zhang, R. Singapogu, Motion smoothness metrics for cannulation skill assessment: what factors matter?, Front Robot AI 8 (2021) 625003. https://www.ncbi.nlm.nih.gov/pubmed/33937348.
[34] M. Callisaya, B. L, M.D. Schmidt, J.L. McGinley, V.K. Srikanth, Ageing and gait variability - a population-based study of older people, Age Ageing 39 (2010) 191-197.
[35] A. Schwarz, C.M. Kanzler, O. Lambercy, A.R. Luft, J.M. Veerbeek, Systematic review on kinematic sssessments of upper limb movements after stroke, Stroke 50(3) (2019) 718-727. https://www.ncbi.nlm.nih.gov/pubmed/30776997.
[36] R.M. Kanko, E. Laende, W.S. Selbie, K.J. Deluzio, Inter-session repeatability of markerless motion capture gait kinematics, J Biomech 121 (2021) 110422. https://www.ncbi.nlm.nih.gov/pubmed/33873117.
[37] J.A. Deane, E. Papi, A.T.M. Phillips, A.H. McGregor, Reliability and minimal detectable change of the 'Imperial Spine' marker set for the evaluation of spinal and lower limb kinematics in adults, BMC Res Notes 13 (2020). https://www.ncbi.nlm.nih.gov/pubmed/33092633.

[38] C. Bramah, S.J. Preece, N. Gill, L. Herrington, The between-day repeatability, standard error of measurement and minimal detectable change for discrete kinematic parameters during treadmill running, Gait Posture 85 (2021) 211-216. https://www.ncbi.nlm.nih.gov/pubmed/33610824.
[39] E. Niechwiej-Szwedo, M. Nouredanesh, J. Tung, Test-retest repeatability reveals a temporal kinematic signature for an upper limb precision grasping task in adults, Hum Mov Sci 75 (2021) 102721. https://www.ncbi.nlm.nih.gov/pubmed/33271492.
[40] H. Tamura, R. Tanaka, H. Kawanishi, Reliability of a markerless motion capture system to measure the trunk, hip and knee angle during walking on a flatland and a treadmill, J Biomech 109 (2020) 109929. https://www.ncbi.nlm.nih.gov/pubmed/32807306.
[41] M.L. Latash, The bliss (not the problem) of motor abundance (not redundancy), Exp Brain Res 217(1) (2012) 1-5. https://www.ncbi.nlm.nih.gov/pubmed/22246105.
[42] Y. Chen, S. Garcia-Vergara, A. Howard, Number of trials necessary to achieve performance stability in a reaching kinematics movement analysis game, J Hand Ther 33(3) (2020) 371-377. https://www.ncbi.nlm.nih.gov/pubmed/31519383.
[43] G.E. Frykberg, H. Grip, M. Alt Murphy, How many trials are needed in kinematic analysis of reach-to-grasp?-A study of the drinking task in persons with stroke and non-disabled controls, J Neuroeng Rehabil 18(1) (2021) 101. https://www.ncbi.nlm.nih.gov/pubmed/34130716.
[44] J. Seuthe, N. D'Cruz, P. Ginis, R. Blobaum, B. Weisser, G. Deuschl, et al., How many gait initiation trials are necessary to reliably detect anticipatory postural adjustments and first step characteristics in healthy elderly and people with Parkinson's disease?, Gait & Posture 88 (2021) 126-131. https://www.ncbi.nlm.nih.gov/pubmed/34034024.
[45] M. Germanotta, C. Iacovelli, I. Aprile, Evaluation of gait smoothness in patients with stroke undergoing rehabilitation: comparison between two metrics, Int J Environ Res Public Health 19 (2022) 13440. https://www.ncbi.nlm.nih.gov/pubmed/36294017.
[46] P. Gulde, J. Hermsdorfer, Smoothness metrics in complex movement tasks, Front Neurol 9 (2018) 615. https://www.ncbi.nlm.nih.gov/pubmed/30258393.
[47] C.T. Gibbons, P.G. Amazeen, A.D. Likens, Distinguishing two types of variability in a sit-to-stand task, Motor Control 24(1) (2020) 168-188. https://www.ncbi.nlm.nih.gov/pubmed/31525730.
[48] K.J. Brink, K.L. McKenzie, A.D. Likens, Nonlinear analyses distinguish load carriage dynamics in walking and standing: s systematic review, J Appl Biomech 38(6) (2022) 434-447. https://www.ncbi.nlm.nih.gov/pubmed/36170973.
[49] M. Fonseca, X. Gasparutto, G. Grouvel, A. Bonnefoy-Mazure, R. Dumas, S. Armand, Evaluation of lower limb and pelvic marker placement precision among different evaluators and its impact on gait kinematics computed with the Conventional Gait Model, Gait Posture 104 (2023) 22-30. https://www.ncbi.nlm.nih.gov/pubmed/37307761.
[50] M. Fonseca, M. Bergere, J. Candido, F. Leboeuf, R. Dumas, S. Armand, The Conventional Gait Model's sensitivity to lower-limb marker placement, Sci Rep 12(1) (2022) 14207. https://www.ncbi.nlm.nih.gov/pubmed/35987823.

Table 1. Description of the standardized functional tasks

Task	Description
Reach-to-grasp (to pick up the receiver of a landline telephone placed on a table at which participant was sitting).	Participants sat with the proximal third of their thighs supported on the end of a plinth with height equal to the length of an individual participant's shank (head of the fibula to plantar surface of heel). The position allowed for hips, knees, and ankles to be at 90 degrees. Plinth height was recorded and replicated for a participant's second session. Starting heel positions were marked on the floor with a horizontal line to minimize inter-trial variability. A table was placed in front with height set so that participants could sit erect with shoulders relaxed, elbows at 90 degrees and hands on the table with wrist creases aligned to front edge. This starting position was marked on the table. A researcher then marked three positions on the table for placement of the landline telephone. Position one was in a participant's midline at 1.5 times of the forearm's length. The second and third positions were the same distance from the front edge of the table and 1.5 forearm's length to the right and left of position one respectively. Conditions were: dominant hand to midline position (RTG-MD); non-dominant hand to midline position (RTG-MND); dominant hand to contralateral position (RTG-CD); and non-dominant hand to contralateral position (RTG-CND). From the starting position, participants were instructed to "pick up the telephone to answer it when you hear the buzzer". The entire task was reach and pick up the telephone, bring to ear, put telephone back and then return hand to start position.
Sit-to-walk (to a landline telephone on a table 3ms in front of participant)	Participants sat on the plinth in the same position as for RTG. To minimize inter-trial variability: the feet positions on the floor were set by drawing around each one and tape was applied to the plinth to mark the seated position. Participants were asked to sit erect with head directly above their pelvis, arms at the sides with elbows flexed. A telephone was on a standard height table three meters in front of the edge of the plinth. Participants were instructed to go and answer the telephone when the buzzer sounded without using their hands to rise.

Table 2. Participant characteristics at testing session one

	Healthy adults (n = 75)
Age in years – mean (SD), min-max	47.8 (17.4), 20.5-85.4
Sex - male:female	29:46
Dominant side - right:left (number)	67:8
Body Mass Kg – mean (SD)	72.6 (14.1)
Height cm– mean (SD)	170.5 (9.5)

Abbreviations: SD = standard deviation; min = minimum; max = maximum.

Table 3. Mean, standard deviation (SD) and coefficient of variation (CV) for kinemetric metrics of smoothness and hesitation for quality trials derived from the Biokido and Vicon systems at sessions 1 and 2.

			RTG-MD smoothness	RTG-MND smoothness	RTG-CD smoothness	RTG-CND smoothness	STW hesitation	STW smoothness
Biokido system	Session 1	N participants	64	62	65	64	50	50
		$\bar{x}$ n trials	3.4 [1-5]	3.0 [1.5]	3.1 [1-5]	3.0 [1-5]	5.1 [1-10]	5.1 [1-10]
		mean (SD)	-4.92 (0.85)	-5.14 (0.89)	-4.84 (0.93)	-4.94 (1.01)	68.61 (20.67)	-2.36 (0.25)
		SD (SD)	0.54 (0.37)	0.60 (0.39)	0.47 (0.35)	0.50 (0.34)	14.40 (11.89)	0.15 (0.12)
		$\bar{x}$ CV [range]	0.11 [0.00, 0.29]	0.11 [0.00, 0.33]	0.09 [0.02, 0.31]	0.10 [0.02, 0.23]	0.22 [0.00, 0.97]	0.06 [0.01, 0.28,]
	Session 2	N participants	59	57	55	58	43	43
		$\bar{x}$ n trials	3.2 [1-5]	3.0 [1-5]	3.3 [1-5]	3.1 [1-5]	4.9 [1-10]	5.0 [1-10]
		mean (SD)	-4.53 (0.70)	-4.94 (0.99)	-4.43 (0.67)	-4.63 (0.84)	67.75 (21.34)	-2.33 (0.24)
		SD (SD)	0.49 (0.32)	0.54 (0.38)	0.52 (0.42)	0.42 (0.35)	11.01 (6.97)	0.12 (0.05)
		$\bar{x}$ CV [range]	0.11 [0.01, 0.33]	0.11 [0.01, 0.33]	0.11 [0.00, 0.29]	0.09 [0.00, 0.36]	0.18 [0.04, 0.48]	0.05 (0.01, 0.12)
Vicon system	Session 1	N participants	71	71	70	71	62	63
		$\bar{x}$ n trials	3.6 [2-5]	3.6 [1-5]	3.5 [1-5]	3.7 [3-5]	5.6 [2-10]	5.8 [2-10]
		mean (SD)	-4.63 (0.59)	-4.66 (0.66)	-4.47 (0.60)	-4.51 (0.68)	61.4 (15.63)	-2.33 (0.19)
		SD (SD)	0.37 (0.24)	0.37 (0.22)	0.27 (0.18)	0.31 (0.17)	8.74 (5.75)	0.09 (0.07)
		$\bar{x}$ CV [range]	0.08 [0.00, 0.22]	0.08 [0.01, 0.23]	0.06 [0.00, 0.17]	0.07 [0.01, 0.18]	0.14 [0.04, 0.47]	0.04 [0.00, 0.13]
	Session 2	N participants	71	69	71	70	61	61
		$\bar{x}$ n trials	3.7 [2-5]	3.7 [2-5]	3.7 [1-5]	3.7 [2-5]	5.9 [2-10]	6.0 [2-10]
		mean (SD)	-4.45 (0.56)	-4.50 (0.64)	-4.36 (0.58)	-4.44 (0.65)	57.49 (15.39)	-2.30 (0.17)
		SD (SD)	0.31 (0.21)	0.30 (0.17)	0.30 (0.18)	0.30 (0.19)	7.05 (3.14)	0.08 (0.05)
		$\bar{x}$ CV [range]	0.07 [0.00, 0.20]	0.07 [0.01, 0.17]	0.07 [0.00, 0.16]	0.07 [0.00, 0.17]	0.13 [0.02, 0.36]	0.04 [0.01, 0.14]

Abbreviations: RTG-MD = reach-to-grasp with dominant hand to midline placement of telephone; RTG-MND = reach-to-grasp with non-dominant hand to midline placement of telephone (RTG-MND); RTG-CD = reach-to-grasp with dominant hand to contralateral placement of telephone; RTG=CND = reach-to-grasp with non-dominant hand to contralateral placement of telephone; STW = sit-to-walk; n = number; $\bar{x}$ n trials = mean number of trials per individual; SD (SD) = standard deviation of within participant standard deviation; CV = average within participant coefficient of variation.

Table 4. Test-retest reliability and smallest detectable change for kinematic metrics derived from the Biokido and Vicon systems

	Biokido system					Vicon system				
	Number individuals (mean number measurements per individual)	ICC [95% CI]	SD	SEM	SDC	Number individuals (mean number measurements per individual)	ICC [95% CI]	SD	SEM	SDC
RTG-MD smoothness	71 (5.7)	0.40 [0.27,0.49]	1.07	0.83	2.29	74 (7.0)	0.44 [0.31,0.51]	0.76	0.62	1.71
RTG-MND smoothness	71 (5.0)	0.57 [0.44,0.65]	1.04	0.68	1.89	73 (7.0)	0.48 [0.34,0.55]	0.80	0.49	1.37
RTG-CD smoothness	70 (5.4)	0.48 [0.35,0.57]	1.06	0.76	2.12	73 (7.0)	0.56 [0.40,0.61]	0.55	0.50	1.40
RTG-CND smoothness	72 (5.1)	0.52 [0.37,0.59]	1.14	0.79	2.19	74 (7.0)	0.62 [0.45,0.65]	0.68	0.49	1.35
STW hesitation	55 (8.5)	0.47 [0.36,0.59]	31.14	22.67	62.83	67 (10.5)	0.57 [0.46,0.65]	15.79	14.86	41.20
STW smoothness	55 (8.5)	0.59 [0.49,0.69]	0.23	0.15	0.41	67 (10.9)	0.63 [0.53,0.71]	0.19	0.09	0.25

Abbreviations: RTG-MD = reach-to-grasp with dominant hand to midline placement of telephone; RTG-MND = reach-to-grasp with non-dominant hand to midline placement of telephone (RTG-MND); RTG-CD = reach-to-grasp with dominant hand to contralateral placement of telephone; RTG=CND = reach-to-grasp with non-dominant hand to contralateral placement of telephone; STW = sit-to-walk; ICC = intra-class correlation coefficient; 95% CI = 95% confidence intervals; SD = standard deviation; SEM = standard error measurement; SDC = smallest detectable change.

Table 5. Limits of agreement (LOA) for kinematic metrics derived from the Biokido and Vicon systems

	Biokido system		**Vicon system**	
	Number of matched pairs of observations	LOA	Number of matched pairs of observation	LOA
RTG-MD smoothness	154	-0.36 (-2.51 to 1.80)	244	-0.20 (-1.84 to 1.45)
RTG-MND smoothness	118	-0.27 (-2.33 to 1.79)	238	-0.23 (-1.95 to 1.50)
RTG-CD smoothness	138	-0.31 (-2.28 to 1.67)	238	-0.16 (-1.63 to 1.32)
RTG-CND smoothness	126	-0.20 (-2.71 to 2.32)	244	-0.15 (-1.67 to 1.37)
STW hesitation	163	2.13 (-54.82 to 59.09)	288	4.26 (-34.63 to 43.14)
STW smoothness	164	-0.05 (-0.69 to 0.59)	309	-0.03 (-0.41 to 0.35)

Abbreviations: RTG-MD = reach-to-grasp with dominant hand to midline placement of telephone; RTG-MND = reach-to-grasp with non-dominant hand to midline placement of telephone; RTG-CD = reach-to-grasp with dominant hand to contralateral placement of telephone; RTG=CND = reach-to-grasp with non-dominant hand to contralateral placement of telephone; STW = sit-to-walk; LOA = limits of agreement.

Table 6. Concurrent validity Vicon and Biokido systems:

intra-class coefficients plus 95% confidence intervals (ICC [95%CI]) and limits of agreement (LOA)

	Number of matched pairs (number of participants)	ICC [95% CI]	LOA
RTG-MD smoothness	372 (69)	0.52 [0.41,0.60]	0.25 (-1.31,1.82)
RTG-MND smoothness	326 (68)	0.39 [0.15,0.56]	0.57 (-1.19,2.32)
RTG-CD smoothness	346 (66)	0.41 [0.30,0.51]	0.28 (-1.32,1.88)
RTG-CND smoothness	342 (70)	0.56 [0.38,0.68]	0.39 (-1.15,1.92)
STW hesitation	409 (54)	0.52 [0.42,0.61]	-6.37 (-45.81,33.06)
STW smoothness	426 (54)	0.58 [0.52,0.65]	0.00 (-0.42,0.42)

Abbreviations: RTG-MD = reach-to-grasp with dominant hand to midline placement of telephone; RTG-MND = reach-to-grasp with non-dominant hand to midline placement of telephone (RTG-MND); RTG-CD = reach-to-grasp with dominant hand to contralateral placement of telephone; RTG=CND = reach-to-grasp with non-dominant hand to contralateral placement of telephone; STW = sit-to-walk.