RELIABILITY OF TWO-DIMENSIONAL VIDEO-BASED RUNNING GAIT ANALYSIS

Mark F. Reinking, PT, PhD, SCS, ATC¹

Leigh Dugan, SPT¹

Nolan Ripple, SPT¹

Karen Schleper, SPT¹

Henry Scholz, SPT¹

Jesse Spadino, SPT¹

Cameron Stahl, SPT¹

Thomas G. McPoil, PT, PhD¹

ABSTRACT

Background: While two-dimensional (2D) video running analysis is commonly performed in the clinical setting, the reliability of quantitative measurements as well as effect of clinical experience has not been studied.

Purpose: The purpose of this study was to assess the intra-rater and inter-rater reliability of six different raters using 2D video analysis of sagittal and frontal plane kinematic variables while running on a treadmill.

Study Design: Cross-sectional Study

Methods: Running videos from 10 individuals (five female, five male) with a mean age of 22.8 years were selected for analysis. Two raters had over 10 years of experience with running video analysis and the other four raters had no prior experience. Before beginning analyses, the senior investigator conducted two hours of training with all raters to review the measurement procedures and the movement analysis software program. After completing training and one practice analysis, each rater assessed four 60-second video clips for the 10 runners twice (20 total). A minimum of one week separated the two assessments on each runner. The order of the runner analyses were randomly assigned and each rater completed a single analysis within 24 hours. After the rater had completed their initial assessment on all 10 runners, a second analysis was completed one week later with a different order of randomization. Eight sagittal plane (SAG) and four frontal plane (FRT) quantitative variables were measured for the left and right lower extremities on all 10 runners. Intra- and inter-rater reliability was assessed using intraclass correlation coefficients (ICC) and standard error of the measurement (SEM).

Results: The intra-rater ICC values for experienced raters ranged from 0.75 to 0.98 for the SAG and 0.45 to 0.96 for FRT variables. The inter-rater ICC values between the experienced raters ranged from 0.76 to 0.99 for the SAG and 0.82 to 0.98 for FRT variables. The intra-rater ICC values for inexperienced raters ranged from 0.54 to 0.99 for the SAG and 0.08 to 0.97 for FRT variables. The inter-rater ICC values between the inexperienced raters ranged from 0.93 to 0.99 for the SAG and 0.79 to 0.98 for FRT variables. Intra-rater SEM values based on average means of all raters ranged from 1 to 47% of the mean values obtained for the SAG and from 6 to 158% for the FRT variables.

Conclusions: The intra-rater and inter-rater reliability levels were higher for SAG quantitative variables assessed in this study in comparison to FRT variables. Experience does not appear to be a factor when consistency is required with repeated analyses on the same runner.

Level of Evidence: 4, Controlled laboratory study

Keywords: Gait analysis, interrater, intra-rater, reliability, running assessment

CORRESPONDING AUTHOR

Mark F. Reinking

School of Physical Therapy

Regis University

3333 Regis Blvd, G-4

Denver, CO 80221

E-mail: reinking@regis.edu

¹ Rueckert-Hartman College of Health Professions, School of Physical Therapy, Regis University, Denver, CO, USA

The authors declare no conflict of interest in matters associated with this paper.

INTRODUCTION

While two-dimensional (2D) video running analysis is commonly performed in the clinical setting, the ability of a single clinician or multiple clinicians to reliably analyze various lower extremity alignment variables quantitatively in both the sagittal and frontal planes is still unclear. Although three-dimensional (3D) video-based running gait analysis has long been considered the gold standard, the expense of the required equipment and complexity of the analysis software makes this type of running gait analysis prohibitive for most sports physical therapy clinics. In general, a video running analysis assesses various body postures and alignments in both the sagittal and frontal planes. Maykut et al¹ compared frontal plane motion variables captured using 2D and 3D video analysis techniques and reported a strong correlation between both analysis methods when assessing the hip adduction angle. The clinical interest in the hip adduction angle, which is a measure of pelvic drop, is based on findings demonstrating that excessive pelvic drop has been linked to patellofemoral pain syndrome and tibial stress fracture.^{2,3} Wille et al⁴ reported that several SAG kinematic variables during running were predictive of ground reaction force variables and joint kinetics during running. These SAG kinematic variables included center of mass (COM) height at midstance and double float, foot inclination angle at initial contact, horizontal distance from shoe heel to COM line at initial contact, and peak knee flexion angle. Although Wille et al⁴ utilized 3D video running analysis for data collection, they suggested that 2D video analysis would likely be sufficient to capture these kinematic variables given the strong correlations between 3D and 2D SAG motions during running. Schurr et al⁵ reported a high correlation in SAG variables between 2D and 3D video analysis, but poor correlations for the FRT variables. At present, little research has been conducted on the reliability of these analyses based on clinician experience as well as when the analyses are performed over time.

Damsted et al⁶ assessed the reliability of two experienced raters to measure knee and hip flexion angles at initial foot strike of 18 individuals running on a treadmill using 2D video recordings. They reported that intra-rater and inter-rater reliability was sufficient

between the two experienced raters but that clinicians should be aware that measurement variations can occur, especially when measurements are taken over different days. Unfortunately, their conclusions are limited to those two SAG variables. Krummen et al⁷ determined only the inter-rater reliability of 12 physical therapists to assess various sagittal and frontal plane running characteristics of six individuals running on a treadmill. While they reported that the inter-rater reliability between the 12 therapists ranged from fair to strong, their research was published as a platform presentation abstract. Thus, specific details regarding data collection and analysis procedures are unknown. One factor that could have affected the results was that Krummen et al⁷ used an iPad to film the six runners. While the iPad provides an easy tool for recording gait in the clinic, the low frame rate (30 frames per second) available on the iPad used by these researchers could be a factor affecting the results. Most researchers agree that the minimum frame rate that should be used for recording running is at least 120 frames per second.⁸ In a more recent study, Pipkin et al⁹ assessed the reliability of three experienced raters performing a qualitative video analysis of 15 individuals running at a self-selected pace on a treadmill using a single camera filming at rate of 120 frames per second. Raters assessed eight sagittal and seven frontal plane body posture and alignment variables using a 3-point or 5-point scoring scale. Using a weighted Kappa statistic, they reported that the level of intra-rater agreement was substantial to excellent for 11 of the 15 variables, whereas only five of the 15 variables had the same level of inter-rater agreement. Of the five variables achieving a rating of substantial (Kappa greater than 0.60) or higher, three were FRT and two were SAG variables. While this study provides valuable information on the reliability of experienced raters performing a 2D video-based running analysis, there are several factors that restrict the application of the study findings. Since only experienced raters were used, the amount of training and experience required to achieve a satisfactory level of reliability is unknown. Additionally, Pipkin et al⁹ provided each rater with pre-identified still frames to ensure that raters were evaluating the same frame and stride within the video. Since clinicians typically perform a running analysis by randomly selecting a

---

stride for analysis, the pre-identified still frames in the Pipkin et al⁹ study could have positively affected the levels of agreement among raters. Just as importantly, while qualitative analysis provides valuable information to the runner, there are also situations in which quantitative kinematic measurements are of value. These situations may include counseling a runner for injury prevention or when documenting changes in the running kinematics because of rehabilitation following injury. For example, Dierks et al¹⁰ suggested that a knee flexion difference less than 40° between initial contact and midstance during stance phase can be associated with patellofemoral pain.

Based on previous research, the consistency of quantitative measurements for sagittal and frontal plane variables when performing a video-based running analysis has not been assessed throughout the entire running cycle. In addition, the influence of rater experience on the reliability of quantitative measurements is unknown. Therefore, the purpose of this study was to assess the intra-rater and inter-rater reliability of six different raters in assessing various sagittal and frontal plane kinematic variables recorded while running on a treadmill using 2D video analysis. Based on previous research, it was hypothesized that 1) quantitative measurements for SAG kinematics would have higher levels of reliability, irrespective of rater experience, than quantitative measurements for frontal plane kinematics, and 2) experienced raters would demonstrate higher levels of reliability than inexperienced raters when assessing both sagittal and frontal plane kinematics.

METHODS

Video Selection

Running gait analysis videos from 10 individuals (five female) were selected from pre-existing 2D videos recorded as part of a previous research study that assessed running-gait between recreational and competitive runners (high school and college cross country runners). Ten runners were selected based on subject size used in previous running reliability studies^{1,6,9,11,12} and number of raters in this study. The mean age of the 10 runners was 22.8 years, with a range of 15 to 29 years. The videos of the 10 individuals were randomly selected from a video pool of 90

runners who were known to be without injury at the time of video capture. All videos were recorded while the individuals ran on a treadmill (Model Mercury S, Woodway USA Inc., Waukesha, WI 53186) using one camera mounted to a portable tripod (Model# EX FH25, Casio America Inc., Dover, NJ 07801) at a rate of 240 frames per second. All 10 individuals had previous experience running on a treadmill and ran at a self-selected speed. To facilitate the observation of marker movement, each participant wore running shorts with the males wearing no shirts and the females wearing a halter-top. Each participant was initially asked to run for five minutes to acclimate to the testing treadmill. Once acclimated, 9 mm spherical reflective markers were placed on each lower extremity over the following locations: anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), lateral epicondyle of the femur, lateral malleolus, lower posterior calf above the Achilles tendon (two markers), and over the midline of the heel (two markers). To minimize ASIS and PSIS marker movement, elastic self-adhesive wrapping was applied around the individual's waist prior to marker placement. Once all markers were attached, the participant was then asked to start running on the treadmill at his or her pre-selected running speed. Once the subject indicated they were in their typical running pattern, they continued to run at their preferred speed for five minutes while video data was recorded. Video data were collected sequentially using one camera for the following four views; right sagittal (side); left sagittal (side), full posterior (frontal), and posterior leg/heel (frontal). All runners selected for this study had previously met the following inclusion criteria: (1) between the ages of 14 to 40 years; (2) ran at least 18 miles per week for one-year prior to participation in the study; (3) had experience running on a treadmill; (4) no previous history of lower extremity congenital or traumatic deformity or previous surgery that resulted in altered bony alignment; and (5) no acute injury three months prior to the start of the study that led to inability to run at least three consecutive days during that time. The Regis University Institutional Review Board approved the study protocol and all participants provided written informed consent prior to participation in the study, and for subjects under 18 years old, parental consent was obtained.

Video Analysis Procedures

Sixty-second video clips (approximately 10 complete strides per clip) were created from each of the original five-minute videos for the reliability assessment. To ensure that the video clip represented the most acclimated pattern of running, the original video recording for each view (five minutes or 300 seconds in length) was trimmed so that the 60-second video clip began after three minutes or 180 seconds of running on the treadmill. Six raters, two experienced and four inexperienced, were asked to assess the video clips for each of the ten runners. The two experienced raters had over 10 years of experience performing 2D video-based running analyses on both collegiate and recreational runners. The four inexperienced runners were second-year graduate physical therapy students with no prior experience in performing a video-based running analysis.

Prior to initiating the video assessments, all raters whether experienced or inexperienced received a total of three hours of training by the senior investigator. During the initial training two-hour training session, all raters were given written and verbal instructions for performing each of the quantitative measurements used in the study as well as the free-access video analysis software program (Kinovea, version 0.8.15, <http://www.kinovea.org>) used by all raters. At the end of the initial training session, all raters were asked to complete a “practice” 2D running analysis on a runner not included in the 10 runners used for the reliability assessment. The results of analysis on the “practice” runner completed by each rater were then reviewed with all raters during another one-hour training session so that any questions or concerns with the measurement procedures or video analysis software could be addressed. At the completion of this second training session, the reliability assessment for the 10 runners was initiated. Each rater assessed all four video clips for each of the 10 runners twice (20 total assessments) with at least one week separating the two assessments of the same runner. A study coordinator, who did not perform any video assessments, randomly assigned a runner for assessment to each rater. When assigned a runner for analysis, the rater was provided a file with the four video clips and asked to complete the assessment within 24 hours. All raters

were instructed to select a stride for analysis after the third foot strike on the video clip to allow them the opportunity to observe the runner's gait pattern and enhance the rater's ability to identify initial foot contact. Once the video analysis was completed, the rater returned the data sheet with the quantitative measurements to the study coordinator. Once the study coordinator received the video clips and data sheet, they provided the rater with the next runner's file with the video clips to be assessed. The same assignment process was used until each rater completed their initial assessment on all 10 runners, at which time the process was repeated with a different order of randomization of the 10 runners for the repeated set of analyses.

Each rater assessed the following variables on the left and right lower extremities for all 10 runners. SAG variables included; 1) angle of the shoe to treadmill at initial contact, 2) angle of the lower leg at initial contact, 3) knee flexion at initial contact, 4) knee flexion at midstance, 5) position of the COM to the posterior point of shoe at initial contact, 6) knee flexion at midswing, 7) vertical position of the center of mass at midstance, and 8) vertical position of the center of mass at double float. FRT variables included; 1) hip adduction angle at initial contact, 2) hip adduction angle at midstance, 3) rearfoot angle at initial contact, and 4) rearfoot angle at midstance. All angles were measure in degrees and all distance measurements were taken in centimeters.

Data analysis

Intraclass correlation coefficients (ICC) were calculated to determine the consistency of each rater to repeatedly perform the measurements individually (intra-rater; $ICC_{3,1}$) between the two video analysis sessions as well as in comparison to the other raters (inter-rater; $ICC_{2,k}$). The level of reliability for the ICC was classified using the characterizations reported by Landis and Koch.¹³ These characterizations were: *slight*, if the correlation ranged from 0.00 to 0.20; *fair*, if the correlation ranged from 0.21 to 0.40; *moderate*, if the correlation ranged from 0.41 to 0.60; *substantial*, if the correlation ranged from 0.61 to 0.80; and *almost perfect*, if the correlation ranged from 0.81 to 1.00. The inter-rater reliability

Table 1. Participant Demographics					
	Age (years)	Height (cm)	Weight (kg)	BMI (kg/m ²)	Miles per week
Females (n=5)	26.0 ± 3.1	164.8 ± 5.3	54.6 ± 2.6	20.1 ± 0.8	28.2 ± 11.1
Males (n=5)	20.0 ± 3.4	176.4 ± 3.2	65.0 ± 9.3	20.9 ± 2.4	37.0 ± 17.2
ALL participants	23.0 ± 4.6	170.6 ± 7.4	59.8 ± 8.5	20.5 ± 1.8	32.0 ± 15.1

was assessed between the two experienced raters, the four inexperienced raters, and between the experienced and inexperienced raters. For the comparison between the two experienced raters and between the four inexperienced raters, the variable measures for the two sessions were averaged. To calculate the inter-rater reliability between the experienced raters and the inexperienced raters, the average measure of each rater was averaged for the experienced raters (n=2) and for the inexperienced raters (n=4). The standard error of the measurement (SEM) was also calculated as another indicator of intra-rater for all six raters. The SEM is in the same units as the original measurement and represents how the measurements would vary if

measured more than once by each rater.¹⁴ All statistical analyses were performed using SPSS software, Version 23 and an alpha level of 0.05 was established for all tests of significance.

RESULTS

Demographic data for the 10 participants in the study are listed in Table 1. The intra-rater ICC and SEM values for SAG and FRT variables are listed in Table 2 and 3, respectively. The inter-rater ICCs for SAG and FRT variables are provided in Table 4.

Sagittal Plane Variables. The intra-rater ICC values for all raters and all variables ranged from 0.54 to 0.99 with the SEM ranging from 1% to 47% of the mean value for all variables assessed. For the *experienced raters*, the intra-rater ICC values ranged from 0.75 to 0.98 with the SEM ranging from 1% to 38% of the mean values for all variables. For the *inexperienced raters*, the intra-rater ICC values ranged from 0.54 to 0.99 with the SEM ranging from 1% to 47% of the mean values for all variables. For all raters, 70% of all SAG SEM values were less than 10% of the variable

Table 2. Intra-rater Reliability of Sagittal Plane Variables (ICC3,1), Mean (2 measures), and SEM (standard error of measurement)																		
Variable	Rater 1 (12 years exp)			Rater 2 (10 years exp)			Rater 3 (no exp)			Rater 4 (no exp)			Rater 5 (no exp)			Rater 6 (no exp)		
	ICC	Mean	SEM	ICC	Mean	SEM	ICC	Mean	SEM	ICC	Mean	SEM	ICC	Mean	SEM	ICC	Mean	SEM
L Shoe Angle	0.98	6.10	1.32	0.98	6.45	1.17	0.88	7.40	3.06	0.98	6.10	1.59	0.96	6.40	1.86	0.98	7.30	1.34
L Tibial Angle	0.89	6.85	0.75	0.88	6.30	0.85	0.90	5.85	0.86	0.76	6.45	0.92	0.94	6.50	0.52	0.89	7.00	0.40
L Knee Angle IC	0.84	10.30	1.16	0.92	10.60	1.16	0.79	10.75	1.73	0.81	10.25	1.34	0.78	9.85	1.67	0.90	10.80	1.11
L knee Angle MS	0.75	37.90	3.07	0.97	37.90	0.96	0.99	38.45	0.56	0.98	36.85	0.85	0.99	38.05	0.68	0.97	38.80	1.07
L Heel to COM	0.89	14.25	0.91	0.92	13.71	0.83	0.92	12.40	0.83	0.92	13.46	0.95	0.79	14.72	1.46	0.93	15.31	0.84
L Knee Angle MSw	0.92	92.90	3.09	0.98	92.50	1.27	0.99	93.00	1.04	0.97	91.35	1.91	0.92	92.00	2.85	0.99	93.00	0.88
L COM Height MS	0.97	97.15	1.21	0.90	93.84	2.35	0.92	92.89	2.11	0.86	96.14	2.77	0.97	96.26	1.28	0.97	93.96	1.31
L COM Height DF	0.89	104.46	2.46	0.89	101.04	2.68	0.90	100.29	2.49	0.82	103.42	3.46	0.96	103.67	1.58	0.97	102.07	1.33
R Shoe Angle	0.98	4.45	1.71	0.98	4.60	1.19	0.95	6.50	2.79	0.99	4.20	1.15	0.95	5.30	2.48	0.99	5.40	1.11
R Tibial Angle	0.91	4.85	0.72	0.88	4.35	0.66	0.86	4.95	0.69	0.92	4.75	0.89	0.62	4.95	0.46	0.85	5.60	0.34
R Knee Angle IC	0.93	11.95	0.70	0.91	11.80	0.78	0.68	12.15	1.14	0.54	11.65	1.76	0.86	12.40	0.96	0.75	11.05	1.10
R knee Angle MS	0.92	38.40	1.46	0.92	38.00	1.38	0.96	38.55	0.95	0.97	38.30	0.92	0.95	38.90	1.16	0.97	38.20	0.92
R Heel to COM	0.85	12.83	1.47	0.93	11.07	1.02	0.97	11.99	0.75	0.89	12.58	1.45	0.92	13.33	1.12	0.95	12.54	0.86
R Knee Angle MSw	0.85	97.15	1.21	0.92	93.84	2.35	0.99	92.89	2.11	0.99	96.14	2.77	0.92	96.26	1.28	0.98	93.96	1.31
R COM Height MS	0.93	98.53	1.82	0.94	95.67	1.56	0.96	93.41	1.54	0.94	97.75	1.68	0.98	96.49	1.12	0.94	94.26	1.78
R COM Height DF	0.92	105.74	2.09	0.95	103.10	1.58	0.92	100.73	2.10	0.94	105.11	1.78	0.97	104.18	1.40	0.92	102.07	2.14

IC = initial contact; MS = midstance; COM = center of mass; MSw = midswing; DF = double leg flight

Table 3. Intra-rater Reliability of Frontal Plane Variables (ICC3,1), Mean (2 measures), SEM (standard error of measurement)

	Rater 1 (12 years experience)			Rater 2 (10 years experience)			Rater 3 (0 years experience)			Rater 4 (0 years experience)			Rater 5 (0 years experience)			Rater 6 (0 years experience)		
	ICC	Mean (°)	SEM (°)	ICC	Mean (°)	SEM (°)	ICC	Mean (°)	SEM (°)	ICC	Mean (°)	SEM (°)	ICC	Mean (°)	SEM (°)	ICC	Mean (°)	SEM (°)
L Hip ADD IC	0.81	4.00	1.09	0.96	3.65	0.57	0.84	3.55	0.82	0.62	3.30	1.45	0.94	3.75	0.51	0.77	4.95	1.04
L Hip ADD MS	0.77	9.80	1.37	0.74	9.25	1.90	0.81	8.95	1.19	0.88	9.45	1.07	0.92	9.00	0.97	0.88	9.15	0.92
L RF Angle IC	0.81	7.40	3.21	0.96	6.55	1.59	0.97	5.85	1.17	0.97	-5.75	1.02	0.97	-5.45	1.30	0.97	-5.50	1.39
L RF Angle MS	0.80	9.10	2.19	0.95	4.65	1.72	0.88	6.75	1.70	0.85	9.55	2.76	0.73	8.25	3.05	0.90	8.80	2.08
R Hip ADD IC	0.84	5.55	0.86	0.45	5.80	1.56	0.08	4.90	1.90	0.55	6.45	1.89	0.78	6.45	1.19	0.63	5.90	1.07
R Hip ADD MS	0.78	11.75	1.21	0.89	12.10	0.72	0.73	10.65	1.54	0.80	12.30	1.31	0.70	11.40	1.46	0.88	10.90	0.75
R RF Angle IC	0.92	3.35	1.99	0.87	4.15	2.64	0.77	1.95	3.08	0.88	3.40	2.06	0.94	2.60	2.01	0.91	4.00	2.05
R RF Angle MS	0.86	13.05	1.31	0.67	11.90	1.67	0.68	11.40	1.23	0.64	12.90	2.31	0.57	12.50	2.14	0.53	11.45	2.78

ADD = adduction; IC = initial contact; MS = midstance

mean with only 10% of SAG SEM values above 20% of the variable mean.

The inter-rater ICC value between the two experienced raters were all greater than 0.88 except for right knee flexion at initial contact which was 0.76. The inter-rater ICC value between the four

inexperienced raters were all greater than 0.92. The inter-rater reliability between experienced and inexperienced raters ranged between 0.97 and 0.99.

Frontal Plane Variables. The intra-rater ICC values for all raters ranged from 0.08 to 0.97 with SEM values ranging from 6% to 158% of the mean value for all variables assessed. For the *experienced raters*, the intra-rater ICC values ranged from 0.45 to 0.96 with the SEM values ranging from 6% to 64% of the mean values for all variables. For the *inexperienced raters*, the intra-rater ICC values ranged from 0.08 to 0.97 with the SEM values ranging from 7% to 158% of the mean values for all variables. For all raters, 48% of all FRT SEM values were less than 20% of the variable mean and 52% of FRT SEM values were above 20% of the variable mean.

The inter-rater ICC for all variables ranged from 0.84 to 0.98 for the *experienced raters*. The inter-rater ICC value between the four *inexperienced raters* ranged from 0.79 to 0.98. The inter-rater reliability between the *experienced* and *inexperienced* raters ranged between 0.86 and 0.99.

DISCUSSION

The intent of this study was to assess the intra-rater and inter-rater reliability of six different raters, two experienced and four inexperienced, in assessing various sagittal and frontal plane kinematic variables recorded while running on a treadmill using 2D video analysis. Two-dimensional video running analysis is commonly performed in the clinical setting to assess potential kinematic abnormalities

Table 4. Inter-rater Intraclass Correlation Coefficients (ICC3,k) with 95% Confidence Interval (CI)

	Experienced		Inexperienced		Experienced vs. Inexperienced	
	ICC	95%CI	ICC	95%CI	ICC	95%CI
Sagittal	L Shoe Ang	0.99 (.95-1.0)	0.99 (.97-1.0)	1.00 (.99-1.0)		
	R Shoe Ang	0.98 (.92-1.0)	1.00 (.99-1.0)	1.00 (.98-1.0)		
	L Tib Ang	0.92 (.69-.98)	0.95 (.87-.99)	0.99 (.95-1.0)		
	R Tib Ang	0.88 (.53-.97)	0.95 (.88-.99)	0.97 (.77-.99)		
	L Knee Ang IC	0.88 (.53-.97)	0.96 (.89-.99)	0.98 (.94-1.0)		
	R Knee Ang IC	0.76 (.00-.94)	0.92 (.80-.98)	0.98 (.90-.99)		
	L Knee Ang MS	0.97 (.88-.99)	0.98 (.95-1.0)	0.98 (.93-1.0)		
	R Knee Ang MS	0.99 (.96-1.0)	0.99 (.98-1.0)	0.99 (.97-1.0)		
	L Heel to COM	0.96 (.84-.99)	0.93 (.71-.98)	0.99 (.94-1.0)		
	R Heel to COM	0.97 (.60-.99)	0.97 (.93-.99)	0.96 (.83-.99)		
	L Knee Ang MSw	0.95 (.82-.99)	0.99 (.97-1.0)	0.99 (.95-1.0)		
	R Knee Ang MSw	0.96 (.82-.99)	1.00 (.99-1.0)	0.99 (.96-1.0)		
	L COM Hgt MS	0.94 (.00-.99)	0.97 (.91-.99)	0.99 (.96-1.0)		
	R COM Hgt MS	0.93 (.25-.99)	0.97 (.88-.99)	0.98 (.38-1.0)		
	L COM Hgt DF	0.94 (.00-.99)	0.98 (.93-.99)	0.99 (.98-1.0)		
	R COM Hgt DF	0.93 (.49-.99)	0.97 (.89-.99)	0.99 (.57-1.0)		
Frontal	L Hip Add IC	0.85 (.39-.96)	0.93 (.79-.90)	0.97 (.88-.99)		
	R Hip Add IC	0.87 (.47-.97)	0.83 (.57-.95)	0.96 (.84-.99)		
	L Hip Add MS	0.94 (.76-.98)	0.98 (.94-.99)	0.99 (.93-1.0)		
	R Hip Add MS	0.93 (.73-.98)	0.95 (.85-.99)	0.96 (.74-.99)		
	L RF IC	0.82 (.35-.96)	0.98 (.94-.99)	0.95 (.82-.99)		
	R RF IC	0.98 (.93-.00)	0.95 (.88-.99)	0.98 (.93-1.0)		
	L RF MS	0.92 (.51-.98)	0.96 (.89-.99)	0.91 (.67-.98)		
	R RF MS	0.84 (.39-.96)	0.79 (.45-.94)	0.87 (.47-.97)		

Ang=angle, Tib=tibia, IC=initial contact, MS=midstance, MSw=Midswing, COM=center of mass, Hgt=height, Add=adduction, RF=Rearfoot

---

when screening for injury prevention or when documenting changes in the running kinematics because of rehabilitation following injury. Since it is common for a video analysis to be conducted on more than one occasion to assess whether changes in running kinematics have occurred because of an injury or a modification of a runner's gait pattern, the consistency of qualitative measurements assessed on different days is important. In addition, the ability of multiple clinicians with varying levels of clinical experience to reliably perform an assessment of quantitative measurements in both the sagittal and frontal planes has not been previously investigated.

Interpretation of the Reliability of Sagittal and Frontal Plane Variables

In general, all raters consistently measured each SAG variable between the two sessions, one week apart with high levels of reliability ($ICC \geq 0.75$). Based on the classification scheme proposed by Landis and Koch, all the intra-rater ICC variables for the two experienced raters would be classified as *almost perfect* except for Rater 1's ICC for left knee flexion at midstance) which would be classified as *substantial*. For the inexperienced raters, based on the Landis and Koch¹³ classification scheme, the intra-rater ICC for 12 of the 16 variables would be classified as *almost perfect*. Of the remaining four variables, left tibial angle, left knee angle at initial contact, and right tibial angle would be classified as *substantial*. The only intra-rater ICC for the inexperienced raters classified as *moderate* was right knee angle at initial contact (Rater 4). The results for the inter-rater reliability were quite similar in that the ICC values for the two experienced raters were all classified as *almost perfect* except for right knee flexion at initial contact which would be classified as *substantial*. The inter-rater ICC values between the four inexperienced raters as well as between the experienced versus inexperienced raters would all be classified as *almost perfect*.

Based on the ICC values, one could conclude that the intra- and inter-rater reliability among all six raters was very high. However, it is important to take into consideration the SEM values for all raters. For the experienced raters, except for the left and right shoe angle, the SEM values for all other SAG variables

were less than or equal to 15% of the mean value. This would indicate that if a variable mean value was 60°, the degree of variability associated with the measurement by a rater would range between 51° and 69°. For the inexperienced raters, apart from the left and right shoe angle, the SEM values for all other SAG variables were less than or equal to 20% of the mean value. This would indicate that if a variable mean value was 60°, the degree of variability associated with the measurement by a rater would range between 48° and 72°. The only other study that has attempted to evaluate the reliability of quantitative kinematic measures in running was conducted by Damsted et al.⁶ They assessed the ability of two raters to quantify left knee and hip flexion angles at initial contact during running. They reported that 95% confidence limits for knee flexion ranged from three to 8° within a day and from 9° to 14° between days for the two raters. In the current study, the left mean knee flexion angle at initial contact (averaged across all 6 raters) was 10.43° with the average SEM for all raters 1.36°. Thus, the range for left knee flexion in the current study would be 9.07° to 11.79° which is similar to the values reported by Damsted et al.⁶

In general, ICC values were lower for the eight FRT variables. For the experienced raters, based on the classification Landis and Koch scheme,¹³ six of the variables for rater 1 and five of the variables for rater 2 would be classified as *almost perfect* based on the intra-rater ICC values. The intra-rater ICC values for left hip adduction angle at midstance (Rater 1 and 2), right hip adduction at midstance (Rater 1), and right rearfoot angle at midstance (Rater 2) were classified as *substantial*, with right hip adduction at initial contact (Rater 2) classified as *moderate*. Using the Landis and Koch classification scheme, the intra-rater ICC values for only two of the eight variables, left hip adduction at midstance and left rearfoot angle at initial contact would be classified as *almost perfect* for all four inexperienced raters. For the other six variables, the ICC values would be classified as either *substantial* or *moderate*.

The inter-rater ICC for all values was *almost perfect* between the two experienced and four inexperienced raters except for the right rearfoot angle at midstance which was *substantial*. The inter-rater reliability comparing the experienced and

---

inexperienced raters was *substantial*. These findings disagree with Maykut et al¹ who reported ICC intra-rater reliability values for hip adduction to be 0.95 for the right and 0.96 for the left using a single rater. Unfortunately, these authors do not provide any information regarding the level of experience of the rater. In the current study, the ICC values for hip adduction angles at initial contact and at midstance ranged from 0.45 to 0.89 for the experienced raters, and 0.08 to 0.92 for the inexperienced raters.

The SEM values for the frontal plane variables were considerably higher in comparison to the SAG variables ranging from 6% to 64% for the experienced raters and from 7% to 158% for the inexperienced raters (Table 3). Based on these findings, the authors confirmed the first hypothesis which stated quantitative measurements for SAG kinematics would have higher levels of reliability, irrespective of rater experience, than quantitative measurements for FRT kinematics.

Influence of Rater Experience

As noted in the introduction, the influence of rater experience on the reliability of quantitative measurements for SAG and FRT variables during running has not been previously investigated. In the current study, two raters had over 10 years' experience performing running analyses while the other four raters had no previous experience except for a two-hour training session followed by one practice analysis. Based on intra-rater and inter-rater ICC results for the SAG variables, there was minimal differences between the experienced and inexperienced raters for intra-rater and inter-rater reliability. As noted in the results, all SAG variables demonstrated ICC values greater than 0.75 for the experienced raters with only one variable (right knee angle at initial contact) having an ICC value less than 0.75 for the inexperienced raters. Regarding the FRT variables, both the experienced and inexperienced raters demonstrated lower levels of intra-rater reliability as compared to the SAG variables, with several variables assessed demonstrating ICC values less than 0.75 (three variables for the experienced raters and no variables for the inexperienced raters). These findings would suggest that repeated FRT measurements by the same rater, whether experienced or inexperienced, would

be suspect to variation when performed over multiple days. Based on these results, the authors rejected the second study hypothesis which stated that experienced raters would demonstrate higher levels of reliability than inexperienced raters when assessing both sagittal and FRT kinematics. It would appear based on the current study findings, that experience does not appear to be a factor when consistency is required for repeated analyses on the same runner. In addition, the findings would suggest that the same clinician can consistently repeat a quantitative running analysis using the sagittal variables assessed in this study irrespective of experience, assuming the level of training used in the current study. When performing an analysis on FRT variables, these results indicate that the same clinician should always perform the analysis, and to be aware of the possibility of greater measurement variability.

It is important to note that the findings of this study are limited to the eight SAG and four FRT kinematic variables assessed in this study. Thus, caution should be used when attempting to generalize these findings to other conditions or kinematic measures obtained using 2D video analysis. Additionally, the reliability in this study may be greater than in clinical running analyses as a result of the use of pre-placed reflective markers. These markers aided in the angle measurements during the running analysis. The methodology of this study, however, has addressed previous limitations noted in the literature⁹ by having both experienced and inexperienced raters perform a running analysis by reviewing the recorded video in slow-motion or frame-by-frame to select the specific points in the running cycle for analysis rather than using pre-identified still frames. Future research is recommended to compare the reliability of qualitative and quantitative running assessment, and the clinical usefulness of each in making clinical decisions.

CONCLUSION

This study is one of the first to assess the reliability of experienced and inexperienced raters assessing quantitative SAG and FRT variables while running on a treadmill using 2D video analysis. The intra-rater and inter-rater reliability levels were higher for SAG variables assessed in this study in comparison

to FRT variables. Experience did not appear to be a factor when consistency is required with repeated analyses on the same runner. The results of the study indicate that the same clinician can consistently repeat a quantitative running analysis using the SAG variables assessed in this study irrespective of experience. Because of poor reliability, the same clinician should always assess FRT variables and be cognizant of greater measurement variability.

REFERENCES

1. Maykut JN, Taylor-Haas JA, Paterno MV, DiCesare CA, Ford KR. Concurrent validity and reliability of 2d kinematic analysis of frontal plane motion during running. *Int J Sports Phys Ther.* 2015;10(2):136-146.
2. Neal BS, Barton CJ, Gallie R, O'Halloran P, Morrissey D. Runners with patellofemoral pain have altered biomechanics which targeted interventions can modify: A systematic review and meta-analysis. *Gait Posture.* 2016;45:69-82.
3. Pohl MB, Mullineaux DR, Milner CE, Hamill J, Davis IS. Biomechanical predictors of retrospective tibial stress fractures in runners. *J Biomech.* 2008;41(6):1160-1165.
4. Wille CM, Lenhart RL, Wang S, Thelen DG, Heiderscheit BC. Ability of sagittal kinematic variables to estimate ground reaction forces and joint kinetics in running. *J Orthop Sports Phys Ther.* 2014;44(10):825-830.
5. Schurr SA, Marshall AN, Resch JE, Saliba SA. Two-Dimensional Video Analysis Is Comparable to 3d Motion Capture in Lower Extremity Movement Assessment. *Int J Sports Phys Ther.* 2017;12(2):163-172.
6. Damsted C, Nielsen RO, Larsen LH. Reliability of video-based quantification of the knee- and hip angle at foot strike during running. *Int J Sports Phys Ther.* 2015;10(2):147-154.
7. Krummen K, Kelly S, Briggs M. Reliability of lower extremity 2-dimensional video running analysis (Abstract). *J Orthop Sports Phys Ther.* 2016;46:A42.
8. Souza RB. An Evidence-Based Videotaped Running Biomechanics Analysis. *Phys Med Rehabil Clin N Am.* 2016;27(1):217-236.
9. Pipkin A, Kotecki K, Hetzel S, Heiderscheit B. Reliability of a Qualitative Video Analysis for Running. *J Orthop Sports Phys Ther.* 2016;46(7):556-561.
10. Dierks TA, Manal KT, Hamill J, Davis IS. Proximal and distal influences on hip and knee kinematics in runners with patellofemoral pain during a prolonged run. *J Orthop Sports Phys Ther.* 2008;38(8):448-456.
11. Alenezi F, Herrington L, Jones P, Jones R. How reliable are lower limb biomechanical variables during running and cutting tasks. *J Electromyogr Kinesiol.* 2016;30:137-142.
12. Doma K, Deakin GB, Sealey RM. The reliability of lower extremity and thoracic kinematics at various running speeds. *Int J Sports Med.* 2012;33(5):364-369.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
14. Rothstein JM. *Measurement in physical therapy.* New York: Churchill Livingstone; 1985.