Test-Retest Reliability of Stepscan System Gait Metrics

By Patrick Connor, PhD, P.Eng. (March 2022)

Abstract

Computerized gait assessment is changing the industry, making assessments more consistent and efficient for clinicians and researchers alike. It is important, however, to understand just how much data is required to reach the level of consistency or reliability that allow for correct clinical decisions. The present work examines gait data collected on a Stepscan® system for a collection of 27 subjects who were each recorded twice about one week apart. The test-retest reliability of the samples was evaluated at various numbers of strides to paint a picture of how reliability changes with different amounts of data. Spatial and Pressure metrics achieve good-excellent levels of reliability, while the temporal metrics fall more squarely in the good levels range, dipping slightly into the moderate range when fewer strides are used. Overall, we see that at least 8 strides should be recorded to achieve acceptable levels of reliability from most metrics, and that there are diminishing returns in terms of reliability by 20 strides.

Introduction

The nature of providing gait assessments based on clinical observation alone leaves open the potential for variations from clinician to clinician based on levels of experience and training. Computerized gait-assessment aims to provide an objective means of measuring gait information, while lowering the threshold of training and experience required to conduct gait assessments. The added objectivity is only useful inasmuch as the data it provides represents an accurate picture of a person's gait. If an individual is assessed initially and later at a follow-up, it is important that any change in the measured gait parameters represent a real change in the individual's gait. To ensure this, there are two kinds of testing that medical devices, including gait analysis platforms, tend to undergo. The first is a test of validity, where the system is either compared to a previously validated "gold standard" device or other process that ensures that the device *accurately* collects data and computes its parameters correctly. The Stepscan® system has been validated in a separate study[1], showing extremely high ICCs for basic spatial and temporal parameters (>0.999 ICCs, (2,k)). The second test is a test of reliability or repeatability: If an individual is tested with the device twice with some time interval in between, how similar will the results be? Any differences are expected to be caused by some combination of the variances in the device/usage protocol and the subject's performance. The present work covers this form of testing. Given the strongly accurate nature of the Stepscan® system from its validity testing, where it was compared to basic manual distance and temporal measures rather than from an existing medical device, the present *reliability* testing process is primarily focused on understanding how many gait cycles worth of data is needed to achieve stable, consistent metrics.

A number of similar studies have been done on related technologies[2-6]. For example, in a test-retest reliability study[2] of the GAITRite® system, van Uden and Besser evaluated a variety of gait metrics for 21 individuals, recording sessions one week apart. Each session had the subject walk across the 4.8 m long GAITRite® mat, starting 2 m before and walking beyond by 2 m. They did this 8 times so that they could guarantee capturing eight left and eight right strides, which apparently are "appropriate for representing gait characteristics by mean values as representative of normal gait"[7]. They evaluated step length, stride length, stride width (base of support), toe out angle, step time, stride time, swing time, stance time, single support time, double support time, and gait velocity. The present work is modeled somewhat after this study's protocol.

Materials and Methods

Participants

There were 27 individuals (14 male, 13 female) that participated in this study, ranging from approximately 20 to 65 years old (mean age of 37 years). The majority of participants were recorded at a local fitness gym and participated either before or after their regular workout. Some participants reported past injury, from which a few were still recovering. This issue did not disqualify them from participating, and it will be considered in the discussion of results. All participants provided informed consent to take part in our protocol, which was designed in accordance with the standard practices listed in Canada's TCPS2 statement.

Equipment, Setup, and Data Collection Procedure

A Stepscan® system is composed of 60 cm x 60 cm pressure sensitive tiles that can be connected in a series to create a pedway or connected in other ways to create a platform of most any practical shape or size. The entire surface area is covered in active sensors at 5 mm intervals, resulting in 14400 sensors per tile, which are scanned at 100 Hz. The sensors are pressure sensitive, which allows the system to capture not only the spatial and temporal parameters of gait, but also underfoot pressures. For the purposes of the present study, we will look at the reliability of the spatial, temporal, and pressure parameters for Gait.

Figure 1 illustrates the setup that was used in this study. A 5-tile, 3.0 m long Stepscan® system was setup. At each end, an inactive tile (same size and shape) was also placed, followed by a ramp, followed by a tape line approximately 2 m away from the edge of the nearest *active* tile. These measures were taken to help individuals reach a natural walking pace/pattern by the time they reached the active tiles.

Starting behind the tape line, participants were instructed to walk across the system, past the other taped line, turn around and repeat for approximately 3 minutes of walking. This typically resulted in 50 or more gait strides. Participants returned about one week later (average of 6.3 days) and repeated the procedure. After the second visit, participants received an assessment report as compensation that contained their gait metrics. Data was recorded using the Stepscan® clinical software application (Release 2.5.26). All participants walked unshod.

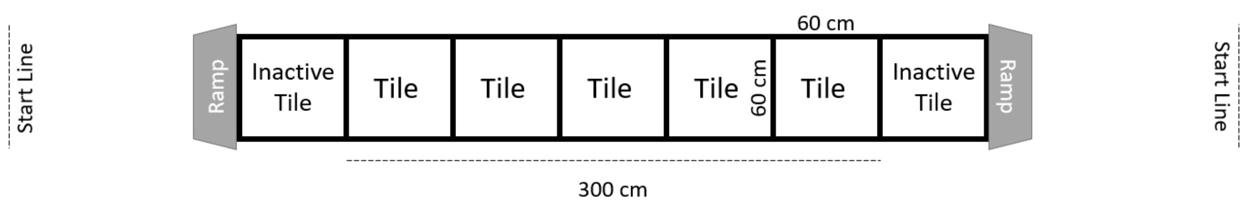

Figure 1, Equipment Setup

Analyses

The data was processed in the Stepscan® software to properly label footprints, delete partial footprints, etc., according to standard Stepscan® data processing practices. In this "full" dataset, there was an average of 56 strides, ranging between 42 and 74 strides. The data was further processed to cut back on the number of strides, so that we could examine the effect different numbers of strides has on the reliability of various metrics. In this way, the data was reprocessed to include only the first 40, 20, 14, and 8 strides. The intention was also that we might estimate (interpolate) an appropriate number of strides required for a given metric, if the threshold of sufficient reliability were to fall between two specific numbers of strides.

Like van Uden and Besser, we will also evaluate step length, stride length, stride width (base of support), toe out angle, step time, stride time, swing time, stance time, single support time, double support time, and gait velocity. To this we will also add evaluating the reliability of regional pressure distribution metrics to address the pressure measurement side of the Stepscan® system. The way in which all of these evaluated metrics are determined by the Stepscan® software is described in Table 1.

Standard measures of test-retest reliability are intra-class correlation coefficients. Here we report the ICC(2,k) formulation (Two-way random, average measures, absolute agreement)[8], which is commonly used in test-retest scenarios.

Table 1, Evaluated gait measures

Gait Metric	*Stepscan® Metric Measurement/Calculation*
Step Length	For each stride, a left and right step length are computed. The step length is the distance from the geometric center of one footprint to the geometric center of the next footprint along the walking direction, which is defined by the line connecting the geometric centers of the first and third footprints in the associated stride.
Stride Length	The distance between the geometric centers of the first and third footprints of a stride.
Stride Width	Calculated as the perpendicular distance between the line connecting the first and third footprint geometric centers in a stride and the geometric center of the second footstep.
Toe out Angle	Calculated as the angle between the mid-line of the footprint (the major axis of an optimally fitting an ellipse) and the line connecting the geometric center of the same footprint to the geometric center of the footprint two steps ahead (the walking direction).
Step Time	Computed as the difference between the initial contact times of consecutive footsteps.
Stride/Gait Cycle Time	Computed as the difference between a stride's first and third footprint initial contact times.
Swing Time	The difference between a foot's stance time and the duration of the gait cycle it initiates.
Stance Time	The time between initial contact and toe off

Single Support Time	The time between the toe off and next initial contact of the opposite foot minus the zero support time. Here all participants walked, so the zero support time is 0.
Double Support Time	The time between the initial contact of one foot and the subsequent toe off of the opposite foot.
Velocity	Stride length / Stride time
Average Foot Pressure	The average of the pressures under each left or right footprint, averaged over all of the left or right footprints in a recording.
Lateral/Medial Pressure Percentage	The percentage of force applied to the lateral/medial halves of the footprint as bisected by the major axis of an optimally fitted ellipse.
Fore/Rear Pressure Percentage	The percentage of force applied to the fore/rear halves of the footprint as bisected by the minor axis of an optimally fitted ellipse.
Pressure Time Integral	The sum of the maximum pressure in each single footstep image multiplied by the time difference between it and the previous image.

Results

Table 2 contains the ICC reliability measures for the metrics described in Table 1 for each number of gait strides processed. Results are averaged across left and right versions of a metric where applicable (e.g. the step length ICC and confidence interval bounds are averaged from the corresponding left and right results). Averages are also taken for the medial/lateral and fore/rear metrics.

The results can be broken down into spatial, temporal, and pressure categories. Let us examine each of these in turn. Figure 2 presents the ICC(2,k) results for the spatial metrics (the 95% CI is not shown here for clarity; see Table 2 for that information), where an exponentially declining curve (i.e., an arctan curve) is fitted to the data points, which are clearly marked at 8, 14, 20, 40, and 56 strides. The results in this figure suggest that capturing 8 strides alone is sufficient to achieve excellent (> 0.90) reliability[8] for all of the spatial metrics.

Figure 3 illustrates analogous results for temporal metrics. Here, the reliability levels range from moderate to excellent, depending on the number of strides involved. Most fall in the range of having good reliability (0.75-0.90). Most notably, single and double support time fall into the moderate range (0.5-0.75) for the 8 strides results but enter the good range by 14 strides. The fitted arctan curves suggest that 14 strides are needed to enter the good reliability range for both of the support time metrics.

Figure 4 shows a mixed set of results for the pressure metrics. The average foot pressure demonstrates excellent reliability, as do the medial/lateral pressure percentages aside from the 8 strides case dipping slightly into the good reliability range. The fore/rear pressure percentages, however, span much of the good reliability range, where only 8 strides are again at the lower end.

Table 2, Test-retest reliability ICC values by metric and number of strides

Gait Metric	*ICC(2,k) with 95% CI, by Number of Strides*				
	8 Strides	*14 Strides*	*20 Strides*	*40 Strides*	*All Strides*
Step Length	.932 [.850, .969]	.933 [.856, .970]	.944 [.878, .974]	.945 [.881, .975]	.950 [.890, .977]
Stride Length	.933 [.852, .970]	.934 [.856, .970]	.945 [.881, .975]	.946 [.882, .975]	.951 [.893, .978]
Stride Width	.915 [.813, .961]	.934 [.854, .970]	.941 [.871, .973]	.952 [.896, .978]	.956 [.903, .980]
Toe out Angle	.968 [.905, .987]	.987 [.938, .995]	.991 [.970, .997]	.993 [.866, .972]	.993 [.982, .997]
Step Time	.854 [.681, .933]	.869 [.714, .940]	.873 [.723, .942]	.900 [.779, .954]	.918 [.821, .963]
Stride/Gait Cycle Time	.863 [.704, .937]	.873 [.724, .942]	.877 [.731, .944]	.902 [.783, .955]	.919 [.822, .963]
Swing Time	.788 [.530, .904]	.883 [.746, .947]	.884 [.749, .947]	.893 [.769, .951]	.913 [.812, .960]
Stance Time	.761 [.485, .890]	.829 [.629, .921]	.855 [.684, .933]	.877 [.733, .944]	.899 [.779, .954]
Single Support Time	.688 [.317, .857]	.780 [.516, .900]	.821 [.604, .919]	.903 [.788, .956]	.887 [.748, .949]
Double Support Time	.639 [.224, .834]	.765 [.493, .892]	.810 [.588, .913]	.833 [.639, .924]	.850 [.675, .931]
Velocity	.864 [.704, .938]	.873 [.723,.942]	.879 [.737, .945]	.895 [.773, .952]	.910 [.804, .959]
Average Foot Pressure	.953 [.897, .979]	.963 [.919, .983]	.967 [.927, .985]	.976 [.948, .989]	.977 [.950, .990]
Lateral/Medial Pressure Percentage	.910 [.802, .959]	.967 [.928, .985]	.968 [.929, .985]	.970 [.935, .986]	.972 [.939, .987]
Fore/Rear Pressure Percentage	.788 [.440, .910]	.855 [.602, .939]	.876 [.610, .951]	.893 [.684, .957]	.903 [.714, .961]
Pressure Time Integral	.909 [.798, .958]	.906 [.794, .957]	.922 [.830, .965]	.942 [.873, .974]	.950 [.890, .977]

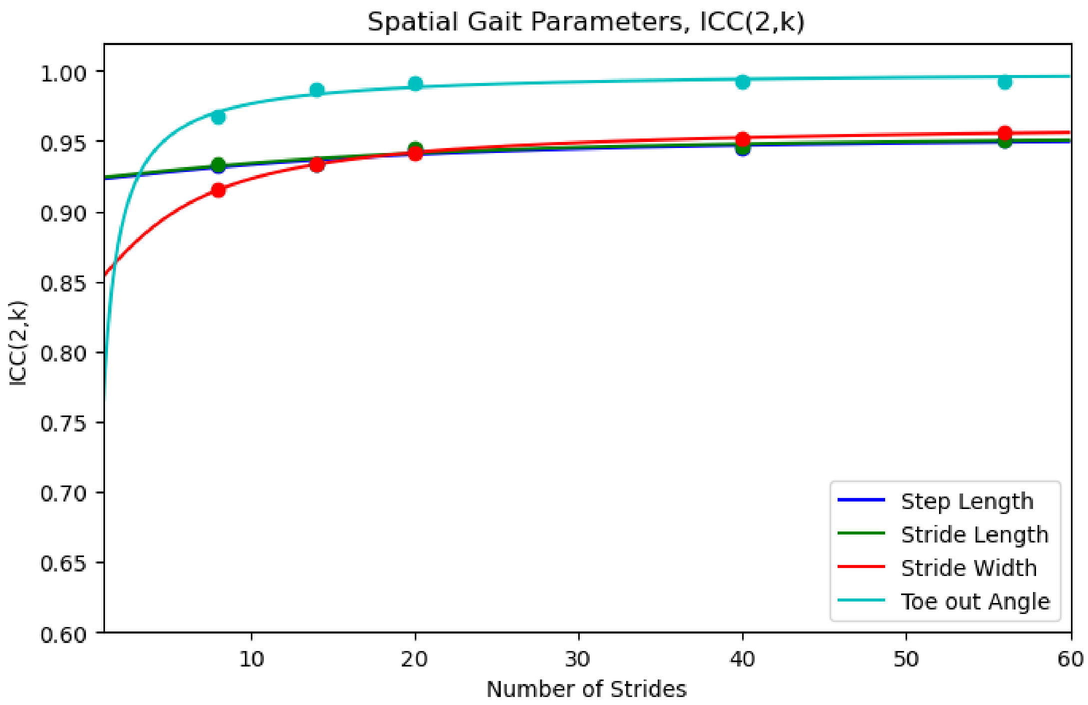

Figure 2, Spatial Gait Parameter ICCs

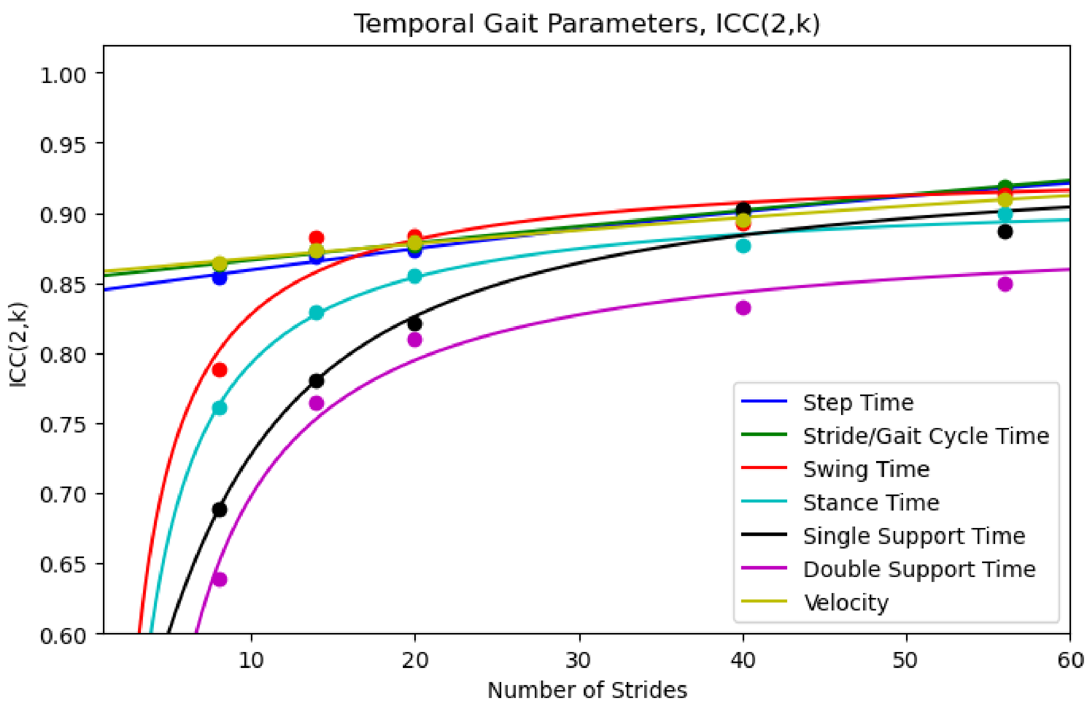

Figure 3, Temporal Gait Parameter ICCs

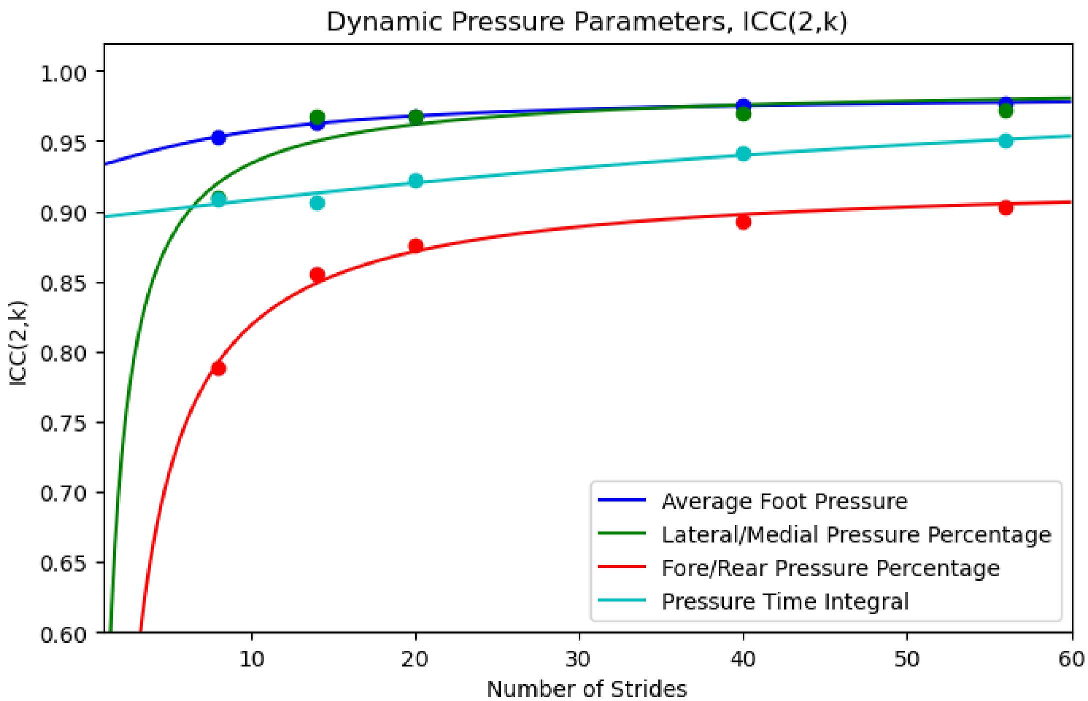

Figure 4, Pressure Parameter ICCs

Discussion

In relation to other test-retest reliability studies, the present study results are comparable. In the study by van Uden and Besser, there were 8 trials (or passes) which would have given, on average, about 48 strides. An average step length is 60-65 cm, so for their 4.8 m active sensor area we could expect 8 steps or 6 strides per pass (6*8=48). This corresponds to approximately the same amount of data as our data set with the full number of strides collected (their 48 vs. our 56 strides). Comparing only the metrics evaluated by both, the spatial metrics except for the base of support/stride width were extremely close. The Stepscan® base of support had an ICC of 0.96 whereas the other study achieved a 0.79[*]. For the temporal values, the Stepscan® results were typically a little lower. Coincidentally, the single and double support values in the van Uden and Besser study were among their least reliable metrics, as was seen here. When compared generally to the spatial and temporal metrics found in a study on the Win-Track platform[3], results here compare favourably.

Unfortunately, in regard to plantar pressure data, we did not find an adequate comparison in the literature. Two studies[12,13] looking at other plantar pressure measurement devices were incompatible for comparison. One used far fewer steps and both divided the foot into more finely resolved areas than the lateral/medial and fore/rear divisions that Stepscan® uses. There are a number of additional

[*] Notably, the base of support or stride width has been noted in other studies[9,10] involving the GAITRite® as being among its least reliable or least consistent metrics.

studies that involve insole-based plantar pressure measurement, but the difference in their form or design to a pressure-sensitive flooring system precludes a relevant comparison.

How much gait data is enough? According to van Uden and Besser, a former study involving 8 left and 8 right strides (16 strides total) was found to be sufficient data to achieve reliable gait metrics. The International Foot and Ankle Biomechanics community issued a set of voluntary standards[11] for plantar pressure measurement devices in which they indicated after interviewing clinical practitioners that 10 footprints minimum were typically recorded per subject for plantar pressure analysis. In the present work, it would appear that a bare minimum of 8 strides (4 passes on a standard 4-tile Stepscan® system) should be used to determine gait metrics with a reasonable degree of reliability. We find here, also, that there are diminishing returns for most metrics by around 12 or 14 strides (6 or 7 passes). Beyond 20 strides, it would appear there is very little to gain in terms of reliability.

Although not yet expressed in our results, metrics could be slightly less reliable on one side (e.g., left) than the other. The most extreme case of this was seen in the left single support time vs. the right single support time, where the 8 and 14 strides reliability results were substantially lower (by 0.322 and 0.163, respectively) for the left side than for the right. When reviewing the 8 strides data, it was found that there were unequal numbers left and right strides. There were 16 samples with 5 left strides and 3 right strides, 29 samples with 4 left and 4 right, and only 8 samples with 3 left and 5 right. Thus, there were more left strides on average. Left strides are L-R-L, so the single support time is for the right foot. Thus, there is more data for right single support time measures than for left, which may explain the discrepancy between them.

A useful way of summarizing gait metrics is to show the symmetry between left and right sides for particular metrics, and is frequently employed in Stepscan® software reports. Notably, we found that symmetry indices (SI), which are calculated as (right - left)/(left + right)*100, often appear to have a greatly lowered reliability than the metrics on which they are based. The same behaviour sometimes occurs when percentages (left/right, or percentage duration of the gait cycle) are invoked. In both cases, this is believed to be due to the fact that normalizing metrics leads to a reduction in variation, which in turn lowers the apparent reliability. To overcome this effect, a population would need to be found where individuals are consistent but where the cohort generates greater variation for such normalized metrics.

As was noted above, participant data was included in this study even when participants were recovering from an existing malady that likely impacted their gait, and when they were recorded before their workout in one session and recorded after their workout in the other session. We expect that this inclusion would tend to lower the ICC results overall. Yet, these participants' data were used in the study to provide an *n* that would give statistical power so as to be sufficient and comparable to other test-retest reliability studies. It could also be argued that such circumstances are true-to-life and such results provide a better representation of what a clinician can expect in real world use of the Stepscan® system.

Conclusions

In this test-retest reliability study, we found that Stepscan® parameters achieve good to excellent reliability overall, aside from two metrics dipping into the moderate range at the lowest number of strides. At only 8 strides, we found that all but a few general-purpose parameters are highly reliable. By including 12 or 14 strides and no more than 20 strides the levels of reliability plateau suggesting that this amount of data should be appropriate for most use cases involving spatial, temporal, and pressure metrics of gait going forward.

References

1. Connor PC. Concurrent validity of Stepscan system measurements. [White paper], Stepscan Technologies Incorporated. 2022 Feb.
2. Van Uden CJ, Besser MP. Test-retest reliability of temporal and spatial gait characteristics measured with an instrumented walkway system (GAITRite®). BMC Musculoskeletal Disorders. 2004 Dec;5(1):1-4.
3. Ramachandra P, Maiya AG, Kumar P. Test-retest reliability of the Win-Track platform in analyzing the gait parameters and plantar pressures during barefoot walking in healthy adults. Foot & ankle specialist. 2012 Oct;5(5):306-12.
4. Menz HB, Latt MD, Tiedemann A, San Kwan MM, Lord SR. Reliability of the GAITRite® walkway system for the quantification of temporo-spatial parameters of gait in young and older people. Gait & posture. 2004 Aug 1;20(1):20-5.
5. Godi M, Turcato AM, Schieppati M, Nardone A. Test-retest reliability of an insole plantar pressure system to assess gait along linear and curved trajectories. Journal of neuroengineering and rehabilitation. 2014 Dec;11(1):1-8.
6. Beckwée D, Degelaen M, Eggermont M, Gonzalez-Rodriguez M, Lefeber N, Vaes P, Bautmans I, Swinnen E. Validity and test-retest reliability of the Stride Analyzer in people with knee osteoarthritis. Gait & posture. 2016 Sep 1;49:155-8.
7. Besser MP, Kmieczak K, Schwartz L, Snyderman M, Wasko J, Selby-Silverstein L. Reliability of temporal spatial gait parameters means in adults without impairment. *Annual Meeting of the Gait and Clinical Movement Analysis Society.* Dallas, Texas, Gait Posture. 1999.; 9(2):113.
8. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine. 2016 Jun 1;15(2):155-63.
9. Roche B, Simon AL, Guilmin-Crépon S, Boizeau P, Andriss B, Alberti C, Presedo A, Ilharreborde B, Husson I. Test-retest reliability of an instrumented electronic walkway system (GAITRite®) for the measurement of spatio-temporal gait parameters in young patients with Friedreich's ataxia. Gait & Posture. 2018 Oct 1;66:45-50.
10. Thorpe DE, Dusing SC, Moore CG. Repeatability of temporospatial gait measures in children using the GAITRite electronic walkway. Archives of physical medicine and rehabilitation. 2005 Dec 1;86(12):2342-6.
11. Giacomozzi, C., Keijsers, N., Pataky, T., & Rosenbaum, D. (2012). International scientific consensus on medical plantar pressure measurement devices: technical requirements and performance. *Annali dell'Istituto superiore di sanita*, *48*, 259-271.
12. Zammit, G. V., Menz, H. B., & Munteanu, S. E. (2010). Reliability of the TekScan MatScan® system for the measurement of plantar forces and pressures during barefoot level walking in healthy adults. *Journal of foot and ankle research*, *3*(1), 11.
13. Gurney JK, Kersting UG, Rosenbaum D. Between-day reliability of repeated plantar pressure distribution measurements in a normal population. Gait & posture. 2008 May 1;27(4):706-9.