Test-Retest Reliability and Inter-Tester Reliability of Kinematic Data from a Three-Dimensional Gait Analysis System

Hitoshi Tsushima, Meg E. Morris and Jennifer McGinley

Abstract. This study aimed to determine the test-retest reliability and inter-tester reliability of kinematic measures in a three-dimensional gait analysis system. Using a VICON 140® three-dimensional motion analysis system, kinematic data for lower extremities during walking were collected by 2 testers (senior physical therapists) for 6 unimpaired adults (age = 20 to 52; mean = 35.2 ± 6.2). The study was conducted using a repeated measures design consisting of two testing sessions per day on two separate testing days. The reliability of joint angle data collected by two different testers on two different days was compared for 2 sessions (days) × 2 testers × 5 trials. Skin markers were placed on 15 defined pelvis and lower body locations in accordance with the VICON Clinical Manager model. Prior to the study commencing, the two physical therapists practiced marker placement for a 3 month period. The first measurements (T1) were carried out by two testers on the same day. The second measurement session (T2) was performed within two weeks using an identical procedure. Coefficients of multiple correlation (CMC) were calculated to evaluate the consistency between the kinematic variables across testers and sessions. Both test-retest and inter-tester reliability were high for motion in the sagittal plane (Ra = 0.971 to 0.994), the frontal plane (Ra = 0.759 to 0.977), and the transverse plane (Ra = 0.729 to 0.899), excluding pelvic tilt. Reduction of variability of marker placement appears possible with standardization and understanding of the placement method. These findings provide evidence of the reliability of using three-dimensional motion analysis for measuring human gait.

Key words: gait analysis, 3D-motion analysis system, reliability of kinematic data, marker placement, physical therapy

Over the last 30 years, many investigations have been conducted with normal subjects, describing the kinematics of gait. Recently, the study of human motion has rapidly advanced, due to the development of three-dimensional motion analysis systems. With these systems, the movement of each segment of the human body can be recorded and analyzed in the sagittal, frontal and transverse planes. Gait analysis using these systems has been recognized as a valuable tool in the assessment of gait disturbances and objective evaluation of treatment. However, the reliability of this form of gait analysis requires further investigation.

Reliability refers to the repeatability or reproducibility of measurement procedures. There are two major factors that can influence the reproducibility of measurements in gait analysis. The first is the variability in people that can occur during repeated gait performances. The second relates to errors that may occur as part of the measurement process. Measurement error can occur from many sources, such as the localisation of bony landmarks, placement of markers, movement of markers with movement of the skin,
and accuracy of the motion analysis system itself.

While the accuracy of three-dimensional motion analysis systems has improved with technological innovation\(^5\)\(^-\)\(^7\), there is limited research available on the reliability of data produced from these tools\(^8\)\(^-\)\(^11\). Winter\(^8\) reported intrasubject repeatability of kinematic data of normal subjects, utilizing the coefficient of variation (CV). The results showed that intrasubject repeatability was better on the same day of testing compared with results from different days. Kadaba and co-authors\(^10\) presented a comprehensive evaluation of intrasubject repeatability of gait based on repeated analyses using 40 unimpaired adult subjects. To quantify the repeatability, the adjusted coefficient of multiple correlation was used as a statistical measure to evaluate the similarity of kinematic waveforms. Results indicated that sagittal plane kinematic data provided high intrasubject repeatability both on the same test day and between test days. On the contrary, frontal and transverse plane data measured over two consecutive days had poor repeatability, especially for pelvic tilt. This was attributed by the authors as partly due to error in marker alignment. Although the study noted consistency of gait measurement on the same test day, further research was necessary in order to determine the repeatability of marker placement between test occasions. A similar study on intrasubject repeatability was conducted by Steinwender et al.\(^11\) using unimpaired and spastic children. They also suggested that errors due to marker placement contributed to lower between-day repeatability.

In summary, limited research on the reliability of three-dimensional gait analysis has reported results within one day and between days. The reliability between days was not high, while a high level of reliability was apparent for analysis on the same day.

Although these findings imply that error of marker replacement influences the reliability of gait analysis, this type of error has infrequently been quantified. When measuring changes over time in gait using this equipment, marker replacement cannot be avoided. Therefore, it is very important to quantify the measurement error caused by marker replacement. It is difficult to separate the variability of performances and measurement error in gait analysis completely. Accordingly, we were interested in the reproducibility of kinematic data from repeated tests, over multiple sessions. Highly trained physical therapists were used to collect the data in order to minimise error associated with maker placement. The purpose of this study was to further assess the test-retest and inter-tester reliability of kinematic data in three-dimensional gait analysis from the viewpoint of marker placement.

### Method

#### Subjects

Six unimpaired persons (3 males and 3 females) participated in this study. Their mean age, body height and body weight were 34.8 ± 12.5 years, 173.3 ± 9.8 cm and 68.5 ± 9.6 kg respectively. None of the subjects were obese, and none had a history of significant musculoskeletal, neurological, cardiovascular or psychological disease. In addition, they all had normal strength and range of motion in the lower extremities and were willing and able to provide informed consent. Two senior physical therapists experienced with the marker placement procedures also participated in this research as testers.

#### Research design and instrumentation

The study was conducted using a repeated measures design consisting of two testing sessions per day on two separate testing days. The reliability of joint angle data collected by two different testers on two different days was compared for 2 sessions (days) × 2 testers × 5 trials.

In this study, the ‘intra-subject’ reliability refers to the repeatability of performance over 5 trials within each test session without removal of markers. The ‘test-retest’ reliability refers to the repeatability from one test session to the next with reattached markers. ‘Inter-tester’ reliability refers to the repeatability between different tester’s sessions in the same day with reattached markers.

A VICON 140\(^{TM}\) three-dimensional motion analysis system (Oxford Metrics Ltd., Oxford, UK) and VICON Clinical Manager\(^{TM}\) (VCM) software were used to capture and analyze the walking patterns of the subjects. The system utilized 4 infrared cameras located at angles around a 10-m walkway. These cameras were placed on diagonal lines intersecting at the center of the walkway (Fig. 1). Each walking trial was videotaped by a standard video camera located towards the center of the walkway to assist with analysis of the data.

All tests in this study were carried out at the Gait Laboratory of the Geriatric Research Unit at the Kingston Centre with the approval of the Kingston Centre Research and Ethics Committee, Southern Health Network, Victoria, Australia.

#### Procedures

Subjects were required to wear shorts and to walk without shoes. Spherical, passive reflective markers (25 mm diameter) were attached directly to the skin of each subject’s bilateral lower extremities using double-sided adhesive tape. All markers were placed in accordance with the VCM model on 15 points of bony landmarks, including the sacrum at the level of S2, bilateral anterior superior iliac spines (ASIS), lateral thighs, the axis of the knee joints, the...
shanks, lateral malleoli, heels, and the second metatarsal heads (Fig. 2). The ASIS and sacral markers were also placed directly on the skin, not on the shorts. After recording anthropometric measures such as leg length and pelvic width, markers were placed by Tester 1 following the defined protocol.

The subjects were instructed to walk at self-selected speed along a 10-m walkway and they practiced walking in order to get familiar with the surroundings. Then the movement of each joint was captured as the subject walked on the walkway at his or her own pace 5 times. A single gait cycle of kinematic data was recorded for each trial. All markers were then removed. After 30 minutes, tester 1 checked that all traces of the marker placement on the subject’s skin had disappeared. Tester 2 then repeated the placement of markers and data collection following identical procedures. Subsequent test sessions were conducted in an identical manner for the same subjects on a second day within 2 weeks of the first session. Each session was videotaped, to assist with interpretation of the data. Prior to the study commencing, the two physical therapists practiced marker placement for a 3 month period.

When gait data is processed in VCM, a static trial that calculates the actual joint center and alignment of the axis prior to the dynamic trial, which calculates the gait variables. In the static trial, the actual joint centers and axis of the lower limb joints are determined according to the anthropometric measures and the positional data of each marker and the Knee Alignment Device (KAD). The KAD is a spring-loaded G-shaped clamp and has three standard-size fixed markers mounted on the tip of rods of equal length placed at right angles in three-dimensional directions. The kinematic data processed by VCM is shown as the angle between two segment axes projected into the related plane.

Data analysis

Walking speed, cadence and stride length were processed by VCM. The mean and standard deviation of each spatiotemporal parameter were obtained for each test session. For each subject, the coefficient of variation (CV) was calculated as a measure of the reproducibility within a session and between sessions. For CV within a session, the mean and the standard deviation of the gait parameters from 5 trials were used, while for CV between sessions, those of the gait parameters from all 20 trials were used. The coefficient of variation was averaged over the number of subjects for each parameter. The mean and standard deviation of within-session and between-session CVs were also computed.

Although one frequently used estimate of reliability is the intraclass correlation coefficient (ICC), using the ICC as a statistical method to clarify reproducibility and similarity for kinematic variables over the whole gait cycle is difficult because those parameters are expressed as waveforms. It was thus considered appropriate to use the adjusted coefficient of multiple correlation (CMC) recommended by Kadaba et al.\(^\text{10}\) to evaluate the similarity of the wave-form parameters over the whole gait cycle. Firstly, the qualitative similarity between waveforms was evaluated visually. Next, the output data from the VCM was processed with Excel, and the CMC was calculated using formula described by Kadaba et al.\(^\text{10}\) (Table 1).

In Table 1, Eq.(1) shows the CMC for evaluation of the repeatability of waveforms within a test session, while Eq.(4) shows the CMC for evaluation of the similarity of waveforms between test sessions. When the waveforms are similar, \(R_a\) in Eq.(1) and (4) is nearly equal to 1. If the
Table 1. Correlation of multiple coefficient (modified from Kadaba, et al.\textsuperscript{11}) reproduced with copyright

<table>
<thead>
<tr>
<th>Subjects</th>
<th>CMC formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of the parameters</td>
<td></td>
</tr>
</tbody>
</table>
\[ t: \text{time sequence} \rightarrow T \]
\[ i: \text{number of sessions} \rightarrow M \]
\[ j: \text{number of trials} \rightarrow N \]

\[ Y_{ij} \text{is the } j\text{th time point of the } i\text{th run on the } i\text{th test session.} \]

CMC for assessing repeatability within a test session

\[ R_A = 1 - \frac{\sum \sum \sum_{i=1, j=1, t=1}^{M, N, T} (Y_{ij} - \bar{Y_i})^2 / MT(N-1)}{\sum \sum \sum_{i=1, j=1, t=1}^{M, N, T} (Y_{ij} - \bar{Y})^2 / MNT - 1} \] ...Eq.(1)

\[ \bar{Y_i} \text{is average at time point } t \text{ on the } i\text{th test session.} \]
\[ \bar{Y} \text{is the grand mean on the } i\text{th session} \]

\[ \bar{Y_i} = \frac{1}{N} \sum_{j=1}^{N} Y_{ij} \] ...Eq.(2)

\[ \bar{Y} = \frac{1}{NT} \sum_{i=1}^{M} \sum_{t=1}^{T} Y_{ij} \] ...Eq.(3)

CMC for assessing repeatability between test sessions

\[ R_A = 1 - \frac{\sum \sum \sum_{i=1, j=1, t=1}^{M, N, T} (Y_{ij} - \bar{Y_i})^2 / T(MN-1)}{\sum \sum \sum_{i=1, j=1, t=1}^{M, N, T} (Y_{ij} - \bar{Y})^2 / MN(T-1)} \] ...Eq.(4)

\[ \bar{Y_i} \text{is average at time point } t \text{ over } M \text{ } N \text{ gait cycle} \]
\[ \bar{Y} \text{is the grand mean over time} \]

\[ \bar{Y_i} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} Y_{ij} \] ...Eq.(5)

\[ \bar{Y} = \frac{1}{MNT} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{t=1}^{T} Y_{ij} \] ...Eq.(6)

Table 2. Means and standard deviations and coefficient of variation of the spatiotemporal parameters for the subjects

<table>
<thead>
<tr>
<th>Gait parameter</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
<th>Session 4</th>
<th>Average</th>
<th>Within session CV</th>
<th>Between session CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadence (steps/min)</td>
<td>121.61 ± 5.21</td>
<td>122.93 ± 5.14</td>
<td>121.63 ± 7.06</td>
<td>120.91 ± 5.24</td>
<td>121.77 ± 5.46</td>
<td>1.76 ± 0.71 (%)</td>
<td>2.19 ± 0.89 (%)</td>
</tr>
<tr>
<td>Walking speed (m/s)</td>
<td>1.48 ± 0.04</td>
<td>1.50 ± 0.09</td>
<td>1.48 ± 0.06</td>
<td>1.47 ± 0.06</td>
<td>1.46 ± 0.06</td>
<td>2.42 ± 1.04 (%)</td>
<td>3.76 ± 1.44 (%)</td>
</tr>
<tr>
<td>Stride length (m)</td>
<td>1.43 ± 0.11</td>
<td>1.43 ± 0.13</td>
<td>1.43 ± 0.09</td>
<td>1.43 ± 0.10</td>
<td>1.43 ± 0.11</td>
<td>3.62 ± 4.39 (%)</td>
<td>4.25 ± 4.25 (%)</td>
</tr>
</tbody>
</table>

Session 1: Day1-Tester1, Session 2: Day1-Tester2, Session3: Day2-Tester1, Session 4: Day2-Tester2.

(Mean ± SD)

waveform is dissimilar, \( R_a \) is nearly equal to zero. Eq.(1) was used to evaluate the repeatability of waveforms within each test session as intra-subject reliability. Test-retest and inter-tester reliability were evaluated using Eq.(4). That is to say, Eq.(4) was used for the repeatability between different testers’ test sessions in same day as well as the repeatability between same tester’s sessions in different test day.
Results

Repeatability of spatiotemporal parameters

The mean and standard deviation and the coefficients of variation (CV) for each spatiotemporal parameter are shown in Table 2. The CVs of all parameters were low values of less than 5%. The CV for cadence was lower than that of walking speed and stride length. Repeatability of spatiotemporal parameters within a session was higher than between sessions.

Qualitative analysis based on graphic waveforms

Figure 3 shows the kinematic variables for 5 walking trials for a unilateral (left side) lower limb in a representative subject. In the graphs, one sees that the similarity of the curve is very high when comparing the 5 trials, suggesting that the reproducibility of each joint angle change for the subject was high.
Figure 4 provides a comparison of the mean results for four test sessions (Day 1 by tester 1, Day 1 by tester 2, Day 2 by tester 1, and Day 2 by tester 2) for the same subject who was presented in Fig. 3. The waveforms of kinematic data (averaged for five trials) for the four sessions are shown together in one graph. Although the variability of the movements in the transverse plane is somewhat larger than that of the movement in other planes, the waveforms for any joint movement are generally convergent and highly similar. Therefore, test-retest and inter-tester reliability appear to be high.

Quantitative analysis by the coefficient of multiple correlation

The CMC was calculated to objectively evaluate the reproducibility for each joint. The CMC intra-subject, test-
retest and inter-tester, means and standard deviations in each joint movement for all subjects are shown in Table 3 respectively. These values were averaged from the CMCs that were calculated for every subject. The CMC for each movement ranged from 0.729 to 0.997 with respect to the intra-subject, test-retest and inter-tester reproducibilities except for the pelvic tilt. The CMC for pelvic tilt ranged from only 0.349 to 0.536 with respect to intra-subject, test-retest and inter-tester reproducibilities.

As expected, the CMCs for intra-subject reproducibility were higher than for test-retest and inter-tester reproducibilities. The CMC exceeded 0.80 for all movements except pelvic tilt. For intra-subject reproducibility, the CMC for hip flexion in the sagittal plane was the highest value ($R_{a} = 0.996$ to 0.997). In contrast, the CMCs for knee rotation and foot rotation in the transverse plane were relatively low ($R_{a} = 0.729$ to 0.799).

The CMC values for test-retest and inter-tester reproducibilities were rather similar, with ranges of 0.747 to 0.994 and 0.729 to 0.992 respectively. Although these CMCs for test-retest and inter-tester reproducibilities were relatively high, a hierarchy of reproducibility due to the

### Table 3. Means and standard deviations of the coefficient of multiple correlation (CMC) for the joint angle motion of the subjects

<table>
<thead>
<tr>
<th>Joint angle motion</th>
<th>Intra-subject</th>
<th>Test-retest*</th>
<th>Inter-tester**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left</td>
<td>Right</td>
<td>L-R Average</td>
</tr>
<tr>
<td><strong>Sagittal plane</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic tilt</td>
<td>0.536 ± 0.124</td>
<td>0.514 ± 0.064</td>
<td>0.525 ± 0.095</td>
</tr>
<tr>
<td>Hip flexion/extension</td>
<td>0.996 ± 0.001</td>
<td>0.997 ± 0.001</td>
<td>0.997 ± 0.001</td>
</tr>
<tr>
<td>Knee flexion/extension</td>
<td>0.994 ± 0.002</td>
<td>0.995 ± 0.001</td>
<td>0.994 ± 0.002</td>
</tr>
<tr>
<td>Ankle dorsiflexion/plantar</td>
<td>0.981 ± 0.004</td>
<td>0.981 ± 0.005</td>
<td>0.981 ± 0.005</td>
</tr>
<tr>
<td><strong>Frontal plane</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic obliquity</td>
<td>0.986 ± 0.004</td>
<td>0.982 ± 0.009</td>
<td>0.984 ± 0.007</td>
</tr>
<tr>
<td>Hip abduction/adduction</td>
<td>0.982 ± 0.005</td>
<td>0.979 ± 0.007</td>
<td>0.980 ± 0.006</td>
</tr>
<tr>
<td>Knee varus/valgus</td>
<td>0.925 ± 0.023</td>
<td>0.926 ± 0.015</td>
<td>0.925 ± 0.019</td>
</tr>
<tr>
<td><strong>Transverse plane</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic rotation</td>
<td>0.926 ± 0.079</td>
<td>0.914 ± 0.101</td>
<td>0.920 ± 0.087</td>
</tr>
<tr>
<td>Hip rotation</td>
<td>0.938 ± 0.027</td>
<td>0.917 ± 0.032</td>
<td>0.927 ± 0.030</td>
</tr>
<tr>
<td>Knee rotation</td>
<td>0.804 ± 0.106</td>
<td>0.910 ± 0.059</td>
<td>0.857 ± 0.099</td>
</tr>
<tr>
<td>Foot rotation</td>
<td>0.858 ± 0.059</td>
<td>0.910 ± 0.058</td>
<td>0.884 ± 0.062</td>
</tr>
</tbody>
</table>

* ‘Test-retest’: the average included both testers’ sessions.
** ‘Inter-tester’: the average included the Day 1 session and the Day 2 session
plane of movement was confirmed. Movement in the sagittal plane produced the highest CMC ($R_s=0.971$ to 0.994), and movement in the frontal plane ($R_s=0.759$ to 0.977) and transverse plane ($R_s=0.729$ to 0.899) followed.

**Discussion**

The reliability of gait analysis is an essential consideration in both clinical and research utilization of quantitative gait analysis data. The factors contributing to variability in gait analysis results can be summarized as:
1) limitations in the accuracy of the measurement hardware and software,
2) variability of walking across trials,
3) reproducibility of the marker placement across sessions (test-retest) and across testers (inter-tester),
4) artifacts originating from skin-marker movement.

Prior research into the reproducibility of marker replacement included only a few reports on test-retest reliability in three-dimensional motion analysis, except by Kadaba et al.\(^{10}\) and Steinwender et al.\(^{11}\). There are even fewer reports on inter-tester reliability. Unreliability can be due in part to the tester’s placement of markers, resulted in difficulty in accurately specifying bony landmarks, inaccuracy of marker installation due to rough estimates, and errors in marker configuration with repeated measurements\(^{13-15}\).

From the results of the present study, the CVs of spatiotemporal parameters in this study were very low within a single session. In fact, the walking performances of the subjects were highly consistent. As expected, the ‘intra-subject’ CMC values were higher than the ‘test-retest’ and ‘inter-tester’ values. The repeatability of the 5 walking trials was thus very high in each session, without marker replacement.

One can see that the similarity of the waveforms in the gait cycle was also high as determined by comparison of each parameter after marker re-attachment. The results confirmed that the reproducibility of each joint motion was high for the sagittal plane, the frontal plane, and the transverse plane for both test-retest and inter-tester reliability even though the markers were replaced. The CMC values reported in this study were higher than those found by Kadaba et al.\(^{10}\) and Steinwender et al.\(^{11}\). This result may be attributed to methodological factors such as the smaller subject numbers in this study, the standardization of the procedure of marker placement, and sufficient tester practice in marker placement. If careful marker placement is carried out according to a standardised protocol, reliability may be enhanced.

In this study, a poor CMC value for pelvic tilt was obtained similar to results of prior studies\(^{10,11}\). Two factors may have contributed to these low values. The primary factor relates to the calculation of the CMC formula, in that the CMC value decreases if the range of values is small. The second factor may be that error is easily produced by the sacral wand marker, as pointed out by Kadaba\(^{10}\). Generally, the range of pelvic tilt during gait was very small, and the average range of motion during the gait cycle was $2.3^\circ$ in the subjects. In comparison, the maximal difference of pelvic tilt measures between sessions was larger than the average range of motion. The problem in the computation of the CMC value may have resulted from the increase in this relative difference.

From these findings, the following recommendations for marker placement are therefore made in order to optimize reliability of three-dimensional and kinematic data:
- testers should be trained in palpation techniques to find bony landmarks accurately;
- marker application points should be clearly defined;
- application points should be marked using a skin-pen cil;
- wand-marker use should be double checked for pelvic tilt measurements.

The metric estimate of tolerable error was not investigated in this study. An essential task is to know how much error in the placement of markers is tolerable when the kinematic data in three-dimensional motion analysis is used to assess the effects of treatment. Further research should be conducted to provide metric estimates of permissible error ranges. Artifacts from body surface markers are a problem indicated by Cappello et al." which is a factor unrelated to the subject’s condition because it can be attributed to body shape and the movement of the subject’s skin. Since the body shape of all the healthy young subjects in the present research was similar, it is assumed that this artifact produced little effect. However, further investigation may be appropriate to reduce skin-marker movement error, as this clinical measurement is used for patients of various body types.

**Conclusion**

In this study, high levels of test-retest and inter-tester reliability were obtained by standardizing marker placement methods and procedures. Of all the factors that influence the reliability of a three-dimensional motion analysis system, reduction of variability of marker placement appears possible with standardization and understanding of the placement method. The results obtained in this study provide a basis for investigation of metric estimates of the amount of error incurred by marker placement, which will be used in future studies. These findings support the reliability of using three-dimensional motion analysis for measuring human gait.
Acknowledgment

The authors wish to thank Ms. Elise Cullis for her technical assistance with data collection and the staff of the Geriatric Research Unit at Kingston Centre for their support. The authors also wish to extend special thanks to the subjects of this study.

References