Original Research

Accuracy and Reliability of Single-Camera Measurements of Ankle Clonus and Quadriceps Hyperreflexia

Check for updates

Keith Macon, MS [a], Dustin Hoang, BS [a],
Lauren Elizondo, MSPT [b,c], Kerri Kallus, DPT [c,d],
James Sulzer, PhD [a], Kathleen Manella, PT, PhD [c]

[a] Department of Mechanical Engineering, University of Texas at Austin, Austin, TX
[b] Spero Rehab, Austin, TX
[c] University of St. Augustine for Health Sciences, Austin, TX
[d] St. David's Rehabilitation Hospital, Austin, TX

Abstract *Objective:* To evaluate the accuracy and reliability of a simple, single-camera smartphone-based method, named the Reflex Tracker (RT) system, for measuring reflex threshold angles related to ankle clonus and quadriceps hyperreflexia.

Design: A prospective comparison study using a high-fidelity reference standard was constructed employing a $2 \times 2 \times 2$ factorial design, with factors of rater (tester) type (student and experienced physical therapist), joint (ankle and knee), and repetition (2 per condition).

Setting: This multicenter study was conducted at 4 outpatient rehabilitation clinics.

Participants: A convenience sample of 14 individuals with a neurologic condition presented with 20 lower limbs that exhibited ankle clonus and/or quadriceps hyperreflexia and were included in the study. Also participating in the study were 8 student and 8 experienced physical therapist raters (testers) (N=16).

Interventions: Not applicable.

Main Outcome Measures: The plantar flexor reflex threshold angle (PFRTA) related to ankle clonus and the quadriceps reflex threshold angle (QRTA) related to quadriceps hyperreflexia were quantified.

Results: PFRTA and QRTA results were compared between the smartphone RT method and synchronous 3-dimensional inertial measurement unit (IMU) sensor motion capture. Mean difference

(bias) was minimal between RT and IMU measurements for PFRTA (bias≤0.2°) and QRTA (bias≤1.2°). Intrarater reliability for PFRTA ranged from 0.85-0.90 using RT and from 0.85-0.87 using IMU; QRTA ranged from 0.97-0.98 using RT and from 0.96-0.99 using IMU. Intersensor reliability for PFRTA and QRTA was 0.97 and 0.99, respectively. Minimum detectable change for PFRTA ranged from 7.1° - 8.7° and for QRTA ranged from 6.1°-8.3°.

Conclusions: RT performed comparable to IMU for accurate and reliable measurement of PFRTA and QRTA to quantify ankle clonus and quadriceps hyperreflexia in clinical settings.

Abnormal reflex activity such as hyperreflexia is common after upper motor neuron injuries.[1] Ankle clonus and quadriceps spasm, forms of hyperreflexia, affect independence and quality of life.[2,3] Precise, reproducible measures of plantar flexor and quadriceps hyperreflexia are necessary to evaluate the effectiveness of interventions directed at normalizing reflex excitability.[4,5] The Modified Tardieu Scale uses kinematic measurements of "catch angles" or reflex threshold angles (RTAs) that predict functional impairment.[6] However, this scale has exhibited poor reliability for testing plantar flexor and quadriceps spasticity.[6,7] The RTA is the joint angle at reflex onset. The ankle clonus drop test elicits the plantar flexor RTA (PFRTA), and the quadriceps pendulum test elicits the quadriceps RTA (QRTA). For the ankle, a larger angle at reflex onset (less dorsiflexion) is associated with presence of clonus.[8] For the knee, a smaller angle at reflex onset (less knee flexion) is associated with presence of quadriceps spasm.[9] The PFRTA and QRTA are valid and reliable quantifiers of spasticity.[8-10] However, PFRTA and QRTA measurements require high-fidelity motion capture and are thus limited to laboratory settings with expensive equipment.[8,10]

Joint kinematics can be measured accurately with a variety of existing technologies. Portable technology such as inertial motion capture systems are accurate[11-13] but require calibration procedures that may be difficult for people with disabilities[14,15] and can be cost prohibitive. Time-of-flight sensing, such as the Microsoft Kinect, lacks accuracy and reliability and undersamples human motion.[16,17] To enable the routine use of the ankle clonus drop test and quadriceps pendulum test in clinical settings, a new motion capture system that is accurate, affordable, and easy for clinicians to use is required.

The purpose of our study was to develop a clinically accessible, accurate, and robust sensing system for quantifying hyperreflexia to use as a tool to evaluate treatment efficacy. Our goal was to use a smartphone camera and custom tracking and signal processing application to measure RTAs associated with the ankle clonus drop test and the quadriceps pendulum test in individuals with hyperreflexia. We evaluated smartphone video capture RTA measures with high-fidelity 3-dimensional inertial motion capture measures and assessed reliability of PFRTA and QRTA outcome measures performed by student and experienced physical therapists.

Methods

Study design

A prospective comparison study using a high-fidelity reference standard was constructed employing a 2 × 2 × 2 factorial design, with factors of rater (tester) type (student and experienced physical therapist), joint (ankle and knee), and repetition (2 per condition).

Setting, participants, and raters

Fourteen individuals with ankle clonus or quadriceps hyperreflexia in 1 or both legs were recruited by a sample of convenience from 4 outpatient rehabilitation clinics in Austin, Texas, from April-December 2019. Participant demographic characteristics are presented in table 1. Eligibility criteria included individuals ≥10 years of age with chronic upper motor neuron disorder and clinical signs of ankle clonus and/or quadriceps hyperreflexia who tolerated sitting and supine positions. Individuals with active medical conditions or undergoing Botox or baclofen treatment were excluded. For participants with bilateral clinical signs, the PFRTA and QRTA tests were performed on both legs, resulting in testing of 20 impaired legs. Also participating in the study were 8 student and 8 experienced physical therapist raters who performed the tests. The study was approved by the University of St. Augustine for Health Sciences Institutional Review Board; all participants gave written informed consent.

Test procedures

Participants were tested at 1 of 4 local outpatient physical therapy clinics in Austin, Texas. A portable commercial wireless inertial measurement unit (IMU) system, MVN BIOMECH,[a] provided a high-fidelity reference standard to compare with our smartphone single-camera Reflex Tracker (RT) system index test measures. The IMU MVN BIOMECH system demonstrates differences of less than 1° root mean square (RMS) difference for knee and ankle joint sagittal plane angles during walking and stair climbing compared with optical motion capture.[11] Another study verified this finding with less than 1° RMS difference and approximately 1% difference in knee flexion velocity during kicking.[12]

Standard clinical equipment was used during testing and included an adjustable height bench, a step platform, a Dycem nonslip sheet,[b] and pillows for participant comfort. The test area required space for positioning of the camera phone 1 m from the participant facing perpendicular to their sagittal plane. Seven IMUs were applied with Velcro bands to pelvis, thighs, shanks, and feet following Xsens recommended procedures.[18] The only exception was the upper leg

Table 1 Participant characteristics.

Participant	Leg Tested	Leg No.	Sex	Age (y)	Diagnosis	Months From Onset	Functional Status	Assistive Device	Orthotic
1	Left	001	Male	36	Stroke	32	Ambulatory	LBQC	AFO
2	Right	002	Male	44	Stroke	109	Ambulatory	SPC	AFO
3	Left	003	Female	55	Stroke	113	Ambulatory	SPC	AFO
4	Left	004	Male	37	SCI	96	Nonambulatory	WC	None
4	Right	005							
5	Right	006	Male	54	Stroke	46	Ambulatory	None	AFO
6	Right	007	Female	45	TM	25	Nonambulatory	WC	None
6	Left	008							
7	Left	009	Male	48	Stroke	26	Ambulatory	None	None
8	Right	010	Male	41	Stroke	22	Ambulatory	None	None
9	Left	011	Female	33	Stroke	23	Ambulatory	None	None
10	Right	012	Male	27	SCI	81	Ambulatory	None	AFO
11	Right	013	Female	36	SCI	129	Nonambulatory	WC	AFO
11	Left	014							
12	Right	015	Female	36	MS	300	Ambulatory	SPC	AFO
12	Left	016							
13	Left	017	Male	37	MS	52	Nonambulatory	WC	None
13	Right	018							
14	Right	019	Female	48	MS	9	Nonambulatory	Walker	None
14	Left	020							

Abbreviations: AFO, ankle-foot orthosis; LBQC, large base quad cane, MS, multiple sclerosis; SCI, spinal cord injury; SPC, single point cane; TM, transverse myelitis; WC, wheelchair.

IMU, which was placed on the middle of the anterior thigh to avoid accidental contact by tester hands or the exam table. The IMUs were calibrated prior to testing each leg using the N-pose option in Xsens MVN Studio software.[c] Following the Xsens calibration protocol,[19] the participant stood in a static N-pose for 10 seconds before returning to a seated or supine position to begin testing. For participants with walking ability, the N-pose was followed by a brief walk to further improve IMU calibration accuracy. For the camera-based motion capture, neon green adhesive stickers[d] were applied to the test leg at the greater trochanter, lateral knee joint line (proximal to fibular head), lateral malleolus, lateral posterior heel, and lateral fifth metatarsal head (fig 1A, B).

The ankle clonus drop test elicits a reproducible clonic response at the ankle.[8,10,20] Performed in sitting, the test leg is lifted until the foot is 2 in above a platform and released allowing the ball (metatarsal heads) of the foot to strike the platform edge. To conduct the quadriceps pendulum test, the individual, positioned on an adjustable height table in supine with lower legs dangling, is elevated such that the foot will not contact the ground; the tester grasps the foot and raises it moving the knee into full extension before releasing. For each test, the test response is recorded for 15 seconds and then the leg is placed in a rest position. To avoid fatigue and test repetition effects, a 1-minute rest period occurs between each repetition (see appendices 1 and 2 for detailed test procedures provided to all testers). The RTA is extracted as the outcome measure from each test.

We used a 2 × 2 × 2 factorial design, with factors of rater (tester) type (student and experienced physical therapist), joint (knee and ankle), and repetition (2 per condition). Each leg was tested a total of 8 times: 4 ankle clonus drop tests and 4 quadriceps pendulum tests. The order of tests was pseudorandomized with ankle tests first for half of the sessions. For participants with 2 legs tested, the IMUs were recalibrated before testing the second leg.

We used iPhone 6[e] and iPhone 10[e] hardware and native camera apps to record test videos at 60 frames per second. Each video was processed using custom Python software,[f] yielding 60 Hz measurements of ankle or knee joint flexion in the sagittal plane. The reference IMUs recorded 3-dimensional joint angles of the pelvis, both hips, knees, and feet at a sampling rate of 100 Hz. Calculations of expected aliasing errors[21] indicate a maximum error of 1.6° at 60 Hz and 0.6° at 100 Hz for the higher velocity ankle drop test (appendix 3).[21,22]

The RT system converts each frame from RGB to the CIE-LAB color space and then uses the channel and spatial reliability discriminative correlation filter tracking algorithm[23] to obtain the pixel x-y (sagittal plane) coordinates of each green adhesive sticker in every frame of the test video. We developed a graphical user interface that allows the user to select each marker from hip to toe and initialize the tracker's regions of interest. The lighting conditions at each test site varied, affecting marker clarity (fig 1D, E). We implemented an adjustable masking function to filter desired colors and improve tracking performance (fig 1C). The ankle and knee joint angle data were extracted from the marker x-y coordinate data using the law of cosines. We used the SciPy peak-finding function find_peaks[24] to identify the important oscillations in the signal, with a heuristically determined threshold of 120°/s to reliably identify the RTA. The number of oscillations were then counted as the detected peaks that followed the RTA.
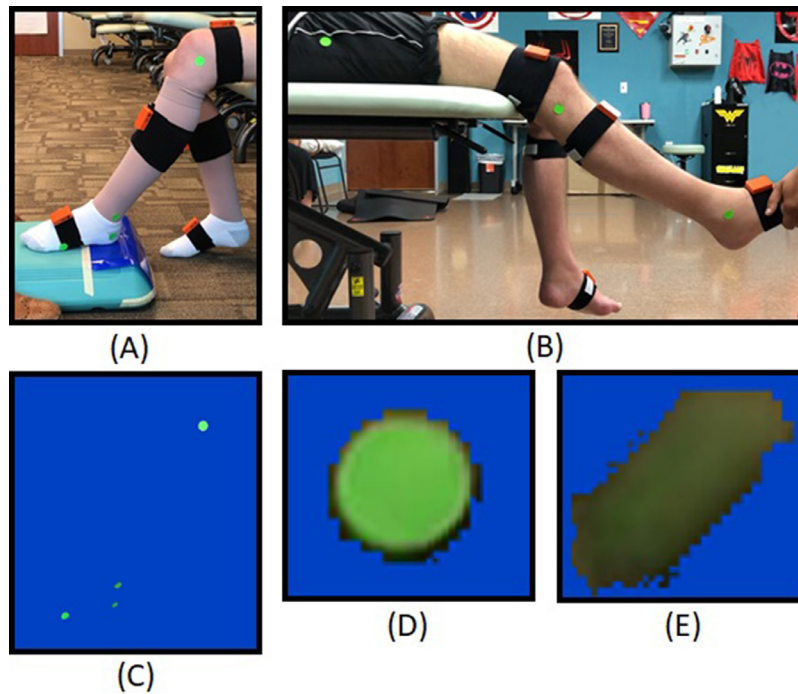
Fig 1 Experimental setups and single-camera perspective. (A) Ankle clonus drop test setup, (B) quadriceps pendulum test setup, (C) ankle drop test setup after applying color masking, (D) stationary marker, and (E) marker experiencing motion blur.

Data acquired

Participant data were acquired during 1 test session consisting of 4 drop test trials and 4 pendulum test trials with a student rater performing the first 2 test trials and an experienced rater performing test trials 3 and 4. Neither RT or IMU test results were available to the participants or raters. Some data were excluded from analysis because of various factors including rater testing error, participant time constraints, and magnetic interference identified by MVN Studio software (fig 2).

Statistical analysis

Bland-Altman plots and 95% limits of agreement (LoAs)[25] were used to examine sensor agreement and bias of PFRTA and QRTA measurements using the first repetition of each test. All Bland-Altman analyses, including confidence interval calculations for bias and LoA, were performed using the *BlandAltmanLeh* package in R. To our knowledge, there are no empirically derived acceptable LoA for PFRTA and QRTA. To estimate acceptable LoAs, we used clinical data from previous studies[8,26] and applied a conservative analysis, Fisher's
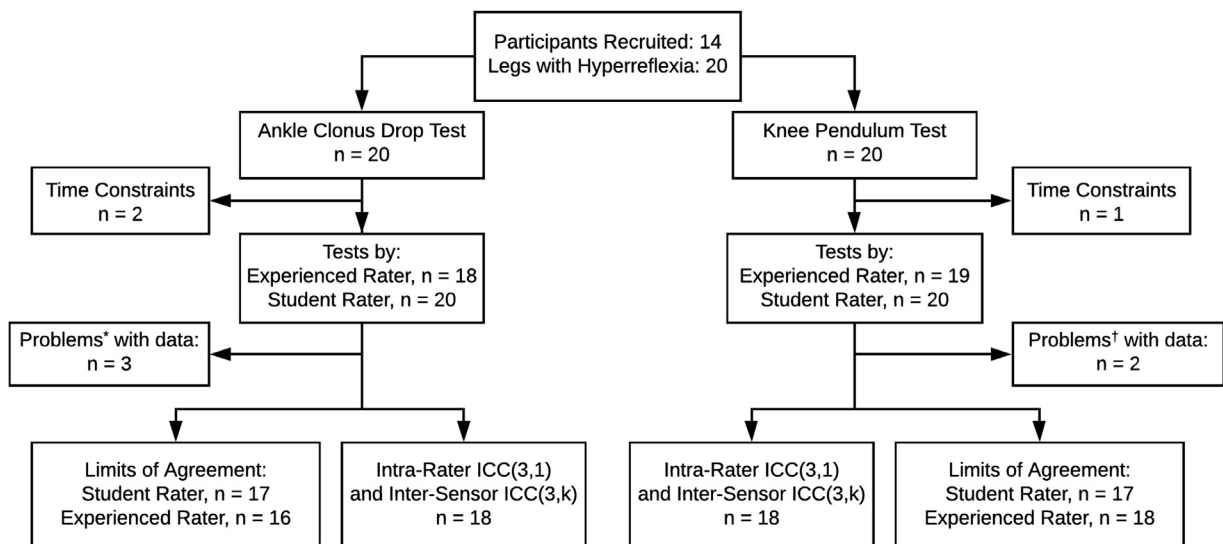

Fig 2 Participant testing and analysis flow diagram. *Problems identified during analysis included the tester contacting the shank IMU during the test, a knee marker coming loose, and magnetic interference. †Problems later identified in knee pendulum test data included a misrecording and magnetic interference.
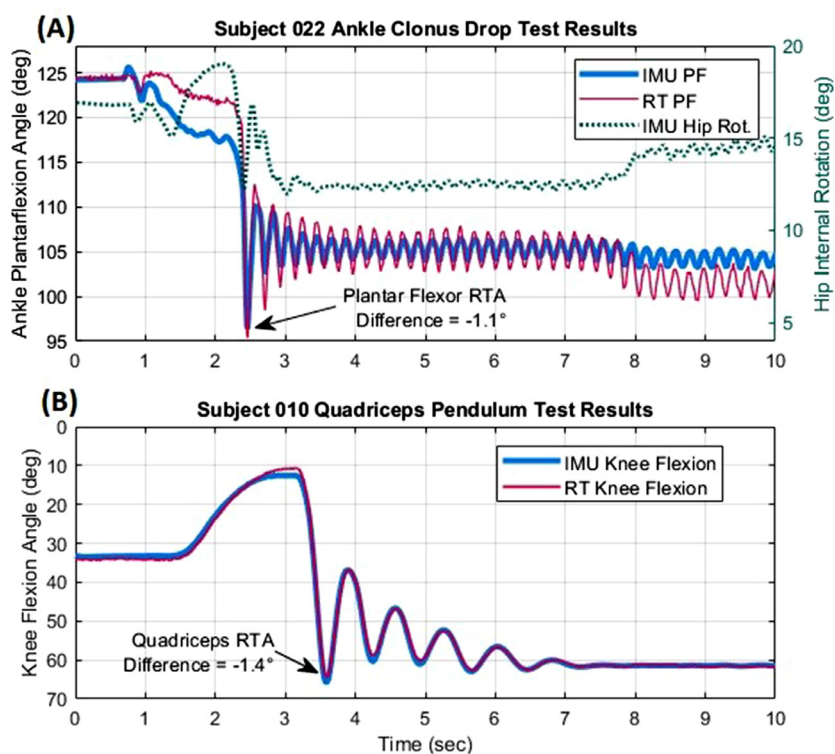
Fig 3 Representative results in 1 participant of the (A) ankle clonus drop test and (B) quadriceps pendulum test measured by the RT system and IMUs. Differences in PFRTA and QRTA measurements by each system are computed relative to baseline values. Hip internal rotation measured by IMUs illustrates 1 source of differences between RT and IMU plantar flexion angle.

least significant difference (LSD),[27] to compute the LSD for PFRTA and QRTA. We then defined acceptable LoAs for PFRTA and QRTA as smaller than the LSD of mean RTA between the impairment level groups in the previous studies.

Reliability of the RT and IMU methods was evaluated and compared using intraclass correlation coefficients (ICCs) and minimum detectable change (MDC$_{95}$).[28] We used ICC(3, 1) to describe the intrarater reliability of both the RT and IMU measurements for each rater type conducting each test.[28] The first and second repetitions performed by each rater were used to calculate each sensor's intrarater reliability using ICC(3, 1). We also used ICC(3, k) to describe the inter-sensor reliability between RT and the IMUs.[28] ICC interpretation was defined as excellent≥0.90, good≥0.75, and moderate to poor<0.75.[29] The MDC$_{95}$[28] was compared with the pre-post changes in group mean values for PFRTA[8] and QRTA[30] reported in previous studies that sought to evaluate treatment efficacy.

To evaluate effects of rater and repetition, we employed a linear mixed effects model based on the factorial experimental design. Each model's random effects included the subject, the subject-rater interaction, and the subject-repetition interaction. The R^g package *afex* was used to construct the model and perform analyses ($\alpha < .05$). This process was repeated with RTA magnitude as a covariate to check for the presence of proportional bias and again with nonflexion joint angles measured using the IMUs as covariates to quantify their effect on sensor disagreement. Each hypothesized model was compared using the R package *stats anova()* function to determine which model provided the best parsimonious fit of the data.

Results

Graphic representation of ankle clonus and quadriceps test response

Representative time series of aligned RT and IMU data are illustrated for Ankle Clonus Drop Test (fig 3A) and Quadriceps Pendulum Test (fig 3B). PFRTA and QRTA differences between the 2 methods were $-1.1°$ and $-1.4°$, respectively. However, in figure 3A, there is a difference in steady-state error after 8 seconds, with the RT ankle angle decreasing below the IMU data. This sudden increase in sensor disagreement can be attributed to a change in hip internal rotation at the 8-second mark affecting the IMU data.

Bland-Altman limits of agreement

Four Bland-Altman plots for PFRTA and QRTA measured by each rater type and each sensor type are illustrated in figure 4. Based on data from earlier studies by Manella and Field-Fote[8] and Fowler et al,[26] for each impairment level group in each study, we calculated the LSD for PFRTA and QRTA of $8.10°$ and $16.4°$, respectively, which defined our acceptable LoA for RT measurements. For the ankle clonus drop test PFRTA, the student rater tests had an RT-IMU bias of $0.18° \pm 1.22°$ (mean±95% confidence interval [CI]) and LoA from $-4.48° \pm 2.12°$ to $4.83° \pm 2.12°$. The experienced rater tests had a bias of $-0.11° \pm 2.04°$ and LoA from $-7.60° \pm 3.53°$ to $7.37° \pm 3.53°$. A 2-tailed F-test showed no significant difference between the student and experienced
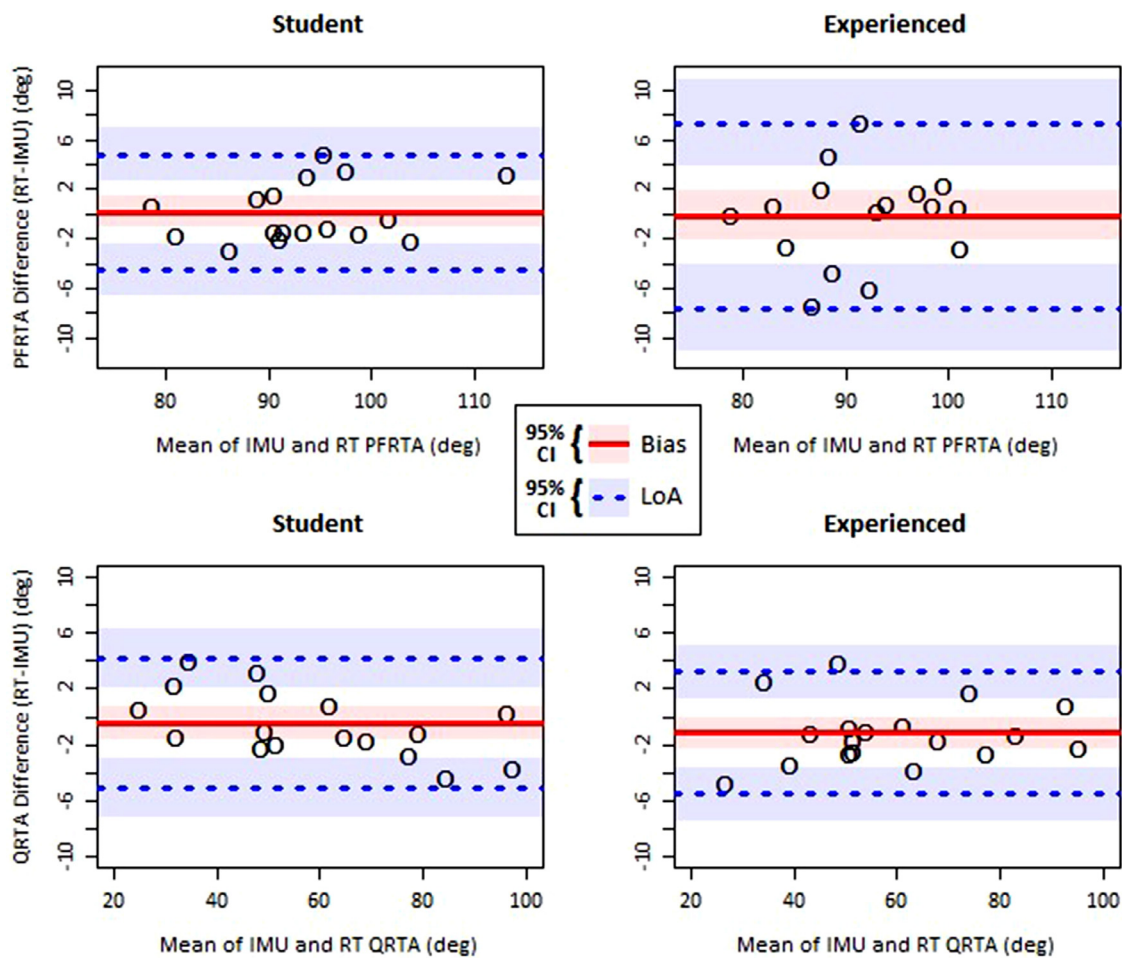
Fig 4   Bland-Altman plots of RTA measurements calculated with the RT system and IMUs. The shaded regions represent separate 95% CIs of the bias and 95% CIs of the LoA.

rater LoAs (P=.084). For the quadriceps pendulum test QRTA, the student rater tests had a bias of $-0.47°\pm1.22°$ and LoA from $-5.14°\pm2.12°$ to $4.20°\pm2.12°$. The experienced rater tests had a bias of $-1.15°\pm1.12°$ and LoA from $-5.54°\pm 1.93°$ to $3.25°\pm1.93°$.

ICC for reliability of PFRTA and QRTA

The intrarater reliability ICCs for each test, each sensor, and each rater type are presented in tables 2A and 2B along with the MDC$_{95}$. The intersensor reliability ICC(3, k) obtained by comparing the average measurements of RT and the IMUs was found to be 0.969 (95% CI, 0.931-0.986) for PFRTA and 0.998 (95% CI, 0.995-0.999) for QRTA. These ICCs reflect the proportion of between-subject and between-sensor variances.

Linear mixed effects model of sources of PFRTA and QRTA disagreement

For the ankle drop test PFRTA disagreement model, we found no significant effects of rater type (β=0.65, SE=0.78, P=.41) or repetition (β=1.02, SE=0.55, P=.073); however, their interaction was significant (β=−1.91, SE=0.74, P=.02). Ankle inversion, treated as a covariate, demonstrated a

trend toward significance (β=−0.066, SE=0.034, P=.058) that likely improved the model's description of the data as indicated by chi-square test (P=.068). Ankle abduction was not a significant covariate and did not improve the model (table 2C).

For the model of quadriceps pendulum test QRTA disagreement, we found no significant effects of rater type (β=0.61, SE=0.40, P=.13), repetition (β=0.59, SE=0.39, P=.14), or their interaction (β=−1.02, SE=0.56, P=.074). The intercept was found to be significant (β=−1.15, SE=0.52, P=.036). Modifying the model to include knee flexion magnitude, defined as the average of RT and IMU measured QRTA, showed no significance and did not improve the model. The final QRTA disagreement model included fixed effects of rater type, repetition, and their interaction effect and is summarized in table 2D.

Discussion

The bias of RT-IMU disagreement in PFRTA and QRTA was within 2° in tests conducted by both student and experienced raters. The LoAs between the 2 rater classes did not significantly differ, and both the student and experienced rater LoAs are within our defined acceptable range based on

Table 2 (A, B) Intrarater reliability of PFRTA and QRTA measurements and (C, D) fixed effects contributing to RT vs IMU disagreement

A. Ankle Clonus Drop Test PFRTA Reliability

Rater Type	Sensor	ICC (95% CI)	MDC$_{95}$ degree (95% CI)
Experienced	IMU	0.85 (0.70-0.93)	7.52 (5.38-9.67)
Experienced	RT	0.90 (0.79-0.95)	7.10 (4.33-9.87)
Student	IMU	0.87 (0.73-0.94)	7.75 (5.21-10.30)
Student	RT	0.85 (0.69-0.93)	8.70 (5.37-12.03)

B. Quadriceps Pendulum Test QRTA Reliability

Rater Type	Sensor	ICC (95% CI)	MDC$_{95}$ degree (95% CI)
Experienced	IMU	0.96 (0.92-0.98)	6.10 (4.11-8.08)
Experienced	RT	0.97 (0.93-0.99)	7.65 (4.82-10.46)
Student	IMU	0.99 (0.97-0.99)	8.40 (5.82-10.97)
Student	RT	0.98 (0.96-0.99)	8.27 (5.73-10.81)

C. Linear Mixed Effects Random Model of PFRTA Disagreement

Fixed Effect	Estimate	SE	DoF	$P(>|t|)$
Intercept	−0.039	0.693	36.0	.956
Rater type (student)	0.654	0.780	28.0	.409
Repetition (rep 2)	1.018	0.547	29.5	.073
Ankle inversion	−0.066	0.034	44.5	.058
Rater type*Repetition	−1.908	0.740	17.2	.020

D. Linear Mixed Effects Random Model of QRTA Disagreement

Fixed Effect	Estimate	SE	DoF	$P(>|t|)$
Intercept	−1.146	0.520	26.7	.036
Rater type (student)	0.613	0.398	50.1	.130
Repetition (rep 2)	0.589	0.391	50.0	.138
Rater type*Repetition	−1.020	0.558	50.1	.074

Abbreviation: DoF, degrees of freedom.

the between-severity group LSD calculated using data reported in earlier work studying the ankle[8] and knee.[26] These findings suggest that RT can assess patients' baseline PFRTA and QRTA with accuracy similar to that of inertial motion capture.

Intrarater reliability of the pendulum test was shown to be excellent for both rater types measured by both sensors, and the intrarater reliability of the ankle clonus drop test ranged from good to excellent (table 2A, B). The higher pendulum test ICCs highlight the presence of intrinsic differences between the ankle clonus drop test and the quadriceps pendulum test. The intersensor reliability when comparing average measurements of the PFRTA and QRTA was excellent (table 2A, B). This suggests that the between-subject RTA variance was substantially higher than the between-sensor variance. Because between-sensor variance is also a key component in finding the LoA, this result demonstrates that RT and IMU have good reliability as well as good agreement relative to the differences between subjects. Our findings concur with Banky et al,[31] who reported good to excellent reliability for smartphone video assessment of soleus

(ICC=0.94) and quadriceps (ICC=0.88) for fast velocity start angles using the Modified Tardieu Scale.[32]

The MDC$_{95}$ for both the PFRTA and QRTA ranged between approximately 6° and 9° (table 2A, B). Manella and Field-Fote[8] reported that in individuals with spinal cord injury who reduced their PFRTA the average change after training was 10.7°. Our findings suggest that RT is capable of reliably identifying typical improvements in patients who respond to treatment and that regardless of rater, there is no significant difference between the IMU and RT systems. Ness and Field-Fote[30] evaluated changes in the QRTA in individuals with chronic spinal cord injury and found an average change of 12.05° after 4 weeks of intervention.

We did not observe a significant effect of rater experience or repetition, suggesting that the RT is repeatable and usable by novices. Although the order of ankle and knee drop tests was randomized across participants, the order of rater type was constant with students always conducting tests before experienced raters. As a result, the interaction effect between rater type and repetition could instead be an effect of testing order.

Study limitations

Although the N-pose and walk calibration is recommended by Xsens for best performance,[18,19] not all participants were able to walk. Use of the N-pose without walk calibration may have resulted in slightly higher RT-IMU disagreement for measurements with nonambulatory participants. The MDC_{95} calculated for RT and IMU characterizes the reliability of same-day repeated measurements. Although some treatments can result in same-day improvement in RTAs,[30] the long-term efficacy of interventions is also of interest. Future work may include test-retest reliability of RT and analysis of RT sensitivity to oscillations during the reflex response. We used inertial motion capture as the basis of comparison for the RT. Although IMUs may approach the accuracy and sensitivity of the criterion standard optical motion capture,[11-13] any system will have some joint kinematic error. Further, whereas some studies found <1° RMS joint angle difference in knee and ankle flexion-extension angles,[11,12] another study using an older version of the system found greater differences at 3°-6°.[33] Because we cannot confirm the exact nature of the error, we can only discuss the agreement between measurement systems, not the true error. During analysis of sources of error, we could not quantify the effects of magnetic distortion or calibration error of the IMUs.

Conclusions

We developed a novel system known as the Reflex Tracker to measure reflex threshold angles unobtrusively and inexpensively in a clinical environment. We found that RT performed with clinically acceptable accuracy and repeatability. We conclude that RT can be used as a tool to effectively measure changes in an individual's PFRTA and QRTA in a clinical environment. Because RT requires only a smartphone camera and simple stickers, it provides both novice and experienced clinicians with easy access to objective measurements of plantar flexor and quadriceps hyperreflexia.

Suppliers

a. MVN BIOMECH; Xsens.
b. Dycem nonslip sheet; Dycem.
c. MVN Studio Software; Xsens.
d. Green adhesive stickers; Avery.
e. iPhone 6 and iPhone 10; Apple Inc.
f. Python software; Python Software Foundation.
g. R; R Foundation for Statistical Computing

Corresponding author

Kathleen Manella, PT, PhD, Doctor of Physical Therapy Program, University of St. Augustine for Health Sciences, 5402 La Crosse Ave, Austin, TX 78748. *E-mail address:* kmanella@usa.edu.

References

1. Gracies J. Pathophysiology of spastic paresis. II: emergence of muscle overactivity. Muscle Nerve 2005;31:552-71.
2. Fee J, Miller F. The Leg Drop Pendulum Test performed under general anesthesia in spastic cerebral palsy. Dev Med Child Neurol 2004;46:273-81.
3. Mayo M, DeForest B, Castellanos M, Thomas C. Characterization of involuntary contractions after spinal cord injury reveals associations between physiological and self-reported measures of spasticity. Front Integr Neurosci 2017;11:2.
4. Adams M, Hicks A. Spasticity after spinal cord injury. Spinal Cord 2005;43:577-86.
5. Patrick E, Ada L. The Tardieu Scale differentiates contracture from spasticity whereas the Ashworth Scale is confounded by it. Clin Rehabil 2016;20:173-82.
6. Mehrholz J, Wagner K, Meibner D, et al. Reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in adults with severe brain injury: a comparison study. Clin Rehabil 2016;19:751-9.
7. Yam WK, Leung MS. Interrater reliability of Modified Ashworth Scale and Modified Tardieu Scale in children with spastic cerebral palsy. J Child Neurol 2006;21:1031-5.
8. Manella K, Field-Fote E. Modulatory effects of locomotor training on extensor spasticity in individuals with motor-incomplete spinal cord injury. Restor Neurol Neurosci 2013;31:633-46.
9. Bohannon R, Harrison S, Kinsella-Shaw J. Reliability and validity of pendulum test measures of spasticity obtained with the Polhemus tracking system from patients with chronic stroke. J Neuroeng Rehabil 2009;6:30.
10. Manella K, Roach K, Field-Fote E. Temporal indices of ankle clonus and relationship to electrophysiologic and clinical measures in persons with spinal cord injury. J Neurol Phys Ther 2017;41:229-38.
11. Blair S, Duthie G, Robertson S, Hopkins W, Ball K. Concurrent validation of an inertial measurement system to quantify kicking biomechanics in four football codes. J Biomech 2018;73:24-32.
12. Zhang JT, Novak AC, Brouwer B, Li Q. Concurrent validation of Xsens MVN measurement of lower limb joint angular kinematics. Physiol Meas 2013;34:N63.
13. Lee J, Shin SY, Ghorpade G, Akbas T, Sulzer J. Sensitivity comparison of inertial to optical motion capture during gait: implications for tracking recovery. In: Paper presented at: IEEE 16th International Conference on Rehabilitation Robotics (ICORR); June 24-28, 2019; Toronto, Canada.
14. Liu Y, Zhang Y, Zeng M. Sensor to segment calibration for magnetic and inertial sensor based motion capture systems. Measurement 2019;142:1-9.
15. Picerno P, Caliandro P, Iacovelli C, et al. Upper limb joint kinematics using wearable magnetic and inertial measurement units: an anatomical calibration procedure based on bony landmark identification. Sci Rep 2019;9:14449.
16. Bonnechère B, Jansen B, Salvia P, et al. Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry. Gait Posture 2014;39:593-8.
17. Guess T, Razu S, Jahandar A, Skubic M, Huo Z. Comparison of 3D joint angles measured with the Kinect 2.0 Skeletal Tracker versus a marker-based motion capture system. J Appl Biomech 2017;33:176-81.
18. BASE by Xsens. Sensor Placement in Xsens Awinda System. Available at: https://base.xsens.com/knowledgebase/s/article/Sensor-Placement-in-Xsens-Awinda-System-1605791447430. Accessed June 29, 2021.

19. Xsens Technologies B.V. MVN user manual: user guide. Available at: https://fccid.io/QILMTW2-3A7G6/User-Manual/Users-Manual-2695756. Accessed June 21, 2021.

20. Manella KJ, Roach KE, Field-Fote EC. Operant conditioning to increase ankle control or decrease reflex excitability improves reflex modulation and walking function in chronic spinal cord injury. J Neurophysiol 2013;109:2666-79.

21. Endevco. Shock measurements: an appropriate sampling rate. Available at: https://link.edgepilot.com/s/41a500ee/VzTHPx8G3kypGyovRgxCKg?u=https://web.archive.org/web/20201124183430/https://www.endevco.com/Our-Resources/Ask-the-Experts/expert-detail/shock-measurements-an-appropriate-sampling-rate. Accessed November 24, 2020.

22. Boyraz I, Uysal H, Koc B, Sarman H. Clonus: definition, mechanism, treatment. Medicinski Glasnik 2015;12:19-26.

23. Lukežič A, Vojíř T, Čehovin Zajc L, Matas J, Kristan M. Discriminative correlation filter tracker with channel and spatial reliability. Int J Comput Vis 2018;126:671-88.

24. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature 2020;585:357-62.

25. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nurs Stud 2010;47:931-6.

26. Fowler EG, Nwigwe AI, Ho TW. Sensitivity of the pendulum test for assessing spasticity in persons with cerebral palsy. Dev Med Child Neurol 2000;42:182-9.

27. Salkind NJ. Fisher's least significant difference test. In: Salkind N, ed. Encyclopedia of research design, Thousand Oaks: SAGE; 2010. p. 492-4.

28. Weir J. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005; 19:231-40.

29. Portney LG, Watkins MP. Statistical measures of reliability. In: Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2009. p. 591-5.

30. Ness L, Field-Fote E. Effect of whole-body vibration on quadriceps spasticity in individuals with spastic hypertonia due to spinal cord injury. Restor Neurol Neurosci 2009;27:621-33.

31. Banky M, Clark RA, Mentiplay BF, Olver JH, Kahn MB, Williams G. Toward accurate clinical spasticity assessment: validation of movement speed and joint angle assessments using smartphones and camera tracking. Arch Phys Med Rehabil 2019; 100:1482-91.

32. Ben-Shabat E, Palit M, Fini N, Brooks C, Winter A, Holland A. Intra- and interrater reliability of the Modified Tardieu Scale for the assessment of lower limb spasticity in adults with neurologic injuries. Arch Phys Med Rehabil 2013;94:2494-601.

33. Robert-Lachaine X, Mecheri H, Larue C, Plamondon A. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. Med Biol Eng Comput 2017;55:609-19.