Test–retest reliability for performance-based outcome measures among individuals with arthrogryposis multiplex congenita

Jaclyn Megan Sions[1]* , Maureen Donohoe[2], Emma Haldane Beisheim-Ryan[3,4], Ryan Todd Pohlig[5], Tracy Michele Shank[3] and Louise Reid Nichols[6]

Abstract

Background: Most individuals with arthrogryposis multiplex congenita, a rare condition characterized by joint contractures in ≥ 2 body regions, have foot and ankle involvement leading to compromised gait and balance. The purpose of this study was to establish between-days, test–retest reliability for performance-based outcome measures evaluating gait and balance, i.e., the 10-m Walk Test, Figure-of-8 Walk Test, 360-degree Turn Test, and modified Four Square Step Test, among adolescents and adults with arthrogryposis multiplex congenita.

Methods: This reliability study included ambulatory participants, aged 10 to 50 years, with a medical diagnosis of arthrogryposis multiplex congenita. Participants completed performance-based measures, in a randomized order, on two separate occasions. Intraclass correlation coefficients with 95% confidence intervals and minimal detectable changes at the 90% and 95% confidence level were calculated.

Results: Participants included 38 community-ambulators with a median of 13 out of 14 upper and lower joint regions affected. Intraclass correlation coefficient point estimates and 95% confidence intervals ranged from .85-.97 and .70-.98, respectively. Minimal detectable changes were 10 to 39% of sample means and were largest for the modified Four Square Step Test.

Conclusions: Among individuals with arthrogryposis, gait speed per the 10-m Walk Test, as well as non-linear walking and dynamic balance assessment per the Figure-of-8 Walk and 360 Degree Turn Tests, have adequate test–retest reliability enabling evaluation of individual patient changes. Changes in groups of ambulatory individuals with arthrogryposis multiplex congenita may be reliably evaluated with all of the studied outcome measures.

Keywords: Arthrogryposis, Clubfoot, Gait, Postural balance, Walking speed

Background

Arthrogryposis multiplex congenita (AMC) is a condition occurring in about 1 in 5,000–10,000 live births characterized by non-progressive joint contractures in 2 or more body regions [1–3]. Clinical presentations are variable, resulting in considerable heterogeneity of individuals with AMC in terms of mobility status, cognitive status, and limitations with activities-of-daily living [1, 4, 5]. Early clinical courses include extensive serial casting, orthotics management, rehabilitation, and often, orthopedic surgeries [6–9]. Care of the lower-extremities focuses on optimizing independence through maximizing ambulatory potential. After optimization, there is a need to monitor for functional degradation, particularly as children transition from adolescence into early adulthood, when they may no longer have access

*Correspondence: megsions@udel.edu
[1] Department of Physical Therapy, University of Delaware, 540 South College Ave., Suite 210JJ, Newark, Delaware 19713, USA
Full list of author information is available at the end of the article

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 2 of 10

to AMC-provider specialists as they age-out of pediatric healthcare facilities [10]. Longitudinal monitoring requires quick and reliable outcome measures with known minimal detectable changes (MDCs) that allow assessment of whether 'true change' exceeding measurement error has occurred [11].

Performance-based outcome measures evaluate an individual's capacity under a given set of conditions, complementing patient-reported outcomes, which evaluate an individual's perceived ability [12]. While self-report measures are discriminative when evaluating if individuals are able or unable to perform a given task, performance-based measures may further stratify levels of physical functioning [12]. Assuming standardized testing procedures, an additional inherent benefit to timed performance-based measures is objectivity. Objectivity may be critical when justifying care, including surgical procedures, orthotics, assistive technology, and rehabilitation, particularly when there is a potential conflict of interest for the examiner and/or patient.

Whereas performance-based functional outcomes, such as the 10-m Walk Test (10mWT), are staples in adult rehabilitation, use in pediatric orthopedics is less common, especially among patients with congenital conditions, including AMC [7]. If non-condition-specific performance-based outcome measures used in other patient populations are established as psychometrically-sound among children and adults with AMC, practitioners may be able to easily track function from adolescence through adulthood, without the need for novel measures, overcoming a primary barrier to adoption of new outcome measures-lack of practitioner familiarity [13]. Further, use of outcome measures that are not condition-specific allows for comparison to normative data, as well as comparison across patient samples; such comparisons can assist with prioritization of patient subgroups most in need of limited resources due to greater risk for poor outcomes.

Among adolescents and young adults with AMC, lower-extremity involvement, especially of the foot and ankle, occurs in 80–90% of individuals [14]. Consequently, gait and balance impairments are ubiquitous. In community-dwelling older adults, reduced gait speed is predictive of adverse health outcomes, such as disability, institutionalization, and mortality [15]. Among pediatric populations, reduced gait speed is associated with worse disease severity and greater disability, [16, 17] and in some populations like Hereditary Motor Sensory Neuropathy, gait speed is a predictor of subsequent functional decline [18]. Given increased challenges to gait stability and symmetry with curved path walking, the use of straight-path walking alone is discouraged [19]. Thus, the purpose of this study was to establish between-days, test–retest reliability and minimal detectable changes for four outcome measures of gait and balance among adolescents and adults with AMC. The four measures were the 10mWT, which evaluates straight-path walking, and the Figure-of-8 Walk Test (F8WT), 360-degree Turn Test (360TT), and modified Four Square Step Test (mFSST), which evaluate non-linear walking.

Methods

Study design and participants

This test–retest reliability study recruited participants, aged 10 to 50 years, with a medical diagnosis of AMC who were ambulatory and able to walk at least 30 m with or without an assistive device. Recruitment occurred through verbal recruitment at community events and print advertisements from April to July of 2019. Exclusion criteria included cognitive-impairment precluding assent/consent, spine or lower-limb surgery in the past 6 months, a history of lower-limb amputation, current dizziness, an acute illness, or a progressive neuromuscular disease. The study was conceptualized in November of 2018, approved by the University of Delaware Institutional Review Board for Human Subjects Research (project number: 1354682; initial approval date: 1/16/2019) and conformed to the World Medical Association's Helsinki Declaration; written informed consent/assent and parental permission (as applicable based on age) was obtained for all participants. See Fig. 1 for the study timeline.

Data collection

Data collections occurred at a University of Delaware clinical research laboratory in Newark, Delaware, and at a Hilton hotel conference space in Norfolk, Virginia. After informed consent, trained research staff conducted standardized interviews for participant characterization; parental input was encouraged for medical history recall among adolescent participants. Participants independently completed the Gillette Functional Assessment Questionnaire to characterize their mobility status, where scores range from 1 = "unable to take steps" to 10 = "walks, runs, climbs on level and uneven terrain without difficulty or assistance" [20]. For further participant characterization, average pain intensity rating over the past 7 days was obtained with the Patient-Reported Outcomes Measurement Information System, where 0 = "no pain" and 10 = "worst imaginable pain" [21]. Participants then completed performance-based outcome measures in a randomized order, and returned 1–10 days later for repeat performance-based testing.
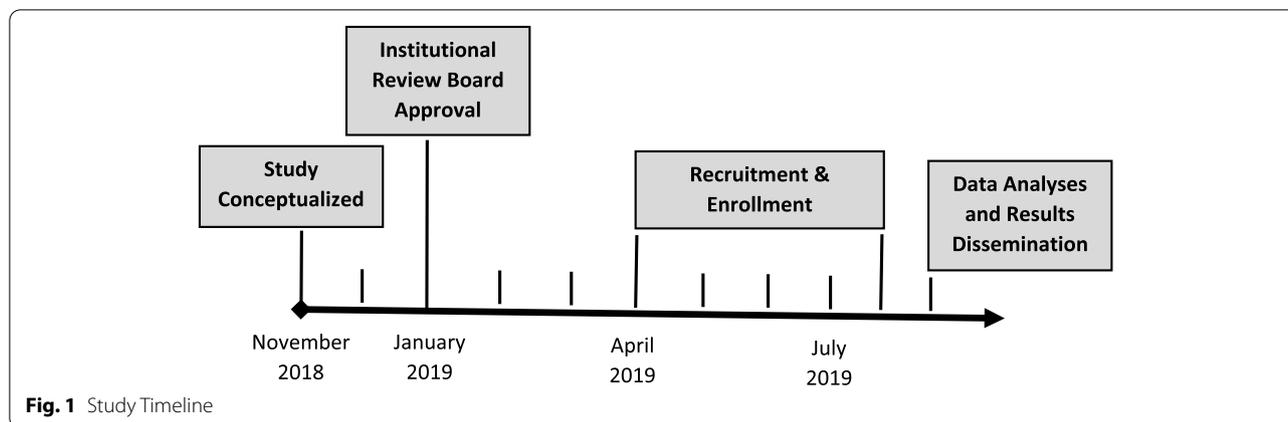
Fig. 1 Study Timeline

Performance-based outcome measures
10mWT
'Usual' gait speed was obtained over the middle 6 m of a 10-m course, allowing 2 m for acceleration and deceleration at either end (Fig. 2) [22]. For some children, adolescent, and adult populations, such as those with neurological conditions and hip dysplasia, between-days test–retest reliability for the 10mWT has been reported [23, 24]. The 10mWT, using ≥ 2 trials, is included in a core set of rehabilitation outcome measures recommended for adults with neurological conditions based on its established psychometric properties and clinical utility [22]. Gait speed was determined from a three-trial average based on prior test–retest reliability research [25].

F8WT
The F8WT, which times curved-path walking around two cones arranged in a figure-of-8 (Fig. 2), is a valid measure of walking skill, and provides complimentary information to gait speed [26]. F8WT can differentiate between adults with lower-limb pathology with varying functional mobility levels [27]. Among adults post-stroke and -total knee arthroplasty, test–retest reliability for time and number of steps, where a greater number of steps indicates poorer performance, has been reported [26, 28–30].

360TT
During community ambulation, up to 50% of steps may include turning, [31] and impaired turning has been associated with recurrent falls [32]. Rehabilitation balance batteries, including the Pediatric Balance Scale, Berg Balance Scale, and Tinetti Performance-Oriented Mobility Assessment, have a 360-degree turn task, but administration time required to complete such batteries may reduce clinical adoption [33]. Therefore, in this study, four 360-degree turns [two to the left; two to the right;

[34] Fig. 2] were completed and average time for completion and number of steps were determined. More steps to complete the turn indicates poorer performance. Test–retest reliability for 360TT, in isolation, has been reported in adult patient populations, such as those with Multiple Sclerosis, Parkinson's disease, and post-stroke [34–37].

mFSST
In various pediatric and adult populations, Four Square Step Test (FSST) test–retest reliability has been reported, as has concurrent, construct, and predictive validity for falls among adults [23, 38–40]. With the FSST, individuals complete multi-directional stepping (i.e., forwards, lateral, and backwards) over canes arranged in a '+' in a specified sequence [38]. Requirement of a specified sequence for a valid trial increases cognitive-load, but the requirement for foot clearance alongside correct sequencing can result in 'invalid' trials, contributing to the FSST's known floor effect [38]. Thus, the mFSST, which substitutes taped lines for canes, was used (Fig. 2) [27, 39]. Aligning with prior research, [39] reliability for 'average' and 'best' performance over two trials, was determined.

Data Analyses
IBM SPSS Statistics 26 (Armonk, NY, USA) was used for all analyses. Descriptive statistics were determined. Wilcoxon Signed Ranks Test was used to evaluate intra-individual differences in pain intensity between testing sessions ($p \leq 0.050$). Intraclass correlation coefficients (ICCs) were used to evaluate between-days, test–retest reliability for performance-based outcome measures using two-way mixed effects, absolute agreement ($ICC_{3,1}$ or $_{3,k}$ models) [41, 42]. Given the nonparametric distribution of performance-based data, Bland–Altman plots were evaluated to ensure ICC analyses were appropriate. ICCs > 0.90 may be considered excellent [41] and
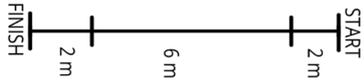
	Equipment: A clear linear pathway at least 10 meters long with lines on the floor indicating the start, finish, and 2-meter marks on either end; stopwatch; gait belt; assistive device (if needed).
	Instructions: "*Walk at your usual pace to the other end of this course. Walk all the way past the other end of the tape before you stop.*"
	[NOTE: Complete 3 trials. Start timer after 1st 2 meters and do not time last 2. 6 meters is divided by total time.]
	Scoring: Average gait speed (meters/second) from 3 trials.

Fig. 2 Standardized Procedures for Performance-Based Outcome Measures
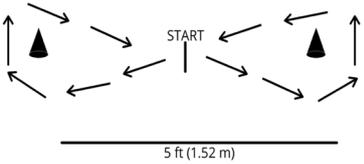
Equipment: 2 cones spaced 5 feet apart with a center mark; a stopwatch; gait belt; assistive device (if needed).

Demonstrate During Instructions: *Walk around these cones in a figure-8 pattern. Start in the middle of the 2 cones. Walk around one cone, then between the cones, and around the other cone, finishing where the walk started. Do this as quickly as you can while trying to complete the walk smoothly without stopping.*

Scoring: Average time (seconds); number of steps.
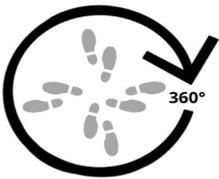
Equipment: stopwatch; gait belt.

Demonstrate During Instructions: *"Turn 360 degrees as quickly and safely as possible."*

[NOTE: Allow the participant to self-select the turn direction for 1st trial; cue for 2nd trial in the same direction. Then do the opposite direction for trials #3 and #4. Assistive device use is not allowed.]

Scoring: Average time (seconds); number of steps.
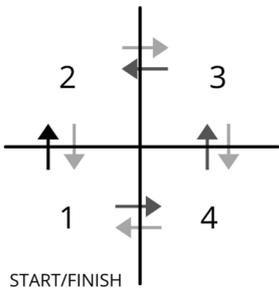
Equipment: Tape marking 4 squares; a stopwatch; gait belt; assistive device (if needed).

Demonstrate during Instructions: "*Try to complete the sequence as fast as possible without touching the lines. Both feet must make contact with the floor in each square. If possible, face forward during the sequence.*"

[NOTE: Sequence #1-2, #2-3, #3-4, #4-1, #1-4, #4-3, #3-2, #2-1. Allow 1 practice trial. During 2 timed trials, do not say sequence aloud as this may pace participant. If the participant touches the line or performs the sequence incorrectly, the trial is considered 'invalid'.]

Scoring: Average or best time (seconds).

desirable for evaluating individual patient changes, [43] while ICCs > 0.70 may be adequate for evaluating group changes [44, 45]. Standard errors of measurements (SEMs) were calculated. The decision to use MDC at the 90% (MDC_{90}) or 95% confidence level (MDC_{95}) may be determinant on the intervention. For example, meeting or exceeding MDC_{90} may be appropriate for evaluating success of conservative interventions, such as rehabilitation, but when evaluating success of surgical interventions with greater inherent risks, MDC_{95} may be more appropriate; [46] thus, both MDC_{90} and MDC_{95} were determined.

Results

Of the 63 interested individuals with AMC screened for study participation, 12 were ineligible due to an inability to walk at least 30 m. Fifty-one enrolled participants completed the first onsite data collection, which was part of a larger cross-sectional research project. Of these 51 participants, 38 were available and agreed to participate in an optional second onsite data collection within 1–10 days. Thus, 38 participants were included in this between-days, test–retest reliability study, exceeding the recommended sample size of at least 30 participants for reliability studies in rehabilitation [47].

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 5 of 10

The sample was largely female, Caucasian, and reported no known genetic cause for their AMC (Table 1). Nearly 50% of the sample had spinal involvement; the median number of affected upper-limb regions was 8 out of 8, and the median number of affected lower-limb regions was 5 out of 6. Home assistive device use was rare, but 29% of the sample reported community assistive device use. The median number of lower-limb orthopedic surgeries was 4, with spinal surgeries and upper-limb surgeries reported much less frequently. Most participants were ambulatory outside the home for community distances (i.e., GFAQ ≥ 8), and had mild-to-moderate pain over the course of last 7 days. Pain was not significantly different between testing sessions 1 and 2 ($p = 0.368$), with approximately two-thirds of the sample reporting 0/10 pain at the time of performance-based testing.

Bland–Altman plots for performance-based measures indicated ICC analyses were appropriate; however, differences between timepoints were significantly different from 0 for F8WT, average mFSST, and best mFSST scores ($t = 2.061$–3.783, $p = 0.001$-0.047). It was not unexpected for a practice effect to occur with repeated trials of novel tasks; while the absolute F8WT and mFSST scores improved between timepoints, individuals' relative standing did not change (all $r > 0.91$). Therefore, ICC analyses were deemed appropriate for evaluating reliability in this subset.

Table 2 presents between-days, test–retest reliability results, SEMs, and MDCs for the sample. ICC point estimates ($ICC_{3,k}$) surpassed 0.90 for average gait speed over 3 trials as evaluated with the 10mWT; average F8WT time and number of steps obtained from 2 trials; and average 360TT and mFSST times, obtained from 4 and 2 trials, respectively. ICC point estimates and 95%CIs were ≥ 0.70 for average number of steps for the 360TT ($ICC_{3,k} = 0.85$; 95%CI: 0.70-0.92) and best time on the FSST ($ICC_{3,1} = 0.87$; 95%CI: 0.71-0.94). MDC_{90} values were 10 to 33% of sample means, while MDC_{95} values were 13 to 39% of sample means, with the largest MDCs

Table 1 Participant Characteristics (*n* = 38)

Age, y[a]	20 (14, 33)
Sex, female	27 (71.1)
Ethnicity, Non-Hispanic/Latino	34 (89.5)
Race, white/Caucasian	31 (81.5)
Height, m[a]	1.56 (1.45, 1.63)
Weight, kg[a]×	53.1 (40.8, 62.4)
Known Genetic Cause of AMC	4 (10.5)
Spinal Involvement ×	18 (48.6)
Upper Limb Regions, 0-8[a]	8 (4, 8)
Lower Limb Regions, 0-6[a]	5 (3, 6)
ASSISTIVE TECHNOLOGY	
Lower Limb Orthotics Use	22 (57.9)
Home Assistive Device Use	1 (2.6)
Cane	1 (2.6)
Community Assistive Device Use	11 (28.9)
Cane	3 (7.9)
Loftstrand Crutches	2 (5.3)
Rollator Walker	1 (2.6)
Wheelchair	5 (13.2)
Spinal	6 (15.8)
Number of Upper Limb[a]↑	0 (0, 3)
Number of Lower Limb[a]↓	4 (2, 7)
GFAQ[a]	9 (8, 9)
7-day Average Pain Rating, 0-10[a]	3 (1, 4)

[a] Data presented as median (25th, 75th percentile) rather than n (% of sample)

× *n* = 1 participant with missing data

β *n* = 2 participants with missing data

↑ upper-limb regions included right and left shoulder, elbows, wrists, and hands

↓ lower-limb regions included right and left hips, knees, and ankle/foot complex

Abbreviations: *GFAQ* Gillette Functional Assessment Questionnaire, *y* years

Table 2 Between-Days, Test–Retest Reliability Results

Outcome Measure	1st Testing Session Mean ± SD	2nd Testing Session Mean ± SD	ICC (95%CI)	SEM	MDC_{90}	MDC_{95}
10mWT, m/sec (*n* = 38)	1.02 ± 0.20	1.01 ± 0.21	.95 (.90, .97)	0.05	0.11	0.13
F8WT, sec (*n* = 37)×	7.77 ± 2.27	7.51 ± 2.01	.96 (.92, .98)	0.43	0.99	1.19
F8WT, steps (*n* = 37)	13.9 ± 3.1	13.7 ± 2.7	.91 (.82, .95)	0.87	2.02	2.42
360TT, sec (*n* = 38)	2.80 ± 1.03	2.84 ± 0.99	.97 (.93, .98)	0.17	0.41	0.48
360TT, steps (*n* = 38)	6.2 ± 1.4	6.2 ± 1.6	.85 (.70, .92)	0.60	1.39	1.66
mFSST-average, sec (*n* = 32)β	11.81 ± 5.21	10.52 ± 4.25	.94 (.80, .98)	1.16	2.70	3.23
mFSST-best, sec (*n* = 38)	12.01 ± 4.75	10.90 ± 4.15	.87 (.71, .94)	1.61	3.73	4.46

× *n* = 1 missing trial during 1st testing session

β *n* = 8 missing trials among 6 participants

Abbreviations: *ICC* Intraclass Correlation Coefficient, *CI* Confidence Interval, *SEM* Standard Error of Measurement, *MDC* Minimal Detectable Change, *10mWT* 10-m Walk Test, *F8WT* Figure-of-8 Walk Test, *360TT* 360-degree Turn Test, *mFSST* modified Four Square Step Test, *m* meters, *sec* seconds

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 6 of 10

for average number of steps during the 360TT and time to complete the mFSST, regardless of 'average' or 'best' performance. Lower MDCs for average mFSST performance support using the average of 2 trials, when available, but 6 participants, i.e., 16% of the sample, had at least 1 invalid trial, suggesting averaging might not be possible in all ambulatory individuals with AMC.

Discussion

Outcome measures enable objective evaluation of functional changes and can inform clinical decisions, predict future ability, and fulfill healthcare documentation requirements [48]. Kennedy and colleagues called for core gait and functional ambulation outcome measures for use in pediatric clinical and research settings [16]. This study is a first step towards a possible core set of performance-based functional outcome measures for use in adolescents and adults born with AMC. Good-to-excellent between-days, test–retest reliability was found for the 10mWT, F8WT, 360TT, and mFSST, which are currently used in other patient populations. ICCs ≥ 0.90 suggest gait speed per the 10mWT, as well as evaluation of curved path walking and dynamic balance per the timed F8WT and 360TT, may be reliably evaluated among individuals with AMC on two separate occasions. Provided MDCs may enable clinicians to determine whether changes surpass measurement error and indicate 'true change' in their patients with AMC, i.e., pre-to-post rehabilitation (using MDC_{90}) and pre-to-post surgery (using MDC_{95}). Additionally, based on ICCs ≥ 0.70, evaluation of quality of movement, i.e., number of steps, as well as dynamic balance per the mFSST may have clinical trials utility in evaluating changes among groups of ambulatory individuals with AMC, although a potential floor effect should be considered when using the average of 2 mFSST trials, due to 'invalid' trials.

While our study did not include a control group, it appears our individuals with AMC are presenting with worse gait speed and dynamic balance when compared to controls and peers with other lower-extremity pathologies. Among typically-developing children and young adults, 'self-selected' gait speeds for 11–30 year olds are, on average, 1.28–1.36 m/sec, [49] which is significantly faster than speeds obtained in our participants with AMC (i.e., 1.01–1.02 m/sec). Further, Scott et al. found gait speeds of 1.2 ± 0.2 m/sec among adolescents and young adults with hip dysplasia ($n = 24$), suggesting gait speeds with AMC, where multiple lower-limb regions are typically involved, are worse [23]. Collectively, results highlight the importance of evaluating and addressing reduced gait speed among individuals with AMC, particularly since 'self-selected'

gait speed is better correlated to perceived gait quality when compared to other performance-based measures, like the 6-Minute Walk Test, in young adults with congenital, mobility-limiting conditions [50]. Scott et al. also reported adolescents and young adults with hip dysplasia ($n = 24$) had FSST times of 6.6 ± 2.5 s, as compared to controls ($n = 21$; 4.0 ± 0.7 s) [23]. Individuals with AMC in our study had mFSST times that were double that of individuals with single-joint involvement and triple that of controls. Hence, with AMC, dynamic balance appears considerably compromised.

Mean 360TT times among young healthy adults ($n = 34$) have been reported to be 2.2 s, [35] which is about 20% faster than timed 360TT among our participants with AMC. Among children and young adults who are typically developing, peak turn velocity during 180 degree turns are, on average, 221–289 degrees/sec [49]. Among our participants with AMC, as evaluated with a 'quick' 360-degree turn, mean turn velocity was about 129 degrees/sec. Combined, data suggests impaired turning with AMC. As community-ambulation requires frequent turning, [31] and impaired turning has been associated with recurrent falls, [32] it may be imperative to incorporate turning into gait training among individuals with AMC.

Our participants with AMC had better FSST performance (median: 10.52–12.01 s) than children, aged 5–12 years, with Down syndrome and cerebral palsy, (i.e., mean: 18.7 ± 5.7 s), [40] which might be due to use of the mFSST without canes in our study and/or impaired cognition or inattention in the aforementioned pediatric study. Conversely, our participants had worse dynamic balance performance as compared to adults with unilateral lower-limb amputation [51] despite being younger and our use of the mFSST; [39] differences might be attributed to multi-region, lower-limb involvement with AMC.

Our between-days, test–retest reliability findings for performance-based tests are generally similar to reports among other patient populations [23, 24, 28, 29, 34–37, 39, 40, 51, 52]. For example, among adolescents and young adults with hip dysplasia and controls ($n = 34$), Scott et al. reported self-selected gait speed test–retest reliability over 10 timed meters of a 14-m course ($ICC_{2,1} = 0.93$; 95%CI: 0.87-0.96), [23] similar to our 10mWT reliability results ($ICC_{3,k} = 0.95$; 95%CI: 0.90-0.97). Among children with neurological conditions, Graser et al. also reported similar between-days reliability using 10-timed meters over a 14-m course ($ICC_{2,1} = 0.90$; 95%CI: 0.80-0.95). [24] A lower MDC_{95}, i.e., 0.13 m/s in our study and 0.18 m/sec in another study, [52] as compared to 0.35 m/sec in the Scott et al. study, [23] may be due to trial averaging.

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 7 of 10

Between-days test–retest reliability for the F8WT, performed at a given individual's self-selected speed, has been reported by Hess et al. for older adults ($n = 18$; time: ICC $= 0.84$; 95%CI: 0.62-0.94; number of steps: ICC $= 0.82$; 95%CI: 0.59-0.93) [26] and among individuals post-stroke ($n = 35$; time: $ICC_{2,1} = 0.98$; 95%CI: 0.96-0.99) [28]. Better F8WT reliability among our participants with AMC (time: $ICC_{3,k} = 0.96$; 95%CI: 0.92-0.98; steps: ICC $= 0.91$; 95%CI: 0.82-0.95) compared to the Hess et al. study [26] may be secondary to our larger sample size and averaging two trials. For F8WT at fast speed using 2 loops, among older women, Jarnlo and Nordell reported test–retest reliability comparable to our study ($n = 30$; $ICC_{3,1} = 0.93$; 95%CI: 0.85-0.97) [29]. To our knowledge, comparative MDC values are unavailable.

The timed 360TT has published between-days test–retest reliability among individuals with Multiple Sclerosis ($n = 61$; $ICC_{2,2} = 0.91$-0.96; 95%CI: 0.86-0.97; $MDC_{95} = 1.5$ s), [35]. Parkinson's Disease ($n = 14$; ICC $= 0.80$; lower bound of 95%CI: 0.66), [34] and post-stroke ($n = 37$; $ICC_{3,2} = 0.82$-0.95; 95%CI: 0.66-0.98; $MDC_{95} = 0.8$–1.2 s) [36]. Between-days test–retest reliability for number of steps is reported in Parkinson's Disease ($n = 14$; ICC $= 0.77$, lower bound of 95%CI: 0.61) (34) and among older adults (ICC $= 0.92$) [37]. We report similarly good-to-excellent reliability for 360TT time ($ICC_{3,k} = 0.97$, 95%CI: 0.93-0.98) and number of steps ($ICC_{3,k} = 0.85$; 95%CI: 0.70-0.92), but a lower MDC_{95} for 360TT time (i.e., 0.5 s), suggesting changes in turning speed may be more easily identified among individuals with AMC as compared those with other neurological conditions.

While many studies report FSST or mFSST *within-day* test–retest reliability, only a few report *between-days* test–retest reliability, [23, 39, 40, 51] which better parallels 'evaluations' and 're-evaluations' in clinical practice, upon which patient improvements are determined. Among children with neurological conditions ($n = 30$), FSST between-days, test–retest reliability ($ICC_{1,1} = 0.54$-0.89; 95%CI: 0.24-0.95) is reported [40]. Among adolescents and young adults with hip dysplasia and controls ($n = 34$) and adults post-stroke ($n = 17$), between-days, test–retest reliability for average FSST performance ($ICC_{2,1} = 0.93$; 95%CI: 0.87-0.96; $MDC_{95} = 1.66$ s) and best mFSST performance ($ICC_{3,1} = 0.90$; 95%CI: 0.68-0.97) are reported [23, 39]. Between-days, test–retest reliability for best FSST performance is also reported among adults with unilateral lower-limb amputation ($n = 60$; $ICC_{2,1} = 0.97$; 95%CI: 0.94-0.98; $MDC_{90} = 2.0$ s) [51]. Our mFSST reliability (ICC $= 0.87$-0.94; 95%CI: 0.71-0.98) was comparable to aforementioned adult studies;

(23, 39, 51). MDCs, i.e., 2.70–4.46 s, were less than those reported among children with neurological conditions, i.e., 5.29 s. [40].

Study strengths and limitations

Our gait speed assessments were scientifically robust as we used a static starting position for all trials, allowed 2.0 m for acceleration, and provided standardized examiner instructions ('usual pace'); failure to control for any of these factors may negatively impact test–retest reliability. [53] Nevertheless, we could not establish the minimal clinically important difference (MCID) for gait speed, or the other performance-based measures, given the study design. Based on a systematic review of gait speed among individuals with pathology, however, the MCID is likely around 0.1 m/s, [54] which is similar to the MDCs calculated in our study.

Study strengths include recruitment of a mixed sample of adolescents and adults with AMC, as well as selection of clinically-feasible outcome measures, which may enhance clinician adoption [55]. But, we acknowledge some additional limitations. First, without access to medical records, we could not confirm self-reported data, including whether participants were amyoplasia- or distal-type arthrogryposis, as defined by Hall et al [5]. The extent of limb involvement, however, would suggest the majority of our participants might be classified within the amyoplasia-type subgroup. Second, we standardized testing based on current practice, where individuals complete 1–3 recorded trials; we did not evaluate for practice effects (by completing trials until fatigue), which might have resulted in underestimation of performance. Third, we did not specifically target a care-seeking sample, who might have had worse mobility status or greater between-days pain fluctuations, which might have increased floor effects for some measures or negatively influenced test–retest reliability. Finally, performance in a laboratory, or clinical setting, may not reflect real-world performance; for example, gait speed among minors with developmental disorders has been reported to be slower in the real-world when compared to laboratory-obtained gait speed [56].

Conclusions

This study supports subsequent research evaluating linear and curved-path walking, as well as dynamic balance, via the 10mWT, F8WT, 360TT, and mFSST, among individuals with AMC. Future studies with larger sample sizes may seek to establish MCIDs, evaluate floor and ceiling effects among a more diverse sample in terms of mobility and musculoskeletal pain, and determine responsiveness of these performance-based measures

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 8 of 10

to commonly employed interventions, such as bracing, rehabilitation, and surgery, in this patient population. In the interim, practitioners may adopt the 10mWT, F8WT, and 360TT and use provided MDCs when evaluating individuals with AMC for objectively determining intervention effectiveness.

Abbreviations
AMC: Arthrogryposis Multiplex Congenita; F8WT: Figure-of-8 Walk Test; FSST: Four Square Step Test; GFAQ: Gillette Functional Assessment Questionnaire; ICC: Intraclass correlation coefficients; MCID: Minimal clinically important difference; MDC: Minimal detectable change; MDC$_{90}$: Minimal detectable change at 90% confidence level; MDC$_{95}$: Minimal detectable change at 95% confidence level; mFSST: Modified Four Square Step Test; SEM: Standard error of measurement; 10mWT: 10-m Walk Test; 360TT: 360-Degree Turn Test.

Acknowledgements
Not applicable.

Authors' contributions
JMS, MD, EHB, TMS, and LRN were involved in participant recruitment and data collections. JMS, EHB, and RTP analyzed the data. JMS and EHB wrote the first manuscript draft. All authors assisted with results interpretation, read and approved the final manuscript.

Authors' information
Not applicable.

Availability of data and materials
The dataset generated and analyzed during the current study is not publicly available due the prevalence of arthrogryposis multiplex congenita and the limited geographic regions where data collections occurred, resulting in the potential for individual participants to be identified. The dataset is, however, available from the corresponding author on reasonable request at megsions@udel.edu through use of data sharing agreement that takes additional steps to protect participant confidentiality.

Declarations

Ethics approval and consent to participate
The study was approved by the University of Delaware Institutional Review Board for Human Subjects Research (project number: 1354682; initial approval date: 1/16/2019). All individuals signed a written assent/parental permission form/consent form prior to participation.

Consent for publication
Not applicable.

Competing interests
LRN is an educational speaker for Smith and Nephew and Orthopediatrics. All other authors declare they have no competing interests.

Author details
[1]Department of Physical Therapy, University of Delaware, 540 South College Ave., Suite 210JJ, Newark, Delaware 19713, USA. [2]Therapeutic and Rehabilitative Services Department, Nemours Alfred I. duPont Hospital for Children, 1600 Rockland Road, Wilmington, Delaware 19803, USA. [3]Department of Physical Therapy, University of Delaware, 540 South College Ave., Suite 144A, Newark, Delaware 19713, USA. [4]VA Eastern Colorado Geriatric Research, Education, and Clinical Center (GRECC), VA Eastern Colorado Health Care System, 1700 North Wheeling Street, Aurora, CO 80045-7211, USA. [5]Biostatistics Core, University of Delaware, 102B STAR Tower, Newark, Delaware 19713, USA. [6]Orthopedics Department, Nemours Alfred I. duPont Hospital for Children, 1600 Rockland Road, Wilmington, Delaware 19803, USA.

References
1. Dahan-Oliel N, Cachecho S, Barnes D, Bedard T, Davison AM, Dieterich K, et al. International multidisciplinary collaboration toward an annotated definition of arthrogryposis multiplex congenita. Am J Med Genet C Semin Med Genet. 2019;181(3):288–99. https://doi.org/10.1002/ajmg.c.31721.
2. Mennen U, van Heest A, Ezaki MB, Tonkin M, Gericke G. Arthrogryposis multiplex congenita. J Hand Surg Br. 2005;30(5):468–74. https://doi.org/10.1016/j.jhsb.2005.06.004.
3. Dahan-Oliel N, van Bosse HJP, Bedard T, Darsaklis VB, Hall JG, Hamdy RC. Research platform for children with arthrogryposis multiplex congenita: Findings from the pilot registry. Am J Med Genet C Semin Med Genet. 2019;181(3):427–35. https://doi.org/10.1002/ajmg.c.31724.
4. Eriksson M, Villard L, Bartonek A. Walking, orthoses and physical effort in a Swedish population with arthrogryposis. J Child Orthop. 2014;8(4):305–12. https://doi.org/10.1007/s11832-014-0597-9.
5. Hall JG. Arthrogryposis (multiple congenital contractures): diagnostic approach to etiology, classification, genetics, and general principles. Eur J Med Genet. 2014;57(8):464–72. https://doi.org/10.1002/ajmg.c.31716.
6. Stilli S, Antonioli D, Lampasi M, Donzelli O. Management of hip contractures and dislocations in arthrogryposis. Musculoskelet Surg. 2012;96(1):17–21. https://doi.org/10.1007/s12306-012-0180-9.
7. Gagnon M, Caporuscio K, Veilleux LN, Hamdy R, Dahan-Oliel N. Muscle and joint function in children living with arthrogryposis multiplex congenita: A scoping review. Am J Med Genet C Semin Med Genet. 2019;181(3):410–26. https://doi.org/10.1002/ajmg.c.31726.
8. Dai S, Dieterich K, Jaeger M, Wuyam B, Jouk PS, Perennou D. Disability in adults with arthrogryposis is severe, partly invisible, and varies by genotype. Neurology. 2018;90(18):e1596–604. https://doi.org/10.1212/WNL.0000000000005418.
9. Kimber E, Tajsharghi H, Kroksmark AK, Oldfors A, Tulinius M. Distal arthrogryposis: clinical and genetic findings. Acta Paediatr. 2012;101(8):877–87. https://doi.org/10.1111/j.1651-2227.2012.02708.x.
10. Fishman LN, DiFazio R, Miller P, Shanske S, Waters PM. Pediatric orthopaedic providers' views on transition from pediatric to adult care. J Pediatr Orthop. 2016;36(6):e75-80. https://doi.org/10.1097/BPO.0000000000000629.
11. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther. 2006;86(5):735–43.
12. Kasper JD, Chan KS, Freedman VA. Measuring physical capacity. J Aging Health. 2017;29(2):289–309. https://doi.org/10.1177/0898264316635566.
13. Pattison KM, Brooks D, Cameron JI, Salbach NM. Factors influencing physical therapists' use of standardized measures of walking capacity poststroke across the care continuum. Phys Ther. 2015;95(11):1507–17. https://doi.org/10.2522/ptj.20140267.
14. van Bosse HJP, Ponten E, Wada A, Agranovich OE, Kowalczyk B, Lebel E, et al. Treatment of the lower extremity contracture/deformities. J Pediatr Orthop. 2017;37(Suppl 1):S16–23. https://doi.org/10.1097/BPO.0000000000001005.
15. van AbellanKan G, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. J Nutr Health Aging. 2009;13(10):881–9. https://doi.org/10.1007/s12603-009-0246-z.
16. Kennedy RA, Carroll K, McGinley JL, Paterson KL. Walking and weakness in children: a narrative review of gait and functional ambulation in paediatric neuromuscular disease. J Foot Ankle Res. 2020;13(1):10. https://doi.org/10.1186/s13047-020-0378-2.

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 9 of 10

17. Eriksson M, Gutierrez-Farewik EM, Brostrom E, Bartonek A. Gait in children with arthrogryposis multiplex congenita. J Child Orthop. 2010;4(1):21–31. https://doi.org/10.1007/s11832-009-0234-1.

18. Kennedy R, Carroll K, Paterson KL, Ryan MM, McGinley JL. Deterioration in gait and functional ambulation in children and adolescents with charcot-marie-tooth disease over 12 months. Neuromuscul Disord. 2017;27(7):658–66. https://doi.org/10.1016/j.nmd.2017.04.005.

19 Belluscio V, Bergamini E, Tramontano M, Formisano R, Buzzi MG, Vannozzi G. Does curved walking sharpen the assessment of gait disorders? An instrumented approach based on wearable inertial sensors. Sensors (Basel). 2020;20(18):5244. https://doi.org/10.3390/s20185244.

20. Novacheck TF, Stout JL, Tervo R. Reliability and validity of the gillette functional assessment questionnaire as an outcome measure in children with walking disabilities. J Pediatr Orthop. 2000;20(1):75–81.

21. Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. J Rehabil Res Dev. 2016;53(6):797–812. https://doi.org/10.1682/JRRD.2015.12.0228.

22. Moore JL, Potter K, Blankshain K, Kaplan SL, O'Dwyer LC, Sullivan JE. A core set of outcome measures for adults with neurologic conditions undergoing rehabilitation: A clinical practice guideline. J Neurol Phys Ther. 2018;42(3):174–220. https://doi.org/10.1097/NPT.0000000000000229.

23. Scott EJ, Willey MC, Mercado A, Davison J, Wilken JM. Assessment of disability related to hip dysplasia using objective measures of physical performance. Orthop J Sports Med. 2020;8(2):2325967120903290. https://doi.org/10.1177/2325967120903290.

24. Graser JV, Letsch C, van Hedel HJA. Reliability of timed walking tests and temporo-spatial gait parameters in youths with neurological gait disorders. BMC Neurol. 2016;16:15. https://doi.org/10.1186/s12883-016-0538-y.

25. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. J Geriatr Phys Ther. 2013;36(1):24–30. https://doi.org/10.1519/JPT.0b013e318248e20d.

26. Hess RJ, Brach JS, Piva SR, VanSwearingen JM. Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test. Phys Ther. 2010;90(1):89–99. https://doi.org/10.2522/ptj.20080121.

27. Beisheim EH, Horne JR, Pohlig RT, Sions JM. Differences in measures of strength and dynamic balance among individuals with lower-limb loss classified as functional level K3 versus K4. Am J Phys Med Rehabil. 2019;98(9):745–50. https://doi.org/10.1097/PHM.0000000000001183.

28. Wong SS, Yam MS, Ng SS. The Figure-of-Eight Walk test: reliability and associations with stroke-specific impairments. Disabil Rehabil. 2013;35(22):1896–902. https://doi.org/10.3109/09638288.2013.766274.

29. Jarnlo GN, Nordell E. Reliability of the modified Figure of Eight–a balance performance test for elderly women. Physiother Theory Pract. 2003;19(1):35–43.

30. Barker KL, Batting M, Schlussel M, Newman M. The reliability and validity of the Figure of 8 Walk test in older people with knee replacement: does the setting have an impact? Physiotherapy. 2019;105(1):76–83. https://doi.org/10.1016/j.physio.2018.07.003.

31. Glaister BC, Bernatz GC, Klute GK, Orendurff MS. Video task analysis of turning during activities of daily living. Gait Posture. 2007;25(2):289–94. https://doi.org/10.1016/j.gaitpost.2006.04.003.

32. Mancini M, Schlueter H, El-Gohary M, Mattek N, Duncan C, Kaye J, et al. Continuous monitoring of turning mobility and its association to falls and cognitive function: A pilot study. J Gerontol A Biol Sci Med Sci. 2016;71(8):1102–8. https://doi.org/10.1093/gerona/glw019.

33. Al-Muqiren TN, Al-Eisa ES, Alghadir AH, Anwer S. Implementation and use of standardized outcome measures by physical therapists in Saudi Arabia: barriers, facilitators and perceptions. BMC Health Serv Res. 2017;17(1):748. https://doi.org/10.1186/s12913-017-2693-2.

34. Schenkman M, Cutson TM, Kuchibhatla M, Chandler J, Pieper C. Reliability of impairment and physical performance measures for persons with Parkinson's disease. Phys Ther. 1997;77(1):19–27. https://doi.org/10.1093/ptj/77.1.19.

35. Soke F, Guclu-Gunduz A, Ozkul C, Cekim K, Irkec C, Gonenli Kocer B. Reliability and validity of the timed 360 Degrees Turn Test in people with multiple sclerosis. Physiother Theory Pract. 2019:1–12. doi: https://doi.org/10.1080/09593985.2019.1641867.

36. Shiu CH, Ng SS, Kwong PW, Liu TW, Tam EW, Fong SS. Timed 360 Degrees Turn Test for assessing people with chronic stroke. Arch Phys Med Rehabil. 2016;97(4):536–44. https://doi.org/10.1016/j.apmr.2015.11.010.

37. Tager IB, Swanson A, Satariano WA. Reliability of physical performance and self-reported functional measures in an older population. J Gerontol A Biol Sci Med Sci. 1998;53(4):M295-300. https://doi.org/10.1093/gerona/53a.4.m295.

38. Moore M, Barker K. The validity and reliability of the Four Square Step Test in different adult populations: a systematic review. Syst Rev. 2017;6(1):187. https://doi.org/10.1186/s13643-017-0577-5.

39. Roos MA, Reisman DS, Hicks G, Rose W, Rudolph KS. Development of the modified four square step test and its reliability and validity in people with stroke. J Rehabil Res Dev. 2016;53(3):403–12. https://doi.org/10.1682/JRRD.2014.04.0112.

40. Bandong AN, Madriaga GO, Gorgon EJ. Reliability and validity of the four square step test in children with cerebral palsy and down syndrome. Res Dev Disabil. 2015;47:39–47. https://doi.org/10.1016/j.ridd.2015.08.012.

41. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

42. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. Qual Life Res. 2019;28(4):1029–33. https://doi.org/10.1007/s11136-018-2076-0.

43. Nunnally JBJ. Psychometric Theory. 3rd ed. New York: McGraw-Hill; 1994.

44. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007;10 Suppl 2:S94-S105. doi: https://doi.org/10.1111/j.1524-4733.2007.00272.x.

45. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2(14):i-iv, 1–74.

46. Donoghue D, Physiotherapy Research and Older People Group, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med. 2009;41(5):343–6.

47. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil. 2005;84(9):719–23. https://doi.org/10.1097/01.phm.0000176452.17771.20.

48. Gaunaurd I, Spaulding SE, Amtmann D, Salem R, Gailey R, Morgan SJ, et al. Use of and confidence in administering outcome measures among clinical prosthetists: Results from a national survey and mixed-methods training program. Prosthet Orthot Int. 2015;39(4):314–21. https://doi.org/10.1177/0309364614532865.

49. Voss S, Joyce J, Biskis A, Parulekar M, Armijo N, Zampieri C, et al. Normative database of spatiotemporal gait parameters using inertial sensors in typically developing children and young adults. Gait Posture. 2020;80:206–13. https://doi.org/10.1016/j.gaitpost.2020.05.010.

50. Bonnefoy-Mazure A, De Coulon G, Armand S. Self-perceived gait quality in young adults with cerebral palsy. Dev Med Child Neurol. 2020;62(7):868–73. https://doi.org/10.1111/dmcn.14504.

51. Sawers A, Kim J, Balkman G, Hafner BJ. Interrater and test-retest reliability of performance-based clinical tests administered to established users of lower limb prostheses. Phys Ther. 2020;100(7):1206–16. https://doi.org/10.1093/ptj/pzaa063.

52. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism. Phys Ther. 2008;88(6):733–46. https://doi.org/10.2522/ptj.20070214.

53. Stuck AK, Bachmann M, Fullemann P, Josephson KR, Stuck AE. Effect of testing procedures on gait speed measurement: A systematic review. PLoS ONE. 2020;15(6):e0234200. https://doi.org/10.1371/journal.pone.0234200.

54. Bohannon RW, Glenney SS. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. J Eval Clin Pract. 2014;20(4):295–300. https://doi.org/10.1111/jep.12158.

55. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a

Sions *et al. BMC Musculoskeletal Disorders*   (2022) 23:121

Page 10 of 10

"Core Outcome Set" - a practical guideline. Trials. 2016;17(1):449. https://doi.org/10.1186/s13063-016-1555-2.

56. Carcreff L, Gerber CN, Paraschiv-Ionescu A, De Coulon G, Aminian K, Newman CJ, et al. Walking speed of children and adolescents with cerebral palsy: laboratory versus daily life. Front Bioeng Biotechnol. 2020;8:812. https://doi.org/10.3389/fbioe.2020.00812.

Publisher's Note