Application of principal component analysis on gait kinematics in elderly women with knee osteoarthritis

Aplicação da análise de componentes principais na cinemática da marcha de idosas com osteoartrite de joelho

Renata N. Kirkwood¹, Renan A. Resende², Cláudio M. B. Magalhães², Henrique A. Gomes³, Sueli A. Mingoti³, Rosana F. Sampaio¹

Abstract

Background: The applicability of gait analysis has been implemented with the introduction of the principal component analysis (PCA), a statistical data reduction technique that allows the comparison of the whole cycle between groups of individuals. Objectives: Applying PCA, to compare the kinematics of the knee joint during gait, in the frontal and sagittal planes, between a group of elderly women with and without diagnosis in the initial and moderate stages of Osteoarthritis (OA).

Methods: A total of 38 elderly women (69.6 ± 8.1 years) with knee OA and 40 asymptomatic (70.3 ± 7.7 years) participated on this study. The kinematics was obtained using the Qualisys Pro-reflex system.

Results: The OA group showed decreased gait velocity and stride length (p < 0.05) and was characterized with higher WOMAC pain score. In the frontal plane, the between-group differences of the components were not significant. In the sagittal plane, three principal components explained 99.7% of the data variance. Discriminant analysis indicated that component 2 and 3 could classify correctly 71.8% of the individuals. However, CP3, which captures the difference in the flexion knee angle magnitude during gait, was the variable with higher discrimination power between groups.

Conclusions: PCA is an effective multivariate statistical technique to analyse the kinematic gait waveform during the gait cycle. The smaller knee flexion angle in the OA group was appointed as a discriminatory factor between groups, therefore, it should be considered in the physical therapy evaluation and treatment of elderly women with knee OA.

Key words: Principal component analysis; gait; kinematics; osteoarthritis.

Resumo

Contextualização: A aplicabilidade da análise de marcha foi aprimorada com a introdução da análise de componentes principais (ACP), uma técnica estatística que reduz o volume de dados, permitindo a comparação de todo o ciclo entre grupos de indivíduos. Objetivos: Comparar, por meio da ACP, a cinemática da articulação do joelho durante a marcha nos planos sagital e frontal, entre mulheres idosas sem e com diagnóstico de leve a moderado de osteoartrite (OA). Métodos: Participaram 38 mulheres idosas (69,6±8,1 anos) com OA de joelhos e 40 assintomáticas (70,3±7,7 anos). A cinemática foi obtida usando o sistema Qualisys Pró-reflex. Resultados: O grupo OA apresentou menor velocidade da marcha e comprimento da passada (p<0,05) e caracterizou-se por maior escore de dor no WOMAC. No plano frontal, as componentes não foram significativas entre os grupos. No plano sagital, três componentes explicaram 90,7% da variância dos dados. A análise discriminante indicou que a componente 2 (CP2) e a 3 (CP3) classificaram corretamente 71,8% dos indivíduos. Entretanto, a CP3, que captava a diferença de amplitude do ângulo de flexão do joelho durante a marcha, foi a variável com maior poder de discriminação entre os grupos. Conclusões: A ACP é uma técnica estatística multivariada, efetiva para a análise das variáveis cinemáticas do ciclo da marcha. A perda de flexão de joelho na marcha do grupo com OA foi apontada como fator discriminante importante entre os grupos, sendo, portanto, uma variável que deve ser considerada na avaliação e tratamento fisioterápicos da mulher idosa com OA de joelho.

Palavras-chave: análise componentes principais; marcha; cinemática; osteoartrite.

Received: 10/02/2010 – Revised: 25/08/2010 – Accepted: 16/11/2010

¹ Physical Therapy Department, School of Physical Education, Physical Therapy and Occupational Therapy, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil
² Physical Therapist
³ Department of Statistics, School of Engineering, UFMG

Correspondence to: Renata Noce Kirkwood, Departamento de Fisioterapia, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6667, Sala 3125, Campus Pampulha, CEP 30270-901, Belo Horizonte, MG, Brasil, e-mail: renata.kirkwood@gmail.com.
Introduction

Osteoarthritis (OA) is a chronic degenerative disease of multifactorial etiology that mainly affects the weight-bearing joints. It is estimated that 4% of the Brazilian population present OA, which is more common in women being the knee as the second most affected joint by the disease, representing a total of 37% of cases. To understand the onset and the progression of OA, it is necessary to know about joint mechanics. The joint mechanics during movement can be estimated by analyzing the gait biomechanics, which involves kinematic and kinetic parameters that are compiled in temporal waveforms allowing the understanding of the gait dynamics. Although this procedure has been used successfully, the large volume of data obtained from a gait temporal wave analysis hinders the interpretation and, often, the appropriate use of information by the physical therapist.

In studies that have assessed the gait kinematics, the analysis was performed by discrete parameters, such as peak values or total amplitude of the curves. The definition of discrete parameters is subjective, and it becomes difficult to extract the same values of all temporal waves, especially in the presence of pathologies. Qualitative comparisons and subjective descriptions of temporal waves have already been explored. A key problem is that these statistical techniques do not take into account the correlation between variables that describe the angular displacement during the gait cycle, and also do not consider the entire cycle in the analysis. When the variables are correlated and the statistical hypothesis is tested by appropriate methods for independent variables, there is the risk of not detecting the effects of the variables precisely due to intercorrelations.

In 1997, Deluzio et al. introduced a new application of the principal component analysis method (PCA) which, in contrast to traditional forms of data analysis, includes a statistical analysis of the entire gait cycle, eliminating the effects of correlations among the variables. The objective of the PCA is to reduce a large amount of correlated variables to a smaller number, retaining most of the original data variability. The new variables, called principal components, are not correlated and ordered, so that the first component explains the largest proportion of data variability. Thus, the values of each component can be calculated by inferential analyses for example, for different groups of individuals. The applicability of the technique has grown, and several studies have already applied the PCA in order to reduce and interpret a large amount of variables from gait analysis of patients with OA, facilitating the decision-making of physical therapy treatments.

However, most of the gait studies in subjects with knee OA, analysed by the PCA, were focused on more advanced stages of the disease. The knowledge of the biomechanical gait characteristics in individuals at early and intermediate stages of knee OA can facilitate early physical therapy treatments that are likely to effectively reduce pain and possibly reduce the disease progression. Therefore, the aim of this study was to compare, through PCA, the angular displacement of the knee joint in sagittal and frontal planes during the gait cycle between two groups of participants, being a group diagnosed with mild to moderate knee OA and an asymptomatic group of participants. Once the components were defined, a discriminant analysis was applied to determine which kinematic variable better discriminates elderly women with mild to moderate OA from asymptomatic elderly women.

Methods

Sample

The study included 38 elderly women diagnosed with bilateral knee OA and 40 asymptomatic elderly women. The recruitment occurred from supporting groups or waiting lists for physical therapy treatment from outpatient clinics and from University Clinics of Belo Horizonte, Brazil. Groups were paired for age. The inclusion criteria were: to be sedentary according to the Center for Disease Control/American College of Sports Medicine (CDC/ACSM), which defines as sedentary a person who does not perform moderate physical activity for 30 minutes, five or more times a week; or vigorous activity for at least 20 minutes, three times a week; not receiving physical therapy treatment for lower limbs for at least six months; did not have falls over the last six months; walk without a walking aid device and had not suffered trauma or surgery on the ankle, knee and hip joints.

The specific criterion for inclusion in the OA group was the diagnosis of bilateral knee OA based on clinical and radiographic criteria of the American College of Rheumatology (ACR), a classification method that provides 91% of sensitivity and 86% of specificity. The radiographic classification was performed according to Kellgren and Lawrence. Only individuals with medial knee OA with mild to moderate degree (grades 2 and 3) were included. The diagnosis and the classification of OA degrees were given by the same examiner. The specific inclusion criterion...
in the control group was absence of history of prolonged pain or other symptoms consistent with OA in any joint of lower limbs. Subjects who had another type of disease in lower limb joints, concomitant neurological diseases or OA in other joints besides the knee were excluded.

**Instrumentation**

**Spatial-temporal and Kinematic Data**

For spatial-temporal and kinematic gait parameters, the motion analysis system Qualisys - ProReflex MCU (Qualisys MEDICAL AB, 411 12 Gothenburg, Sweden) with eight cameras was used. Synchronized to the Qualisys system, two force plates (AMTI) embedded in the walkway were used, which provided the data for determining the events of the gait cycle. Data were collected by the data acquisition software Qualisys Track Manager 1.6.0.x-QTM and then transferred to the Visual 3D software to process. The frequency of collection was 120 Hz.

**Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)**

WOMAC is a specific questionnaire to evaluate patients with knee OA. It consists of three domains: pain, joint stiffness and physical function. The condition is evaluated in the last 72 hours of questionnaire application. The pain subscale score, with five items was calculated using the 5-point Likert scale (0, 25, 50, 75, 100), with lower scores indicating lower levels of pain.

**Procedures**

This study was approved by the Ethics Committee in Research of the Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil (ETIC n°. 005/05), and all participants signed a consent form. The volunteers were subjected to an assessment of demographic aspects, then, the kinematic data collection of the most affected lower limb was performed, considering it as the one of greatest pain levels by subjects with OA, according to the WOMAC pain subscale. In the control group, only the dominant limb was evaluated, defined “as that with which the participant would kick a ball.”

The reflective marks were placed on bone prominences as the highest point of iliac crest, bilateral greater trochanter, medial and lateral epicondyle of the femur, medial and lateral malleolus, calcaneus and head of the 1st and 5th metatarsals. For screening of pelvis, thigh and leg, clusters were used containing reflective marks affixed in a non-colinear form, positioned between the posterior iliac spines at sacrum base and distal aspects of thigh and leg. Volunteers were instructed to walk barefoot on a self-selected speed along a walkway of 7 meters long by 1 meter wide. Ten strides of each subject were collected.

**Data reduction**

Data obtained were transferred to Visual 3D software to build the biomechanical model. They were interpolated and filtered using low pass fourth order Butterworth, with the cutoff frequency of 6 Hz, with the aim of reducing marks’ noise. The force platform data were processed delimiting the initial contact and the toe off and the contact of the same foot on the second platform, and the cycle was standardized from 0 to 100%. The sequence of Cardan angles was used to obtain the angular knee displacement. Flexion and adduction of the knee were considered positive. The variables gait velocity normalized by the individual’s height, stride length and angular knee displacement in the frontal and sagittal planes were assessed.

**Statistical analysis**

Average and standard deviation of WOMAC pain scores were used to characterize the sample and determine the most affected limb. The Shapiro-Wilk test was used to test the normality of anthropometric and spatial-temporal parameters. In the case of normality, Analysis of Variance was used and, in cases where normality was rejected, the nonparametric Mann-Whitney test was applied. According to the number of comparisons, Bonferroni correction was performed with a significance level of 0.05.

PCA was performed with the angular variables of knee joint in the sagittal and frontal plane of the gait cycle. The aim of PCA is to summarize the information contained in the 101 variables, corresponding to the 101 percentages of gait cycle, in a small number of components that explain the greater data variance through linear combinations obtained from those variables. Linear combinations are called principal components, uncorrelated with each other and able to retain as much of the information contained in the original p-variables. This analysis results in coefficients or load vectors that are the weights given to each variable for each component. Through them, the scores are calculated, which are derived by multiplying the load vectors by the angle variables values of each individual during the gait cycle. The coefficients or load vectors represent a specific aspect of temporal waves that allow components interpretation. The scores represent the distance that the temporal wave of an individual to the average for a given
principal component; scores with high values represent a wave closer to the maximum described by the component, and with low values, are closer to the minimum described by the same component

Data interpretation was performed graphically by load vectors and the comparison between the average curves of the individuals who showed the highest and lower scores, as suggested by Deluzio and Astephen. For choosing the number of components, it was considered that 90% of data variability should be explained.

Then, a discriminant analysis with a stepwise method was performed with the scores of each subject in each component in order to determine those that could best discriminate the two groups into OA and asymptomatic. Once the components have been identified, the values of Fisher’s linear coefficient for scores of the selected components were calculated and, then, transformed into an interval ranging from -1 to 1 to facilitate interpretation. The higher the value, the most important is the variable in the discrimination of elderly women with OA from elderly asymptomatic women. All data were analyzed with a significance level of 0.05 and using the SPSS software, version 15 and Minitab, version 16.

Results

Table 1 presents the characteristics of OA and asymptomatic groups. Only BMI was significantly different between groups, since participants with OA were 17% heavier than the asymptomatic group (p<0.05) (Table 1). According to the classification of Kellgren and Lawrence, 14 participants were classified as grade 2 (mild OA), and 24 as grade 3 (moderate OA). In relation to WOMAC questionnaire, the OA group had higher average in pain domain, of 180.6 (SD = 105.6), compared to the asymptomatic group with an average pain of 5 (SD = 17.1).

Regarding temporal and spatial variables of gait, the group with OA showed a smaller stride length, a lower gait speed and greater time spent in stance and swing phases compared to the asymptomatic group (p<0.05) (Table 1).

In the principal components analysis in the frontal plane, three components were needed to explain 91.3% of the total data variability. The component 1 (CP1) explained 70.6%, the component 2 (CP2), 12.3% and the component 3 (CP3), 8.4%. The stepwise discriminant analysis was conducted to determine whether the scores of the three components could predict the condition of belonging or not to the OA group. The result of Wilk’s Lambda was not significant (p>0.05), indicating that scores of the three knee angle components in the frontal plane could not differentiate the conditions of belonging or not to the OA group, ie, groups do not differ with respect to knee angular displacement in the frontal plane.

In the sagittal plane, three components explained 90.7% of data variability. CP1 explained 61.1%; CP2, 18.8% and CP3 10.8%. The stepwise discriminant analysis conducted on scores of the three components showed that the results of Wilk’s Lambda was significant (p<0.05), indicating that only the scores of CP2 and CP3 could differentiate OA and asymptomatic groups, with 71.8% of subjects (56/78) correctly classified by discriminant function.

Figure 1A shows the average angular curves of OA and asymptomatic groups in the sagittal plane. It is observed that among phases of medium stance (30 and 50% of the cycle) and medium swing (60 and 80% of the cycle) of the gait, the OA group exhibits lower angle values than the asymptomatic group. The interpretative analysis of components (Figure 1B) shows that CP2 shows negative values in the stance phase, most markedly in terminal stance and pre-swing (45°-65°), and positive values from the initial swing phase until the medium swing (70°-85°). Therefore, capturing the difference between the knee angle at the end of the stance phase until the initial swing. Regarding the highest score, represented by a subject from the asymptomatic group, and the lowest score of CP2, represented by a subject in the OA group, it is observed that the OA group shows loss of knee extension on the final stance phase and loss of flexion in swing phase (Figure 1C).

Table 1. Anthropometrics and spatial-temporal data during gait in a group of participants with and without diagnoses of knee OA in the beginning of the study (N=78).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group Asymptomatic (N=40) average (SD)</th>
<th>Group OA (N=38) average (SD)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>70.3 (7.7)</td>
<td>69.6 (8.1)</td>
<td>0.698a</td>
</tr>
<tr>
<td>Height(m)</td>
<td>1.5 (0.01)</td>
<td>1.5 (0.01)</td>
<td>0.134b</td>
</tr>
<tr>
<td>BMI m/Kg2</td>
<td>27.2 (0.8)</td>
<td>31.1 (0.9)</td>
<td>0.002a</td>
</tr>
<tr>
<td>Stride Length (m)</td>
<td>1.2 (0.1)</td>
<td>1.1 (0.1)</td>
<td>0.000a</td>
</tr>
<tr>
<td>Velocity (m/sec)</td>
<td>1.1 (0.2)</td>
<td>0.9 (0.1)</td>
<td>0.000a</td>
</tr>
<tr>
<td>Stance Phase (sec)</td>
<td>0.7 (0.1)</td>
<td>0.7 (0.1)</td>
<td>0.000a</td>
</tr>
<tr>
<td>Swing Phase (sec)</td>
<td>0.4 (0.03)</td>
<td>0.5 (0.04)</td>
<td>0.001a</td>
</tr>
</tbody>
</table>

*aAnova (one way) test significant at p<0.0125; *Mann-Whitney test significant at p<0.0017.

Table 2. Normalized discriminant function of the PC2 and PC3 scores in the sagittal plane.

<table>
<thead>
<tr>
<th>Components</th>
<th>Normalized Discriminant Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC2</td>
<td>0.257</td>
</tr>
<tr>
<td>PC3</td>
<td>*0.966</td>
</tr>
</tbody>
</table>

* PC3 with higher discriminant power between groups.
The CP3 shows positive values in the swing phase and negative values in the stance phase, thus captures the range of knee flexion during gait (Figure 1B). In Figure 1D, it is observed that the difference between the highest (asymptomatic group) and the lowest score (group OA) of CP3 is the range of knee flexion throughout the gait cycle. The average knee range of motion in the sagittal plane throughout the gait cycle of the asymptomatic group was 55.8° and the group with OA was 51.9°.

To determine which significant components in the sagittal plane are more important in the discriminant function, the values of the linear Fisher coefficients of CP2 and CP3 were obtained by means of discriminant analysis and transformed to the interval [-1 to 1]. As shown in Table 2, CP3, which represents the range of knee flexion in the gait, has a coefficient of discrimination much greater than the CP2, therefore, it is the variable that best discriminates the two groups of participants with mild or moderate OA from the asymptomatic group. Figure 2 confirms this finding, because in the dispersion diagram the scores of OA and asymptomatic groups were split in the direction of CP3, ie, CP3 is the component that best distinguish the groups in mild to moderate OA from asymptomatic elderly women.

Discussion

This study aimed to compare the angular knee displacement in sagittal and frontal planes during gait in a group of elderly women with knee OA and a group of asymptomatic elderly women through the PCA. PCA is a multivariate statistical technique appropriate for the interpretation of gait cycle variables, because it reduces the volume of data into a smaller number, taking into account all the information contained in the cycle, eliminating the intercorrelations, but retaining most of the information contained in the original data. Other methods of data reduction, using pre-defined parameters for comparison between peaks and ranges, although they are simple techniques, do not include all temporal information of the cycle, beyond the studied parameters are highly correlated.

The discriminant model applied in scores of CP2 and CP3 of the sagittal plane, correctly classified 71.8% of elderly women, ie, the loss of knee extension in stance phase portrayed by CP2 and knee flexion range of motion in the gait cycle portrayed by CP3 are important angular displacement information in detecting elderly women with and without knee OA. Although CP2 had a greater pattern of variation among individuals, the contribution of CP3 in the discriminant analysis was higher, so the impact of knee
flexion loss in the gait cycle is more important in stages of mild to moderate OA. Astephen et al.\textsuperscript{4} showed that the loss of knee flexion range in the stance phase of gait is progressive, ie, increases as the degenerative process progresses. Therefore, clinically the results of this study shows that physical therapy treatment for individuals with mild to moderate degree of OA should focus on gaining range of knee flexion, with the goal of trying to slow the progression of OA or even be used as a marker of clinical comparison between groups.

The presence of knee pain or anterior muscle weakness of thigh may be associated with decreased range of knee flexion during gait. In fact, the OA group in this study, had pain and physical function scores larger than the asymptomatic group. Maly, Costigan and Olney\textsuperscript{26} showed that the smaller the excursion of flexion/extension of knee in the gait, the higher the pain scores of OA individuals. Although a cause and effect relationship between limiting range, muscle weakness and pain in OA is still unclear in the literature, physical therapy treatment focusing on muscle strengthening can improve knee range of motion during gait, and also allow pain relief.

Another important variable is the decrease in gait speed observed in the OA group. Several studies have demonstrated the effect of speed on gait markers\textsuperscript{4,10,27}. The decrease in speed reduces the acceleration of movement and consequently the force momentum\textsuperscript{28}, therefore, it is a strategy that individuals with OA would use in response to pain to reduce joint stress. Recently, Zeni and Higginson\textsuperscript{6} shown that even when the speed variable was controlled, the range of knee flexion was still significantly lower in OA compared to an asymptomatic group. The authors state that knee excursion in the sagittal plane in gait is directly related to the pathology.

However, recently, Robbins and Maly\textsuperscript{29} noted that slowing gait speed in healthy subjects, decreased knee peak load, but there was an increase in the impulse, ie, the knee received a lower overload, but for a longer time. Based on these findings, the authors questioned whether slow the gait speed is an interesting strategy for individuals with OA\textsuperscript{29}. Although in the present study the variable impulse has not been investigated, subjects in OA group spent more time in the stance phase and swing than those from the asymptomatic group. It is possible that the prolonged effect of a lower overload is next to a short effect of intense joint stress, increasing pain. To answer this question, more studies are needed to understand other factors that may contribute to the decrease in gait speed in subjects with knee OA.

The impact of these changes in the range of other joints of the lower limb should be investigated. Limited range of flexion in the stance phase of gait can change the phase of shock absorption and slow the advance of the lower limb. The end of the stance phase and initial swing would affect mainly the toe off phase\textsuperscript{30}. The loss of knee extension in terminal stance phase, as pointed out by CP2, decreases the stability of the support limb, increasing the energy expenditure of all gait\textsuperscript{31}. As the knee range of motion is an important discriminant factor, knowledge of the effect of this loss in other joints of the lower limb in the initial stages of OA may further enhance the rehabilitation programs.

Although this study has identified significant kinematic changes due to OA, further studies are required to understand cause and effect of mechanical gait factors in the pathology, requiring longitudinal studies with a focus in this objective. It would also be important to check these findings to a group of male participants, since the gender factor changes the relationship of gait markers\textsuperscript{23}.

In conclusion, through the PCA, one can restrict the number of variables, making it easier to use statistical techniques to identify gait characteristics, such as loss of knee flexion, which differs elderly women in mild or moderate stages of knee OA from asymptomatic elderly women. This finding will enable physical therapists to develop strategies focused on the found mechanical malfunction, improving the individual’s performance or even helping to slow the OA progression.

References:


