# *Objective Classification of Gait Patterns in Children with Cerebral Palsy*

**Mervic Azupardo**

promotor :
Prof. dr. Herbert THIJS, Mrs. Kaat
DESLOOVERE

Universiteit Maastricht

Eindverhandeling voorgedragen tot het bekomen van de graad
Master of Science in Biostatistics

universiteit hasselt

Universiteit Hasselt
Center for Statistics

# OBJECTIVE CLASSIFICATION OF GAIT PATTERNS IN CHILDREN WITH CEREBRAL PALSY

By: Mervic Azupardo

External Supervisor: Prof. Dr. Kaat Desloovere
Internal Supervisor: Prof. Dr. Herbert Thijs

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biostatistics

HASSELT, 2007

# Certification

This is to certify that this report was written by Mervic Azupardo under our supervision.

……………………….…………..…...Date…………………

**Prof. Dr. Herbert Thijs** **Internal Supervisor**

………………..……........…………..Date……………………

**Prof. Dr. Kaat Desloovere** **External Supervisor**

Thesis is submitted in partial fulfilment of the requirements
for the degree of Master of Science in Biostatistics.

……….………………..…….……. Date…………………

**Mervic Azupardo** **Student**

Hasselt, 2007

# TABLE OF CONTENTS

# I. INTRODUCTION

## A. Background of the Study

“Cerebral” is defined as pertaining to the brain, cerebrum or intellect, and “Palsy” refers to paralysis of muscle or group of muscles. Jointly, Cerebral Palsy (CP) is the term used to describe a group of conditions with motor impairments resulting from brain damage during the early stages of development. Cerebral palsy is non progressive and is usually not diagnosed until a child is about 2 to 3 years of age. About 2 to 3 children in 1,000 over the age of three have cerebral palsy. This is because of the plasticity of a child's central nervous system, or its ability to recover completely or partially after an injury occurs. If a brain injury occurs early, the undamaged areas of a child's brain can sometimes take over some of the functions of the damaged areas.

There are four categories of CP based on the different movement impairments: Spastic, Athetoid, Ataxic and Mixed Forms. Spastic cerebral palsy is characterized by muscles that are stiff and permanently contracted, making awkward movements to varying degrees of severity and in turn limits the patients’ range of motion and causing jerky, unpredictable movements. This approximately affects about 70 to 80 percent of CP patients. Often, CP patients have a hard time moving from one position to another and may also have a hard time holding and letting go of objects. Athetoid cerebral palsy is characterized by uncontrolled movements of the hands, feet, arms, or legs and, in some cases, the muscles of the face and tongue. These uncontrolled movements often interfere with speaking, feeding, reaching, grasping, and other skills requiring coordinated movements. About 10 to 20 percent of CP patients have the Athetoid case. Ataxic CP is a rare form of cerebral palsy which affects an estimated 5 to 10 percent of CP patients. Ataxic cerebral palsy affects sense of balance and depth perception. Typically, persons affected by ataxic cerebral palsy have poor coordination, walk unsteadily and exhibit difficulty when attempting quick or precise movements. Lastly, the Mixed form of CP have symptoms of more than one of the categories mentioned. The most common mixed form includes spasticity and athetoid movements.

CP is also classified according to the affected region of the body: (1)Diplegia, affects either both arms or both legs of the patients; (2) Hemiplegia, affects the limbs on only one side of the body; (3) Quadriplegia, affects all the limbs; (4) Monoplegia, affects only one limb; and (5) Triplegia, affects three limbs.

The cause of CP is still unknown but there are factors identified that can possibly increase the risk. The following are some of the factors that might cause CP: complicated labor and delivery of infants (inborn brain damage, breech presentation, etc.), unhealthy condition of infant (low birth weight, prematurity, Apgar score, nervous system malformation and others), and seizures in the newborn. Due to this prevention like regular prenatal care and proper child care should be noted.

The life expectancy of CP patients depends on the severity of their condition. Research has shown that if a child has severe cerebral palsy, other complications may occur to shorten the patients' life. Otherwise, if the CP patients' medical condition is considered to be typical, the child is likely to have normal life expectancy.

Treatments for CP patients come in various forms. Physical therapy usually begins shortly after the diagnosis is made. Specific exercises are used in physical therapy to prevent the weakening or deterioration of muscles from disuse (disuse atrophy) and to avoid muscle contracture (muscles fixed in a rigid, abnormal position). Drug Therapy is necessary for those who have seizures associated with cerebral palsy, and may be effective in preventing seizures in many patients. Surgery may not be necessary, but it is sometimes recommended to improve muscle development, correct contractures, and reduce spasticity in the legs which in turn can help the patient child achieve his or her optimal level of functioning.

## B. The Data

The assessment was performed according to the standard protocol of the Clinical Motion Analysis Laboratory of the University Hospital, Pellenberg, Belgium. The evaluation included anthropometric measurements, analogue video recording, 3D kinematics and kinetic data collection, dynamic surface electromyography and an extended clinical examination

**Table 1: Selected gait analysis parameters**

| |
|---|
| *Time and distance parameters* |
| Walking velocity (m/s) |
| Cadence (steps/min) |
| Step length (m) |
| Timing of toe off (% of gait cycle) |
| *Ankle and foot kinematics and kinetics* |
| Ankle angle at initial contact (°) |
| Maximum ankle dorsiflexion between loading response and mid-stance (°) |
| Range of ankle motion during push-off (°) |
| Timing of maximal ankle dorsiflexion in stance (°) |
| Maximum ankle dorsiflexion in stance (°) |
| Ankle angle at midswing (°) |
| Mean foot alignment in stance (°) |
| Maximum plantairflexionmoment at pre-swing (Nm/kg) |
| Moment in the ankle at loading response (Nm/kg) |
| Ankle double bump pattern |
| Ankle second rocker |
| Ankle power absorption at loading response (W/kg) |
| Ankle power generation at pre-swing (W/kg) |
| Maximum ankle velocity around toe-off (rad/s) |
| *Knee kinematics and kinetics* |
| Knee flexion at initial contact (°) |
| Shock absorption in stance =Maximum knee flexion in stance – Knee angle at initial contact (°) |
| Maximum knee flexion in stance (°) |
| Maximum knee extension in stance (°) |
| Maximum knee flexion in swing (°) |
| Amount of delayed knee flexion in swing (% of gait cycle, relative to toe-off) |
| Maximum knee flexion moment in stance (Nm/kg) |
| Maximum knee extension moment is stance (Nm/kg) |
| Maximum knee power generation in stance (W/kg) |
| Maximum knee power absorption in stance (W/kg) |
| Maximum knee flexion velocity around toe-off (rad/s) |
| *Pelvic motion* |
| Pelvic mean anterior tilt (°) |
| Range of pelvic motion in sagittal plane (°) |
| Range of pelvic obliquity (°) |
| Range of pelvic rotation (°) |
| Mean of pelvic obliquity (°) |
| Mean pelvic rotation (°) |
| *Hip kinematics and kinetics* |
| Hip angle at terminal stance (°) |
| Range of sagittal hip motion in stance (°) |
| Maximum hip flexion in swing (°) |
| Mean coronal hip angle in stance (°) |
| Mean coronal hip angle in swing (°) |
| Hip rotation angle at initial contact (°) |
| Hip rotation angle at terminal stance (°) |
| Maximum hip abduction moment in stance (Nm/kg) |
| Timing of 0 moment in hip (% of gait cycle) |
| Maximum hip power generation in stance (W/kg) |
| Maximum hip power absorption in stance (W/kg) |
| Hip power generation at terminal stance (W/kg) |
| Maximum hip flexion velocity in swing (rad/s) |

(Christiaens et al, 2006). There were 460 CP patients included in this retrospective study. The patients' gait analyses which took place between January 1997 and December 2006 were utilized.

The following were the four criteria for CP patients' inclusion in the study: (1) children had to be capable to undergo a full barefoot walking gait analysis independently, (2) orthopedic surgery prior to the evaluation time was not allowed [one exception is an Achilles tendon lengthening at least 4 years prior to the evaluation time], (3) the latest Botulinum Toxin Type A (BTX-A) treatment had to be at least 6 months prior to the evaluation time and the children had to be treated by a multidisciplinary team. This implies that all children had to be treated according to the integrated approach, which was based upon the research of Molenaers in 1999. (Christiaens et al, 2006)

There were 46 variables included in this study. All of which are gait analysis parameters namely: Time and distance (4 parameters), ankle and foot kinematics and kinetics (12 parameters), knee kinematics and kinetics (11 parameters), pelvic motion (6 parameters), and hip kinematics and kinetics (13 parameters). The variable gender was not considered in the analysis since it not considered as a risk factor associated with an infant being born with cerebral palsy. Although according to literature in rare cases very low birth weight and abnormal intrauterine size is associated with an increased risk of cerebral palsy for boys.

The purpose of this study is to develop a classification system for Cerebral Palsy patients based on their gait patterns utilizing their kinematics, kinetics and clinical data. The remainder of this section describes some related studies on gait patterns. Section 2 discusses the methodology to be used then proceeding to the results and discussions. The softwares used in this study were the following: SAS software for data reduction techniques namely: Principal Components Analysis (proc princomp), Factor Analysis (proc factor) and Multidimensional Scaling (proc mds). The same software was used for procedures like Hierarchical (proc cluster) and Non-Hierarchical (proc fastclus) clustering. For the normal mixture modeling clustering, the R language software was used specifically utilizing the "mclust" package obtained from http://cran.r-project.org/.

# II. METHODOLOGY

In investigating the method to classify the CP data three approaches were considered, hierarchical, non-hierarchical and normal mixture clustering techniques. These methods have different techniques in classifying the observations into groups, but cross classifications of these methods will be done to check for classifying consistency in the clustering procedures. The data was found to be high dimensional which became a motivation for data reduction techniques namely: principal components analysis (PCA), factor analysis (FA) and multidimensional scaling (MDS). These data reduction techniques will be used as an input in the three clustering procedures.

## A. VARIABLE REDUCTION TECHNIQUES

### A.1. Principal Components Analysis

Principal Components Analysis (PCA) aims at explaining the variance-covariance structure of a set of variables through the use of linear combinations of the acquire variables. Its general objectives are (1) data reduction and (2) interpretation. (Johnson, et al, 2002)

Given $n$ observations on p variables the purpose of the analysis is to find $k$ new variables ($k< p$) which are obtained from the original p variables. The generated $k$ new variables are now called principal components, account for as much as possible of the variation on the acquired $p$ variables while being mutually uncorrelated and orthogonal. The first principal component possesses the largest proportion of the variation in the dependent variable compared to the succeeding principal components which possess some proportion of the variation in accordance with their succession. By using the first few principal components as new variables, rather than using the original variables, the dimension of the dataset is now reduced.

## A.2. Factor Analysis

Factor analysis (FA) focuses on the variance that is analyzed. It is assumed that the variance of a single variable can be decomposed into a "*common variance*" that is shared by the other variables included in the model. And that this variance is unique to a specific variable and includes an error component. In common factor analysis, a small number of factors are extracted to account for the intercorrelations among the observed variables to identify the latent dimensions that explain why the variables are correlated with each other.

## A.3 Multidimensional Scaling

The objective of Multidimensional Scaling (MDS) is to "fit" the original data into a low-dimensional coordinate system such that any similarity or dissimilarity caused by a reduction in dimensionality is minimized (Johnson and Wichern, 2002). In addition, the low dimensional representation should be as approximately correct as possible. To determine this, the MDS defines a function called "stress" where interpretation is indicated in the table below.

Kruskal's informal guidelines for the stress function

| Stress | Goodness of Fit |
|---|---|
| 20 % | Poor |
| 10% | Fair |
| 5% | Good |
| 2.5% | Excellent |
| 0% | Perfect |

*Goodness of fit* of fit refers to the monotonic relationship between the similarities and the final distance (Johnson and Wichern, 2002). In every $q$ dimension generated from the MDS procedure denotes a specific stress value. The stress function decreases as $q$ increases. Graphically, the number of dimensions can be determined by locating a sharp elbow on a scree plot of the stress against the number of dimensions. The purpose of the MDS in this paper is not to graphically display the dimensions but as a way in reducing the high dimensionality of the CP data.

## B. CLUSTERING TECHNIQUES

Grouping items into clusters is a form of summarizing information in a large body of data and at the same time provides additional insight. There are different ways of grouping items but all aiming at grouping observations that are very similar to each other and at the same time dissimilar when compared to observations from another group. Formally, these grouping techniques can be categorized as hierarchical, non-hierarchical or normal mixture clustering. The hierarchical clustering arranges the observation in a form of a hierarchy, hence the name, while the non hierarchical clustering requires a pre-specified number of clusters. Both procedures have specific trade offs in their strengths and weaknesses. The hierarchical allows the observations to cluster themselves by a specified distance but is sensitive to outliers while the non-hierarchical is robust in the possibility of outliers but needs a pre-specified number of clusters on classifying the observation. The normal mixture clustering estimates a model for the data that allows for overlapping clusters, producing a probabilistic clustering rather than deterministic since it allows for quantification of uncertainty in observations to belong to mixture components. The section that follows defines these techniques in more detail.

### B.1. Hierarchical Clustering

Hierarchical clustering techniques proceed by a series of successive fusions or divisions. Agglomerative hierarchical techniques start with the individual objects as distinct clusters which indicate that there are as many clusters as there are objects. The most similar objects are initially grouped based on their degree of similarity. Subsequently, as the similarity decreases, all subgroups generated by the algorithm will combine to form a single cluster. The Divisive technique is the inverse of the agglomerative technique; it initially starts with all objects belonging to a single group then eventually splits according to their dissimilarities.

This study will employ the agglomerative clustering technique through the linkage methods. Each observation starts as a cluster by itself”. Then the two closest clusters are merged to form a new cluster that replaces the two old clusters (previously called cluster by itself). Merging the two closest clusters is repeated until only one cluster is left. The various clustering methods differ in how the distance between two clusters is calculated. In this study the following are employed: single linkage (minimum distance), complete linkage (maximum distance), centroid linkage (cluster centroid) and Ward’s method (ANOVA sum of squares). These distances will be presented for comparison purposes to justify the number of cluster that is most appropriate to the data.

**Single linkage**

Single linkages or “*shortest distance*” algorithm calculates the distances or similarities between pairs of objects. Groups are formed from the individual entities by merging nearest neighbors, where the term *nearest neighbor* connotes the smallest distance or largest similarity. The distance between two clusters is defined by $D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(x_i, x_j)$ **(1)**

Where; $D_{KL}$, any distance or dissimilarity measure between clusters $C_K$ and $C_L$

$C_K$ , *K*th cluster, subset of {1, 2, ... , $n$ }
$C_L$ , *L*th cluster, subset of {1, 2, ... , $n$ }
$x_i$ , *i*th observation (row vector if coordinate data)
$x_j$ , *j*th observation (row vector if coordinate data)

The distance between two clusters is the minimum distance between an observation in one cluster and an observation in the other cluster. By imposing no constraints on the shape of clusters, single linkage tends to sacrifice performance in the recovery of compact clusters in return for the ability to detect elongated and irregular clusters.

### Complete Linkage

In contrast with the above mentioned distance, complete linkages or "*maximum distance*" algorithm estimates the distance between two clusters which is the maximum distance between an observation in one cluster and an observation in the other cluster. Complete linkage is strongly biased toward producing clusters with roughly equal diameters, and it can be severely distorted by moderate outliers (Milligan 1980).

The distance between two clusters is defined by $D_{KL} = \max_{i \in C_K} \max_{j \in C_L} d(x_i, x_j)$ **(2)**

Where; $D_{KL}$ ,any distance or dissimilarity measure between clusters $C_K$ and $C_L$

$C_K$ , $K$th cluster, subset of {1, 2, ... , $n$ }

$C_L$ , $L$th cluster, subset of {1, 2, ... , $n$ }

$x_i$ , $i$th observation (row vector if coordinate data)

$x_j$ , $j$th observation (row vector if coordinate data)

### Centroid Method

In the centroid method, the distance between two clusters is defined as the (squared) Euclidean distance between their centroids or means. The centroid method is more robust to outliers than most other hierarchical methods but in other respects may not perform as well as Ward's method or average linkage (Milligan 1980).

The distance between two clusters is defined by $D_{KL} = \left\| \overline{X}_K - \overline{X}_L \right\|^2$ **(3)**

Where; $D_{KL}$ ,any distance or dissimilarity measure between clusters $C_K$ and $C_L$

$X_K$, mean vector for cluster $C_K$

$X_L$, mean vector for cluster $C_L$

**Ward's Minimum-Variance Method**

In Ward's minimum-variance method, the distance between two clusters is the *ANOVA* sum of squares between the two clusters added up over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation. The sums of squares are easier to interpret when they are divided by the total sum of squares to give proportions of variance (squared semi partial correlations).

The distance between two clusters is defined by $$D_{KL} = B_{KL} = \frac{\left\| \overline{X}_K - \overline{X}_L \right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}} \quad \textbf{(4)}$$

Where; $D_{KL}$ ,any distance or dissimilarity measure between clusters $C_K$ and $C_L$

$X_K$, mean vector for cluster $C_K$

$X_L$, mean vector for cluster $C_L$

Ward's method tends to join clusters with a small number of observations, and it is strongly biased toward producing clusters with roughly the same number of observations. It is also very sensitive to outliers (Milligan 1980).

**B.2. Non-Hierarchical Clustering**

**K – Means Method**

The K-means algorithm according to Hastie et. al., (2001) is one of the most popular iterative descent clustering methods. The said method is intended for data wherein all variables are quantitative, and the squared Euclidean distance $$d(x_i, x_{i'}) = \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2 = \| x_i - x_{i'} \|^2 \quad \textbf{(5)}$$

is chosen as the dissimilarity measure.

The algorithm starts by partitioning the input points into k initial sets, either at random or using some heuristic data. It then calculates the mean point, or centroid, of each set. It constructs a new partition by associating each point with the closest centroid. Then the centroids are recalculated for the new clusters. The algorithm is repeated by alternate application of these two steps until convergence, which is obtained when the points no longer switch clusters.

There are specific criterions that should be considered in determining which clusters suits the data best. One of these is the pseudo F statistics which should have a relatively large value when compared across different levels of cluster/groups. Another is the Cubic Clustering Criterion which actually has guidelines in determining the quality of the clusters generated. The CCC is best used when its values are plotted across the number of clusters. Since the CCC is a criterion in the K-means, it also requires the data to be large, otherwise the values misbehave. There are some guidelines in interpreting CCC. First indication of the peaks on the plot with CCC greater than 2 or 3 indicate good clustering and peaks between 0 and 2 indicate possible clusters but should be interpreted cautiously. It also indicates that there may be several peaks if the data has a hierarchical structure, while very distinct nonhierarchical spherical clusters usually show a sharp rise before the peak followed by a gradual decline. Very distinct nonhierarchical elliptical clusters often show a sharp rise to the correct number of clusters followed by a further gradual increase and eventually a gradual decline. Another indication is that if all values of the CCC are negative and decreasing for two or more clusters, the distribution is probably unimodal or long-tailed. Although very negative values of the CCC, say, -30, may occur due to outliers. Outliers generally should be removed before clustering. If the CCC increases continually as the number of clusters increases, the distribution may be grainy or the data may have been excessively rounded or recorded with just a few digits.

The *r-square* criterion should also be noted and should look for a value that explains as much variance as appropriate for the study. Milligan and Cooper (1985) demonstrated that changes in the R-Square are not very useful for estimating the number of clusters, but it may be useful if you are interested solely in data reduction.

### B.3. Mixtures

Finite mixture modeling has been recognized as models that can provide statistical approach to the practical questions that arise in applying clustering methods (McLachlan and Basford 1998; Banfield and Raftery 1993; Cheeseman and Stutz 1995; Fraley and Raftery 1998). Several strategies were investigated by incorporating hierarchical agglomerative clustering technique and the EM algorithm using the Bayesian Information Criterion (BIC). An advantage of the previous is that it tends to produce clusters that have good partitions without information on the groupings. Likewise for the EM algorithm, good partitions can also be expected. The main difference is that in the EM algorithm pre-specification of the number of clusters is required with several initiations since the likelihood surface tends to have multiple modes. Addressing these two issues, Fraley and Raftery (2002) extended this strategy by allowing selection of the parameterization of the model as well as the number of clusters simultaneously using the BIC.

Given data **y** with independent multivariate observations $\mathbf{y}_1,\ldots,\mathbf{y}_n$, the likelihood for a mixture model with G components is

$$L_{MIX}(\theta_1,\ldots,\theta_G;\tau_1,\ldots,\tau_G \mid \mathbf{y}) = \prod_{i=1}^{n}\sum_{k=1}^{G}\tau_k f_k(y_i \mid \theta_k), \tag{6}$$

Where $f_k$ and $\theta_k$ are the density and parameters, respectively, of the $k$th component in the mixture, and $\tau_k$ is the probability that the observation belongs to the $k$th component ($\tau_k \geq 0; \sum_{k=1}^{G}\tau_k = 1$).

Most commonly, $f_k$ is the multivariate normal (Gaussian) density $\phi_k$ parameterized by its mean $\mu_k$ and covariance matrix $\Sigma_k$:

$$\phi_k(\mathbf{y}|\mu_k,\Sigma_k) = \exp\{-1/2\ (\mathbf{y}_i-\mu_k)^T \Sigma_k^{-1}\ (\mathbf{y}_i-\mu_k,)\} \ / \ \sqrt{\det(2\pi\Sigma_k)} \tag{7}$$

Data generated by mixtures of multivariate normal densities are characterized by groups or clusters centered at the mean $\mu_k$, with increased density points nearer the mean. The covariance $\Sigma_k$ determines the geometric features (shape, volume, orientation) of the clusters.

Each covariance matrix is parameterized by eigenvalue decomposition in the form

$$\Sigma_k = \lambda_k D_k A_k D_k^T, \qquad \textbf{(8)}$$

Where: $D_k$ is the orthogonal matrix of eigenvectors

$A_k$ is a diagonal matrix whose elements are proportional to the eigenvalues of $\Sigma_k$, and

$\lambda_k$ is a scalar.

Also, the orientation of the principal components of $\Sigma_k$ is determined by $D_k$ ,while $A_k$ determines the shape of the density contours; $\lambda_k$ specifies the volume of the corresponding ellipsoid, which is proportional to $\lambda_k^d |A_k|$, where d is the data dimension.

**Table 2. Parameterization leading to different models**

| Identifier | Model | Distribution | Volume | Shape | Orientation |
|---|---|---|---|---|---|
| EII | $\lambda \boldsymbol{I}$ | Spherical | equal | equal | NA |
| VII | $\lambda_\kappa \boldsymbol{I}$ | Spherical | variable | equal | NA |
| EEI | $\lambda \boldsymbol{A}$ | Diagonal | equal | equal | coordinate axes |
| VEI | $\lambda_\kappa \boldsymbol{A}$ | Diagonal | variable | equal | coordinate axes |
| EVI | $\lambda \mathbf{A}_\kappa$ | Diagonal | equal | variable | coordinate axes |
| VVI | $\lambda_\kappa \mathbf{A}_\kappa$ | Diagonal | variable | variable | coordinate axes |
| EEE | $\lambda \boldsymbol{DAD^T}$ | Ellipsoidal | equal | equal | equal |
| EEV | $\lambda \boldsymbol{D}_k \boldsymbol{A} \boldsymbol{D}_k^T$ | Ellipsoidal | equal | equal | variable |
| VEV | $\lambda_\kappa \boldsymbol{D}_k \boldsymbol{A} \boldsymbol{D}_k^T$ | Ellipsoidal | variable | equal | variable |
| VVV | $\lambda_\kappa \boldsymbol{D}_k \boldsymbol{A}_k \boldsymbol{D}_k^T$ | Ellipsoidal | variable | variable | variable |

Models are compared to each other by utilizing the BIC (Bayesian Information Criterion).

# III. RESULTS

## A. Exploratory Data Analysis

The 460 patients included in this study are composed of 198 girls and 262 boys. Correlation was observed for almost all the gait analysis parameters. The data was noted to have no missing observations. It was also observed that the distributions of the gait parameters were mostly skewed (Appendix Section A).

### B. Variables reduction techniques

Due to the high dimensionality of the CP data, data reduction techniques were utilized (PCA, FA and MDS). These techniques will then be used as input in the clustering procedures (Hierarchical, Non-hierarchical and Normal Mixtures) included in this study.

**Principal Component Analysis**

Utilizing the Principal Component Analysis (PCA), the dimension of the dataset can be reduced without too much loss of information. The PCA was used on all the 46 gait parameters. From the 46 variables, 18 principal components were produced accounting for 80.44 percent of variability explained. Principal components (PCs) are arranged in such a way that the first few linear combinations have the largest variation explained. Looking into the details of the linear combinations, the first PC indicated contrasts between the measurement on *knee parameters* and *hip measurements.* The second PC showed that the more description on the *ankle measurements*. The third and fourth PC explained more on the *knee measurements* and *hip measurements,* respectively. While the fifth PC explained the *pelvic* and *hip measurements.* The sixth PC indicated a contrast between the *knee* and *hip measurements,* which may be a description of the walking position of children to balance themselves and walk independently given their CP conditions. These first six PCs explained about fifty percent of the variability of the original data and the succeeding PCs contributed little increments in explaining the variability. For this study, 18 PCs were chosen so as not to allow a large amount of information loss.

**Factor Analysis**

Through Factor Analysis (FA), variables can be reduced in linear combinations. The *varimax* orthogonal rotation was used in this analysis aiming to obtain a clear pattern of loadings identifying the gait parameters and aims to maximize the variance on the new axes, which in turn leads to easier interpretation. There were 6 factors extracted from the FA procedure. These linear combinations convey essential information contained in the original set of variables. In this study the interrelationships are more pronounced in the "*time and distance parameters*" and "*knee parameters*" than in "*hip parameters*".

**Multidimensional Scaling**

From the 46 variables included in the study, the number of dimensions was determined by a criterion called the *stress function*. The stress function for the CP data was set to have about 5% Goodness of fit which resulted to 13 dimensions.

| Dimensions | Goodness of fit |
|---|---|
| 2 | 0.5478 |
| 3 | 0.4406 |
| 4 | 0.3641 |
| 5 | 0.3085 |
| 6 | 0.2576 |
| 7 | 0.2162 |
| 8 | 0.1801 |
| 9 | 0.1444 |
| 10 | 0.1150 |
| 11 | 0.0878 |
| 12 | 0.0652 |
| 13 | 0.0486 |

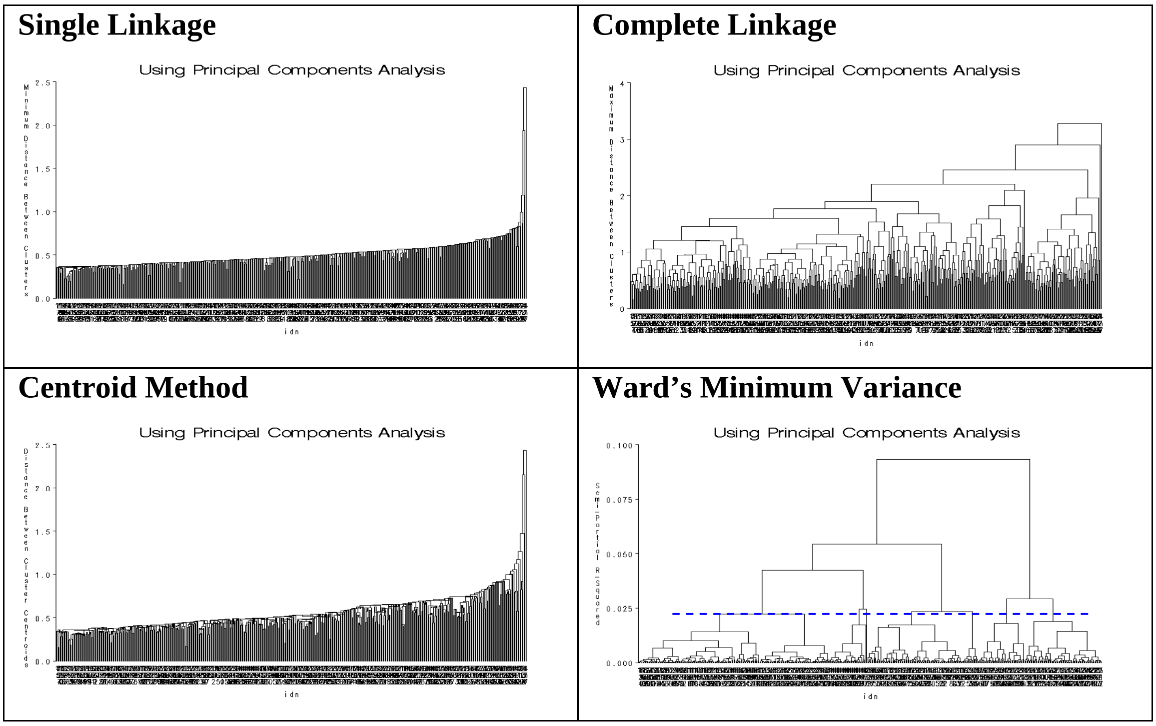
## C. Clustering Results

### C.1. Hierarchical Method

In the hierarchical clustering procedure, all the variables were clustered using several distances allowing the observations to group themselves based on whether it is single linkage (shortest distance), complete linkage (maximum distance), centroid method (cluster centroid) and

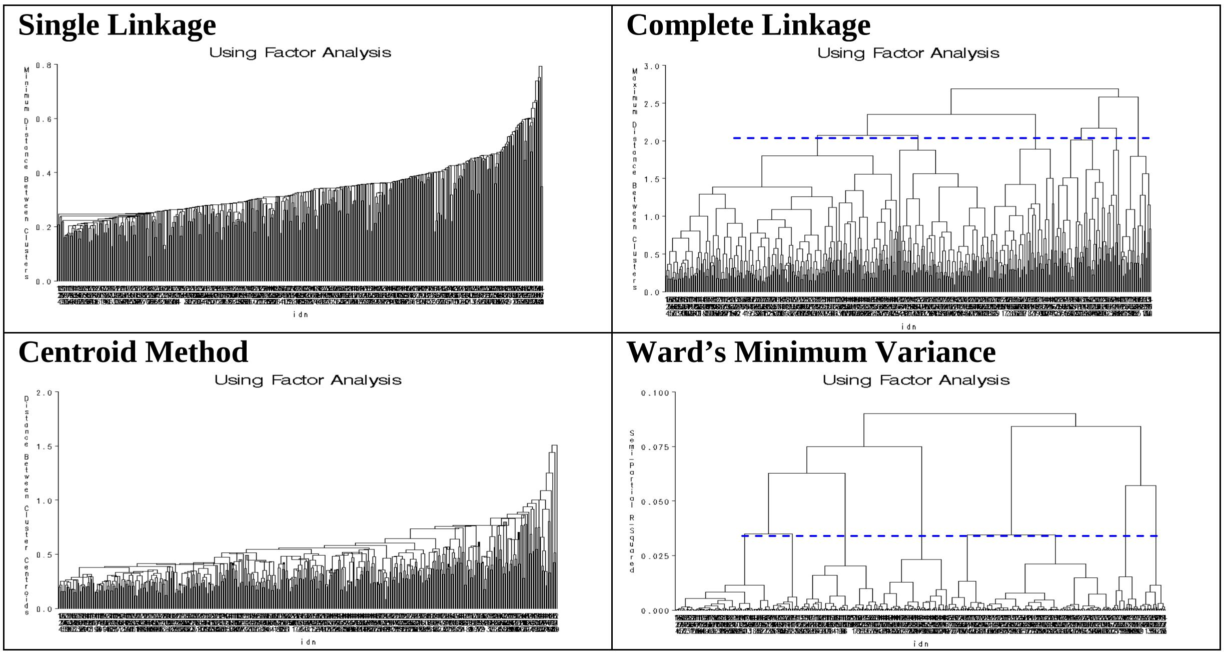
Ward's minimum variance method. The comparisons were undertaken aiming to find parallelism on the suitable number of clusters based on the indication of how the patients were clustered based on the distances from single linkage, complete linkage, centroid method and Ward's minimum variance method.

**Figure1. Hierarchical method using variables from Principal Component Analysis**




Applying the 18 PCs in the hierarchical clustering procedure, the single linkage did not show a good cluster distribution. Likewise for the centroid method, the clustering was not well separated to identify groups. For the complete linkage or the maximum distance a better distribution of clustering was observed when compared to the previous two but it is still not that clear to see a good distribution. Using the Ward's minimum variance method, it showed a more clear clustering procedure than that the single and complete linkages and centroid method. The Ward's distance indicated a suitable 8 clusters to be selected, taking note that the selection of this cluster was based on the proportionality of the distance covered by the cluster along the x-axis, although in this case the proportionality of the cluster is not that obvious.

**Figure 2. Hierarchical method using variables from Factor Analysis**



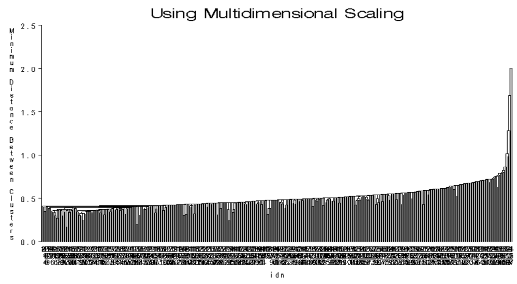

The 6 factors from the factor analysis were then used into the hierarchical clustering procedure. It was observed in the single linkage and centroid method that no identifiable clusters can be generated from these distances. For the complete linkage method it can be seen that there is a better separation of clusters when compared to the previous two distances but considering 7 clusters using this method assigns about 30% of the observations to only one cluster which does not meet the aim of this procedure.  While for the clustering method using the Ward's minimum variance method, 8 clusters may possibly be a suitable number of clusters based on the proportionality of the distance of the cluster to the x-axis.

The MDS procedure generated 13 dimensions which had a stress function equal to about 5%. Looking at the 3 distances namely: single, complete and centroid did not indicate good cluster separation. While for the Ward's minimum variance method, it generated 8 clusters which are somehow proportional.

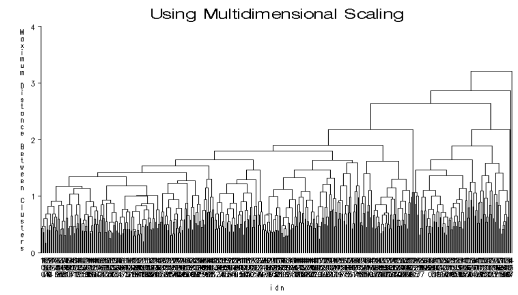
The results of the hierarchical clustering using PCA, FA and MDS variables indicated similar results, but the distances covered by the respective clusters generated by the three methods varies.

**Figure 3. Hierarchical method using variables from Multidimensional Scaling**

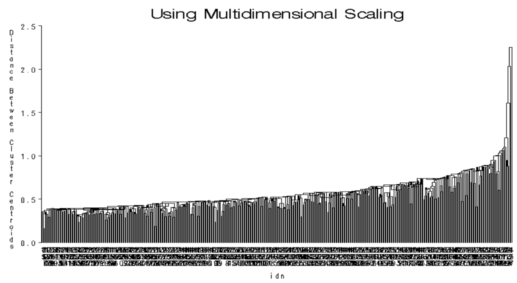
**Single Linkage**



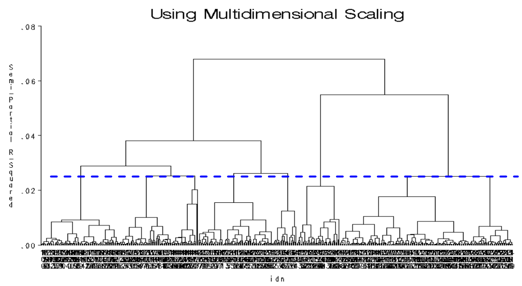

**Complete Linkage**




**Centroid Method**




**Ward's Minimum Variance**




## C.2. Non-Hierarchical Method

The clusters generated for the Non-Hierarchical using the 18 PCs as an input, initial clusters values from 2 until 10 clusters were considered. Taking note of the CC Criterion and the *r-square* values, since these values would indicate whether the number of clusters that was developed in the cluster procedure is good or not. Good clustering can be verified by plotting the CC Criterions for several groups and then see whether there are peaks which might indicate a good number of clusters. On the other hand, as mentioned in the guidelines of choosing good number of clusters, aside from noting the peak, the values of the CC criterion should be in the range of 2-3.

Looking at the plot of the CC Criterion (Figure 4) the values are all negative but among these negative values a peak was observed in the 7 clusters which has an equivalent *r-square* value of 28.55 percent (Figure 5).

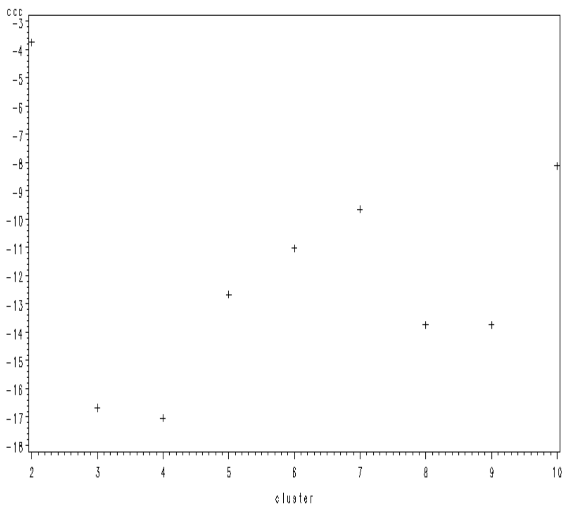



**Figure 4. CC Criterion for Principal Components Analysis on the gait analysis of CP patients.**

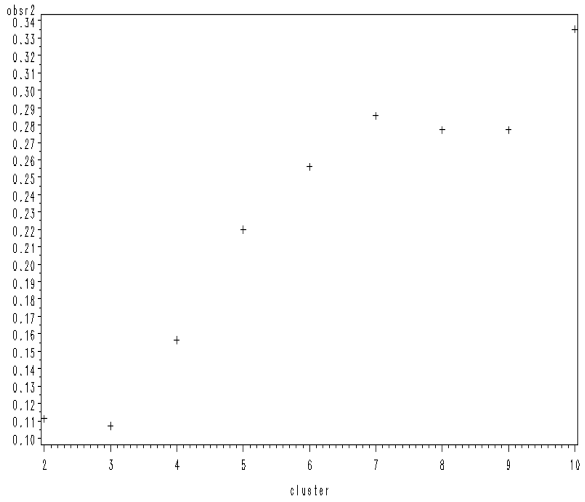



**Figure 5. R-square value for Principal Components Analysis on the gait analysis of CP patients.**

Since the clusters generated from both hierarchical and non-hierarchical methods are not very stable in the sense that they did not satisfy the requirements in arriving at a suitable number of clusters a cross classification was done to be able to pinpoint how much discrepancy is present between the two clustering procedures.

**Table 3. Cluster Summary for Non-Hierarchical Method using variables from PCA.**

| Cluster | Frequency | Nearest Cluster | Distance Between Cluster Centroids |
|---|---|---|---|
| **1** | 1 | 5 | 20.7269 |
| **2** | 44 | 6 | 4.7335 |
| **3** | 1 | 6 | 17.8319 |
| **4** | 98 | 7 | 4.336 |
| **5** | 33 | 4 | 4.7618 |
| **6** | 107 | 7 | 3.8046 |
| **7** | 176 | 6 | 3.8046 |

**Table 4. Cross Classification of Hierarchical and Non-Hierarchical Methods using PCA**

| Non-Hierarchical Method | Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **2** | 9 | 0 | 13 | 5 | 1 | 1 | 15 | 0 | 44 |
| **3** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| **4** | 11 | 7 | 5 | 58 | 0 | 17 | 0 | 0 | 98 |
| **5** | 0 | 0 | 2 | 1 | 5 | 25 | 0 | 0 | 33 |
| **6** | 11 | 4 | 68 | 8 | 13 | 2 | 1 | 0 | 107 |
| **7** | 89 | 59 | 6 | 5 | 16 | 1 | 0 | 0 | 176 |
| **Total** | 120 | 70 | 94 | 77 | 36 | 46 | 16 | 1 | 460 |

From the table above it can be observed that cluster 4 of the hierarchical clustering and cluster 4 of the non-hierarchical clustering both methods have identified 58 patients to belong in the same group. This is approximately equivalent to 75 percent and 60 percent parallelism of the hierarchical and non-hierarchical methods, respectively. Taking note that the cluster numbers in both procedure are mere representations of the generated clusters and it doesn't follow that cluster 1 in hierarchical method also represent same observation in cluster 1 from non-hierarchical method. Also 25 observations were classified in the same group which has 54 and 76 percent matching for cluster 6 of the hierarchical cluster method and cluster 5 in the non-hierarchical cluster method, respectively. It can also be noted that cluster 1 and 2 of the hierarchical method which has 89 and 59 observations, respectively, if joined together may yield better clustering that can parallel the cluster 7 of non-hierarchical method. It was also observed that the hierarchical and non-hierarchical clustering method identified cluster 8 and cluster 1, respectively, that has one observation which is a cluster in itself. Although in cluster 3 of the non-hierarchical clustering procedure it had another one observation which is a cluster in itself. This one observation alone in one cluster should be treated with caution since this observation could actually be behaving very differently from the other characteristics of observations in the other clusters and thus should not be treated as an outlier. This result could be explained by the large distance of the cluster centroids to their respective nearest cluster as noted in Cluster Summary table (Table 4).

Using the factors obtained in FA to initialize in the k-means clustering procedure, several clusters were tried and the criterions and *r-square* value which would aid in finding a good number of clusters were checked. The graph of the CC Criterion indicates a peak in cluster 8 (Figure 6) which in turn has an equivalent *r-square* value of 48 percent (Figure 7).

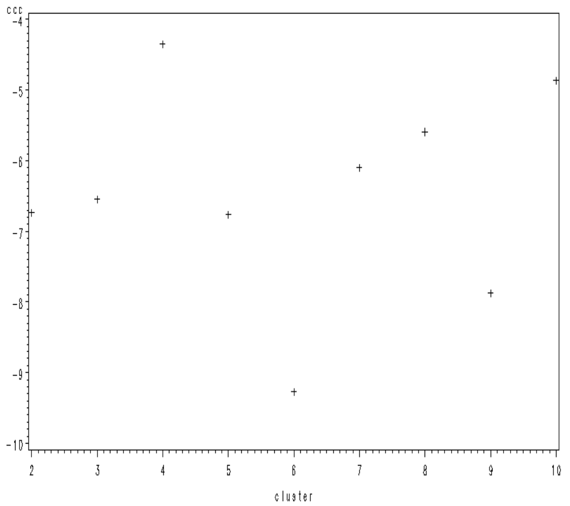



**Figure 6. CC Criterion for Factor Analysis on the gait analysis of CP patients.**

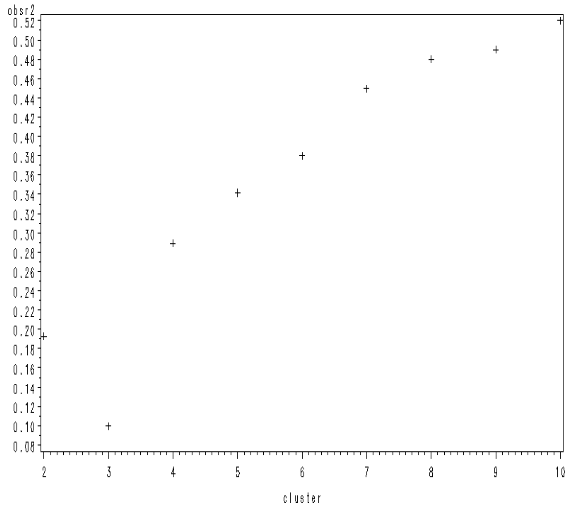



**Figure 7. R-square value for Factor Analysis on the gait analysis of CP patients.**

The negative values of the CC Criterion may indicate that even if a peak is present, the number of cluster may not be that reliable since the CCC requires the value to be between 2 and 3 to be acknowledged as suitable number of clusters.

**Table 5. Cluster Summary for Non-Hierarchical Method using variables from PCA.**

| Cluster | Frequency | Nearest Cluster | Distance Between Cluster Centroids |
|---|---|---|---|
| **1** | 127 | 6 | 1.728 |
| **2** | 49 | 4 | 1.9495 |
| **3** | 54 | 2 | 2.0177 |
| **4** | 47 | 2 | 1.9495 |
| **5** | 46 | 6 | 2.2983 |
| **6** | 88 | 1 | 1.728 |
| **7** | 27 | 2 | 2.6535 |
| **8** | 22 | 3 | 2.4895 |

A cross classification was made to be able to check the consistency of the grouping of patients utilizing the hierarchical and non-hierarchical clustering methods.

**Table 6. Cross Classification of Hierarchical and Non-Hierarchical Methods using FA.**

| **Non-Hierarchical Method** | **Hierarchical Method** | | | | | | | | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 96 | 24 | 0 | 4 | 0 | 3 | 0 | 0 | 127 |
| **2** | 5 | 0 | 10 | 11 | 0 | 21 | 2 | 0 | 49 |
| **3** | 16 | 0 | 3 | 1 | 0 | 7 | 19 | 8 | 54 |
| **4** | 0 | 1 | 0 | 46 | 0 | 0 | 0 | 0 | 47 |
| **5** | 0 | 1 | 43 | 0 | 1 | 1 | 0 | 0 | 46 |
| **6** | 0 | 60 | 9 | 11 | 1 | 0 | 6 | 1 | 88 |
| **7** | 1 | 3 | 0 | 1 | 22 | 0 | 0 | 0 | 27 |
| **8** | 0 | 0 | 1 | 0 | 1 | 3 | 0 | 17 | 22 |
| **Total** | 118 | 89 | 66 | 74 | 25 | 35 | 27 | 26 | 460 |

From the table above, there were a several clusters from the hierarchical and the non-hierarchical clustering methods that showed parallelism in how the patients were classified. There were 22 observations classified in the same group which has 88 and 81 percent matching for cluster 5 the hierarchical cluster method and cluster 7 in the non-hierarchical cluster method, respectively, which indicated very good cross classification. Similar observations were noted for clusters 1, 2, 3, 4 and 8 of the hierarchical clustering methods which showed parallelism ranging from 60% to 90% with clusters 1, 4, 5, 6 and 8 of the non-hierarchical clustering method. This can indicated that the cluster analysis using the factor yield better results utilizing the two clustering procedures.

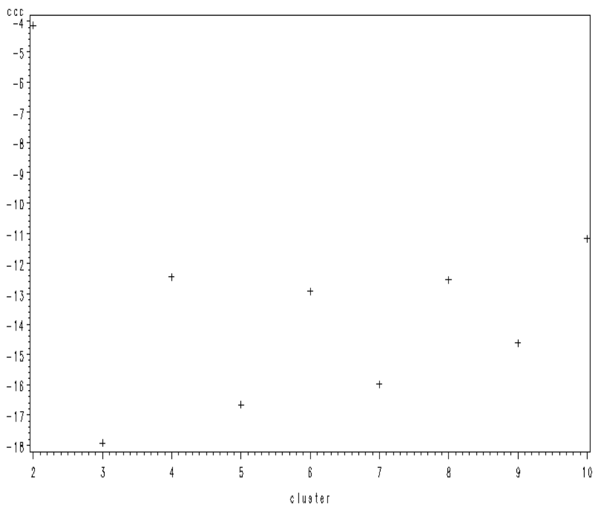



**Figure 8. CC Criterion for Multidimensional Scaling on the gait analysis of CP patients.**

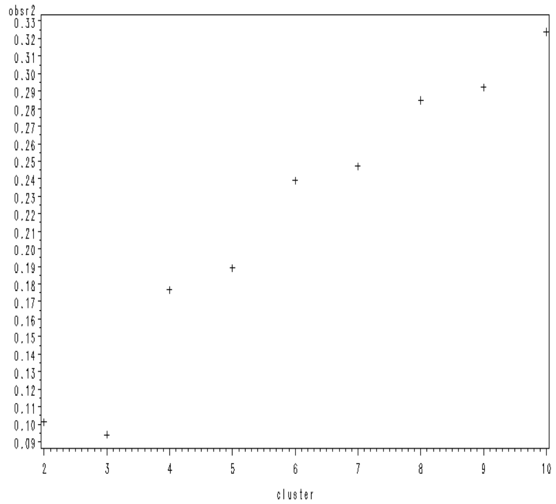



**Figure 9. R-square value for Multidimensional Scaling on the gait analysis of CP patients.**

Utilizing the 13 Dimensions generated from the MDS procedure with cluster values initialized from 2-10 was used to plot the CC Criterion (Figure 5) although the values are negative a peak was observed in the 8 clusters which has an equivalent *r-square* value of 28.48 % (Figure 6).

**Table 7. Cluster Summary for Non-Hierarchical Method using variables from MDS.**

| Cluster | Frequency | Nearest Cluster | Distance Between Cluster Centroids |
|---|---|---|---|
| **1** | 43 | 7 | 4.9299 |
| **2** | 2 | 4 | 11.5038 |
| **3** | 1 | 7 | 21.3212 |
| **4** | 79 | 5 | 4.6143 |
| **5** | 184 | 6 | 4.4912 |
| **6** | 78 | 5 | 4.4912 |
| **7** | 72 | 1 | 4.9299 |
| **8** | 1 | 6 | 17.8536 |

**Table 8. Cross Classification of Hierarchical and Non-Hierarchical Methods using MDS.**

| Non-Hierarchical Method | Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 0 | 13 | 0 | 9 | 0 | 2 | 19 | 0 | 43 |
| **2** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| **3** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| **4** | 9 | 1 | 43 | 4 | 2 | 2 | 3 | 15 | 79 |
| **5** | 57 | 94 | 11 | 6 | 1 | 14 | 1 | 0 | 184 |
| **6** | 5 | 6 | 18 | 1 | 21 | 25 | 2 | 0 | 78 |
| **7** | 5 | 13 | 1 | 22 | 0 | 7 | 17 | 7 | 72 |
| **8** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| **Total** | 76 | 127 | 73 | 42 | 26 | 51 | 42 | 23 | 460 |

From the cross classification in Table 8, very few parallelism was observed. Although it was noted in the non-hierarchical method that two clusters (clusters 3 and 8) had only one observation in the cluster, which may indicate that these two clusters are behaving differently from the characteristics of the other cluster and should not be treated as an outlier. This result could be explained by the large distance of the cluster centroids to their respective nearest cluster (Table 7).

It was also observed that the highest proportion of observation clustered together using the two clustering procedures was noted in cluster 3 from the hierarchical method and cluster 4 from the non-hierarchical which has about 50% parallelism. This can indicate that the MDS procedure is not a very good method in this case since the output of the clusters was not that good and may indicate that the PCA and FA may be better variable reduction techniques that can yield better output when used to initialize the hierarchical and non-hierarchical clustering techniques.

The clusters indicated in the non-hierarchical method were noted based on the criterions specified in the previous chapter. Taking into consideration that these said criterions are mere indicators and not absolute rule in deciding the final number of cluster.

### C.3. Normal Mixture Modeling

Using the 18 PCs as an input the normal mixture model, several numbers of clusters were computed as indicated in the figure below (cluster 2 until 10). With the aid of the BIC criterion a best model of 3 clusters which has the characteristic of a diagonal distribution, volume is variable, shape is equal and the orientation is coordinate axes was suggested.

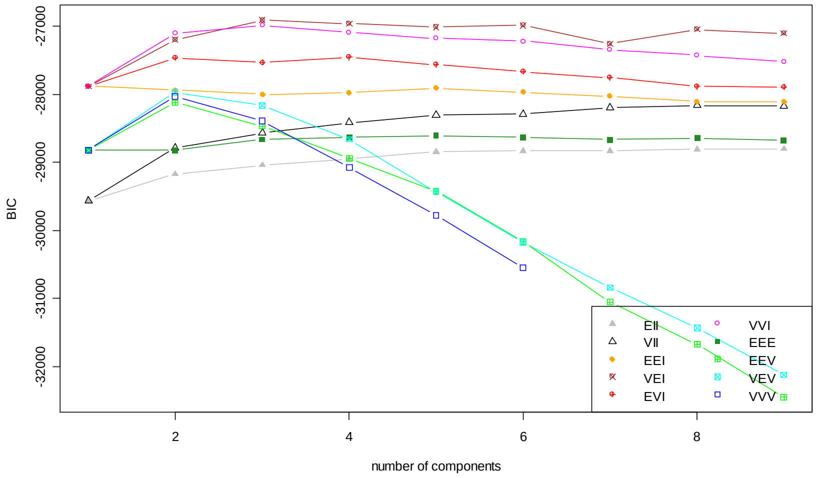



**Figure 10. BIC values for 10 components using PCA variables**

Noting the characteristics of the 3 clusters from the 18 PCs, a plot of the first 2 PCs is shown below. It was noted that the clusters were overlapping each other (Figure 11) also plotting the observation that has uncertainty to belong from one cluster to another (Figure 12) were scattered more on the outermost The ellipses superimposed on both the classification and uncertainty are based on the covariances of the clusters.

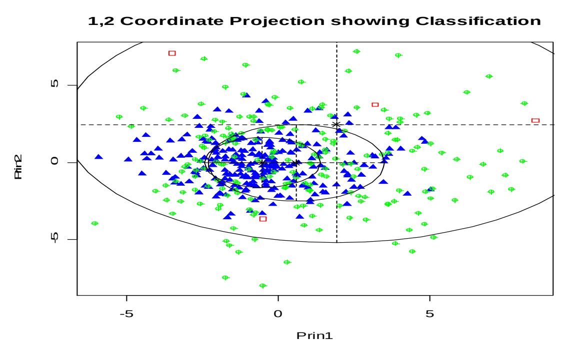



**Figure 11. Plot of cluster classification for PC variables 1 and 2.**

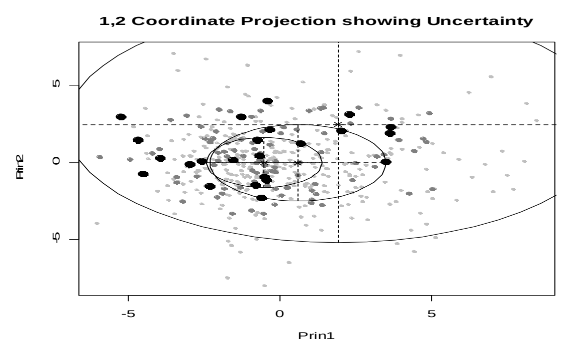



**Figure 12. Plot of PC1 and PC2 on the observations' cluster uncertainty.**

Utilizing the 6 factors from FA generated 3 clusters with characteristics of a spherical distribution, volume is variable and shape is equal was noted.

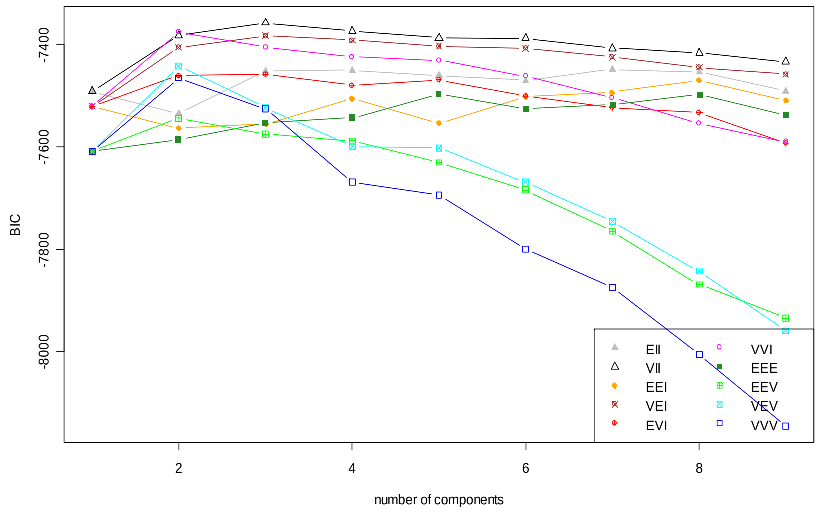



**Figure 13. BIC values for 10 components using FA variables**

Plotting the 3 clusters for factors 1 and 2 showed overlapping clusters but not very diverse like the mixtures using PCA. The ellipses superimposed on both the classification and uncertainties are based on the covariances of the clusters. In addition, it was also noted that most of the observations that has uncertainty is contained inside the cluster ellipses.

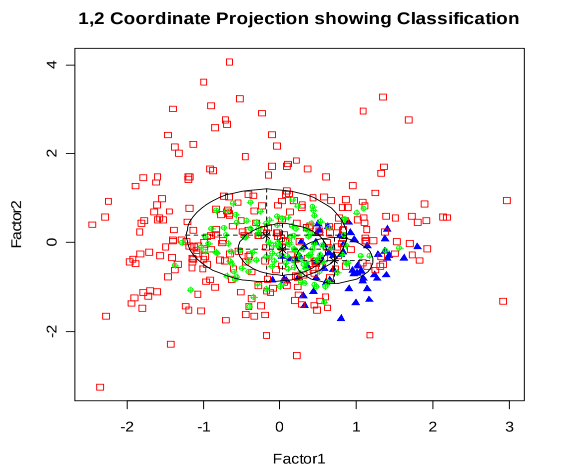



**Figure 14. Plot of cluster classification for FA variables 1 and 2.**

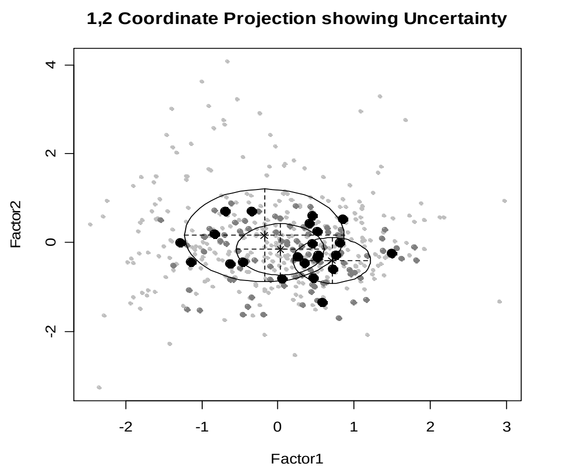



**Figure 15. Plot of factor 1 and factor 2 on the observations' cluster uncertainty.**

The 13 dimensions from the MDS procedure was used as an input in the normal mixture modeling wherein the characteristic was described to have a diagonal distribution, volume is variable, shape is equal and the orientation is coordinate axes. The characteristic of the cluster using the MDS is similar when the 18 PCs are used. The difference is that the MDS generated only 2 clusters for the CP data.

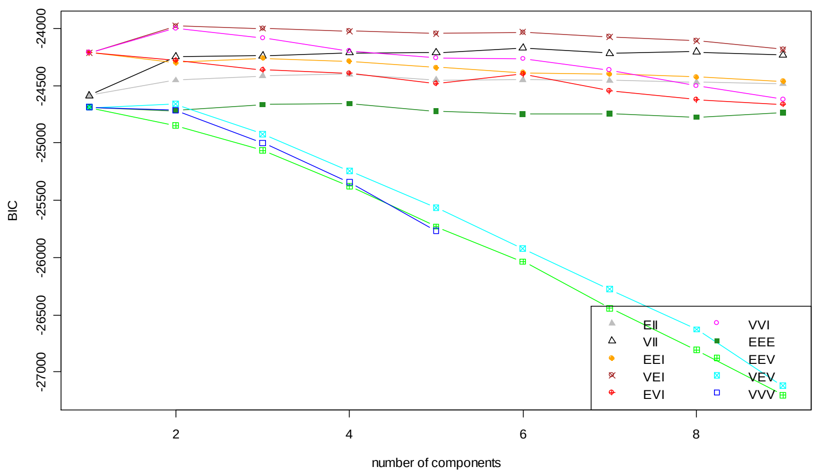



**Figure 16. BIC values for 10 components using MDS variables**

From the figure below (Figure 17) the observations were superimposed together with the ellipses. Taking note that the ellipses superimposed on both the classification and uncertainties are based on the covariances of the clusters. The uncertain observations were also plotted (Figure 18) but were observed to be scattered outside the ellipses.

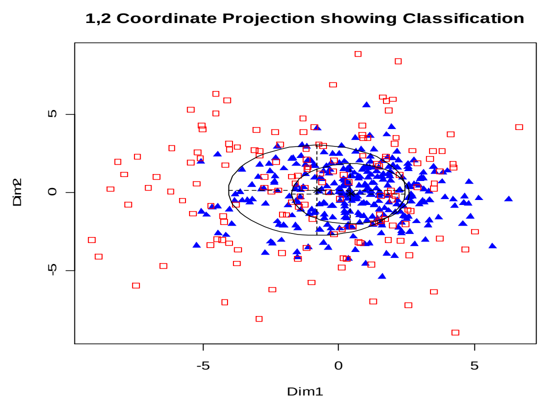



**Figure 17. Plot of cluster classification for MDS dimensions 1 and 2.**

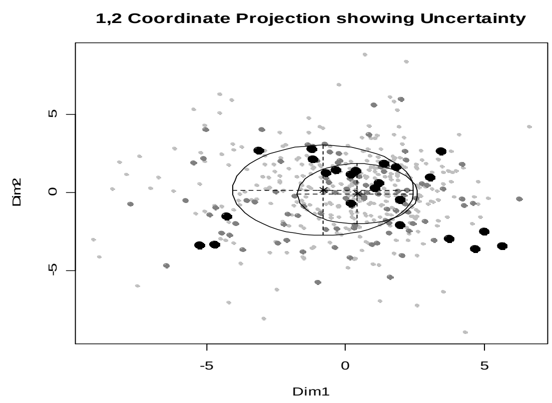



**Figure 15. Plot of dimensions 1 and 2 on the observations' cluster uncertainty.**

Considering the clusters generated from the normal mixture model, cross classifications against the hierarchical and the non-hierarchical clustering methods (Tables 9-14) were shown to be able to verify which observations were classified in one group across the different clustering methods used in this study.

From Table 9, comparing clusters from the normal mixture model (cluster 1) and the non-hierarchical clustering method (cluster 7) have cross classified 146 observations to have the same gait patterns which is about 55% and 83%, respectively. Table 10 compares the normal mixture model with the hierarchical clustering algorithms and has identified about 81% and 37%, respectively, where observations that have similar gait patterns.

**Table 9. Cross Classification of Normal Mixture Model and Non-Hierarchical Methods using variables from Principal Components Analysis**

| Normal Mixture Model | Non-Hierarchical Method | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | |
| **1** | 0 | 12 | 0 | 51 | 6 | 49 | 146 | 264 |
| **2** | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 4 |
| **3** | 0 | 31 | 0 | 47 | 26 | 58 | 30 | 192 |
| **Total** | 1 | 44 | 1 | 98 | 33 | 107 | 176 | 460 |

**Table 10. Cross Classification of Normal Mixture Model and Hierarchical Methods using variables from Principal Components Analysis**

| Normal Mixture Model | Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 97 | 61 | 41 | 36 | 16 | 9 | 4 | 0 | 264 |
| **2** | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 4 |
| **3** | 23 | 9 | 52 | 41 | 19 | 36 | 12 | 0 | 192 |
| **Total** | 120 | 70 | 94 | 77 | 36 | 46 | 16 | 1 | 460 |

Comparing the results using the PC variables across the three clustering procedures, there were 76 observations that were identified to belong to one cluster regardless of the clustering method used. This may indicate that these observations actually have similar gait patterns.

From Table 11 , comparing clusters from the normal mixture model (cluster 3) and the non-hierarchical clustering method (cluster 1) have cross classified 111 observations to have the same gait patterns which is about 88% and 87%, respectively. In addition, cluster 1 of normal mixture model and cluster 7 of the non-hierarchical clustering method have identified 63 observations that have similar gait patterns. While Table 12, compares the normal mixture model (cluster 3) with the hierarchical clustering method (cluster 1) have identified 92 observations that have similar gait patterns.

**Table 11. Cross Classification of Normal Mixture Model and Non-Hierarchical Methods using variables from Factor Analysis**

| Normal Mixture Model | Non-Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 0 | 0 | 0 | 1 | 0 | 63 | 0 | 0 | 64 |
| **2** | 16 | 44 | 47 | 46 | 46 | 21 | 27 | 22 | 269 |
| **3** | 111 | 5 | 7 | 0 | 0 | 4 | 0 | 0 | 127 |
| **Total** | 127 | 49 | 54 | 47 | 46 | 88 | 27 | 22 | 460 |

**Table 12. Cross Classification of Normal Mixture Model and Hierarchical Methods using variables from Factor Analysis**

| Normal Mixture Model | Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| **1** | 0 | 49 | 4 | 8 | 0 | 0 | 3 | 0 | 64 |
| **2** | 26 | 13 | 62 | 63 | 25 | 32 | 22 | 26 | 269 |
| **3** | 92 | 27 | 0 | 3 | 0 | 3 | 2 | 0 | 127 |
| **Total** | 118 | 89 | 66 | 74 | 25 | 35 | 27 | 26 | 460 |

Using the factors as an input in the clustering procedures, two clusters of observations were identified across the three clustering methods used in this study. One composed of 84 observations while the other had 49 observations, these two clusters that were cross classified using the three clustering procedures are actually distinct observations that have identical gait patterns.

Cluster 1 of normal mixture model and cluster 5 of the non-hierarchical clustering method have identified 172 observations (Table 13) that have similar gait patterns which is about 55% and 93%, respectively. While Table 14, compares the normal mixture model (cluster 1) with the hierarchical clustering method (cluster 2) have identified 92 observations that have similar gait patterns.

**Table 13. Cross Classification of Normal Mixture Model and Non-Hierarchical Methods using variables from Multidimensional Scaling**

| Normal Mixture Model | Non-Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 1 | 18 | 0 | 0 | 37 | 172 | 44 | 39 | 0 | 310 |
| 2 | 25 | 2 | 1 | 42 | 12 | 34 | 33 | 1 | 150 |
| Total | 43 | 2 | 1 | 79 | 184 | 78 | 72 | 1 | 460 |

**Table 14. Cross Classification of Normal Mixture Model and Hierarchical Methods using variables from Multidimensional Scaling**

| Normal Mixture Model | Hierarchical Method | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 1 | 67 | 120 | 37 | 32 | 13 | 34 | 4 | 3 | 310 |
| 2 | 9 | 7 | 36 | 10 | 13 | 17 | 38 | 20 | 150 |
| Total | 76 | 127 | 73 | 42 | 26 | 51 | 42 | 23 | 460 |

Cross classifying the results of the three clustering methods using the MDS variables have identified 92 observations that have identical gait patterns. This may indicate that these observations maybe be considered to belong to one cluster or group since they are clustered together using three different procedures.

Summarizing the number of clusters obtained from the three cluster procedures and utilizing PCA, FA and MDS variables, Table 15 is presented to be able to compare the clusters obtained from the different procedures. As in the previous analysis the process of determining the number of clusters for the hierarchical method considers proportionality of the observation meaning avoiding bulk in one cluster alone. So the distances on the x-axis which is represented by the *id* are carefully considered and measured. Although these measurements are quite subjective since the cut off decision for a suitable number of cluster depends on the number of *id's* covered by the cluster. While for the Non-hierarchical cluster procedure, the

CC Criterion and the $R^2$ values were used as indicators in suggesting the plausible number of clusters. The normal mixture however deviated from the numbers of cluster from the two methods. This result may be due to the fact the criteria for an observation to belong to one cluster is based on BIC and not distances unlike in the hierarchical and non-hierarchical case. Furthermore, no restriction was imposed specifically on the covariance structure of the clusters, hence yielding less number of clusters as compared to the hierarchical and non-hierarchical cluster methods. In addition, the fewer number of clusters from the normal mixture model result may be because these few clusters are encompassing smaller clusters like in the hierarchical or non-hierarchical clustering techniques (i.e. Cluster 1 from Normal Mixture Model includes clusters 2, 4 and 6 from the non-hierarchical clustering technique).

**Table 15. Summary for the number of clusters by clustering methods and the input variables**

| Variables | Hierarchical Method | Non-Hierarchical (K-means) Method | Normal Mixture Model |
|---|---|---|---|
| **PCA** | 8 | 7 | 3 |
| **FA** | 8 | 8 | 3 |
| **MDS** | 8 | 8 | 2 |

## IV. DISCUSSIONS AND CONCLUSIONS

The study investigates on classifying the gait patterns in children with cerebral palsy. Three clustering methods were employed to group these CP patients namely: hierarchical, non-hierarchical and normal mixture model. Although these cluster methods have the same purpose which is to group the data, these methods have different approaches in clustering the data. The hierarchical cluster method has two approaches, the agglomerative and divisive way. The agglomerative starts the algorithm considering a single observation as a group in itself and then eventually merges that unit to another unit thereby creating a new group. The merging continues until all the units belong to one group. The divisive works the opposite way; first the units are all clustered in a single group and eventually split until all the units are separated from each other. The agglomerative method was used in this study since it has the advantage of being computationally simple.

One of the strengths of the hierarchical cluster method is its ability to cluster the data without the need to initiate a starting value for the cluster, unlike the non-hierarchical which require a starting value for it to start clustering. A strong point of the non-hierarchical method is its ability to be robust against outliers which is a weak point of the hierarchical cluster method. To be able to get the benefit of both techniques the two cluster methods were employed. There are different criterions for each cluster method. The hierarchical cluster method use different distances to which will determine how the data will be clustered (i.e. single linkage, ward's minimum variance method, etc.). While the non-hierarchical cluster method checks the quality of the clusters based the CC criterion and the *r-square* value.

The normal mixture model allows the selection of the parameterization of the model as well as the number of clusters simultaneously using the Bayesian Information Criterion (BIC). In addition, the normal mixture model incorporates the advantages of the hierarchical agglomerative clustering technique which yield good partition even and doesn't need to be initialized; and the EM algorithm which in this case makes probabilistic rather than deterministic assignments of the observation to cluster centers with the aid of the BIC.

Variable reduction techniques were used in this study specifically, the Principal Component Analysis, Factor Analysis and the Multidimensional Scaling. The three procedures proved very useful since it reduces the 46 gait parameters into a fewer dimensions. Aside from the benefit of reducing the variables, the PCA, FA and MDS in turn transform the original variables into uncorrelated new variables which are crucial in the non-hierarchical cluster method. In addition, the procedure "mclust" limits the number of variables as an input in the procedure to only 26 variables and taking note that this study has 46 variables, hence when variable reduction methods were used all transformed variables from the 3 procedures were all below 26 variables.

There were observations that were classified in similar clusters with the three clustering procedures used in this study. First, using the PC variables across the three clustering procedures, there were 76 observations that were identified to belong to one cluster regardless of the clustering method used. This may indicate that these observations actually have similar gait patterns. Utilizing the factors as an input in the clustering procedures, two clusters of observations were identified across the three clustering methods used in this study. One composed of 84 observations while the other had 49 observations, these two clusters that were cross classified using the three clustering procedures are actually distinct observations that have identical gait patterns. Cross classifying the results of the three clustering methods using the MDS variables have identified 92 observations that have identical gait patterns. This may indicate that these observations maybe be considered to belong to one cluster or group since they are clustered together using three different procedures.

To summarize, based on the results of the three clustering techniques utilized in this study the methods that used variables from the FA procedure yield better results in terms of the criterions and more specifically indicated more parallelism in the cross classification. Although the clustering procedure that used the FA technique proved better than the PCA and MDS it does not however indicate good clustering of the CP data since the criterions specifically for the hierarchical and non-hierarchical were not fully satisfied in terms of the proportionality, CC Criterion and the $R^2$. While for the normal mixture, fewer clusters were observed and were noted to have some overlap among them, which makes it difficult to find distinct clusters to indicate good groupings of the CP data. However, cross classification of the three clustering procedures used in this study have identified two clusters (composed of 84 and 49 observations) that may be indicators of a good grouping (CP patients share very similar characteristics) since these observations were clustered together despite the clustering technique used.

## REFERENCES


C. Fraley and A. E. Raftery, MCLUST: Software for model-based cluster analysis, Journal of Classification, Vol. 16, pages 297-206 (1999).

C. Fraley and A. E. Raftery, Model-based clustering, discriminant analysis, and density estimation, Journal of the American Statistical Association, Vol. 97, pages 611-631 (2002).

C. Fraley and A. E. Raftery, Enhanced Software for model-based clustering, discriminant analysis, and density estimation: MCLUST, Journal of Classification, Vol. 20, pages 263-286 (2003).

C. Fraley and A. E. Raftery, MCLUST: Software for model-based clustering, discriminant analysis, and density estimation, Technical Report No. 415, Department of Statistics, University of Washington, October 2002.

C. Fraley and A. E. Raftery, MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering, Technical Report No. 504, Department of Statistics, University of Washington, September 2006.

C. Fraley and A. E. Raftery, Model-based Methods of Classification: Using the mclust Software in Chemometrics, Journal of Statistical Software, Vol. 18, Issue 6 (2007).

Everitt, B., Cluster Analysis (1980)
2nd edition Halsted Press

Hair, J., et al, Multivariate Data Analysis with Readings (1995)
4th Edition, Prentice Hall International Editions

Hastie, T, Tibshirani, R. and Friedman, J., The Elements of Statitiscal Learning – Data Mining, Inference and Prediction., Springer Series in Statistics (2001)

J. D. Banfield and A. E. Raftery, Model-based Gaussian and non-Gaussian Clustering, Biometrics, Vol. 49, pages 803-821 (1993).

Johnson, R.A and Wichern, D.W, Applied Multivariate Statistical Analysis (2002)
5th Edition , Prentice Hall

Kruskal, J. B., and Wish, M. (1978), *Multidimensional Scaling*, Sage University Paper series on Quantitative Application in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications.

Molenberghs, G., Multivariate Data Analysis Course Notes (2007)
Universiteit Hasselt, Belgium

Verbeke, G., Advanced Modeling Techniques Course Notes (2007)
Universiteit Hasselt, Belgium

Verbeke G, Molenberghs G, Correlated and Multivariate Data Course Notes (2006)
Universiteit Hasselt, Belgium

**Websites**

Work and power in hemiplegic cerebral palsy gait. (1990)
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9183258&dopt=Citation

Common gait abnormalities of the knee in cerebral palsy (1993)
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=8458127&dopt=Citation

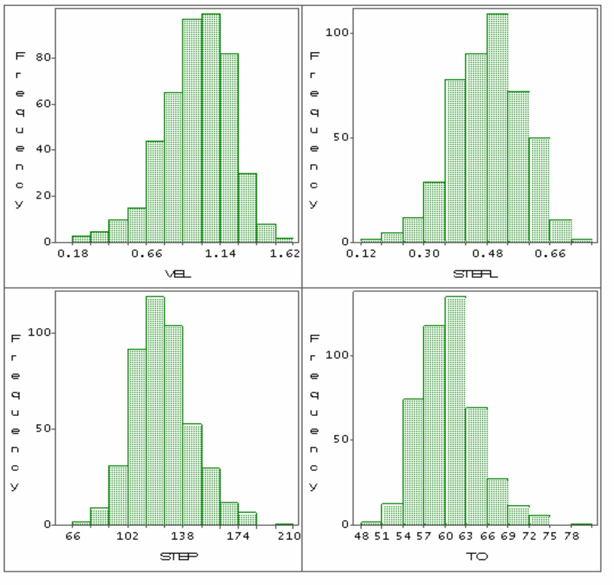
Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer. O'Byrne JM, Jenkinson A, O'Brien TM. Department of Orthopaedic Surgery, Harborview Medical Centre, Seattle, WA 98104-2499, USA.

http://www.originsofcerebralpalsy.com/02-forms/04-spastic.html
http://en.wikipedia.org/wiki/Cerebral_palsy
http://support.sas.com/faq/008/FAQ00893.html
http://www.ucppgmc.com/research.html
http://adc.bmjjournals.com/cgi/content/full/90/5/474
http://www.cerebralpalsy.org/what-is-cerebral-palsy/life-expectancy/

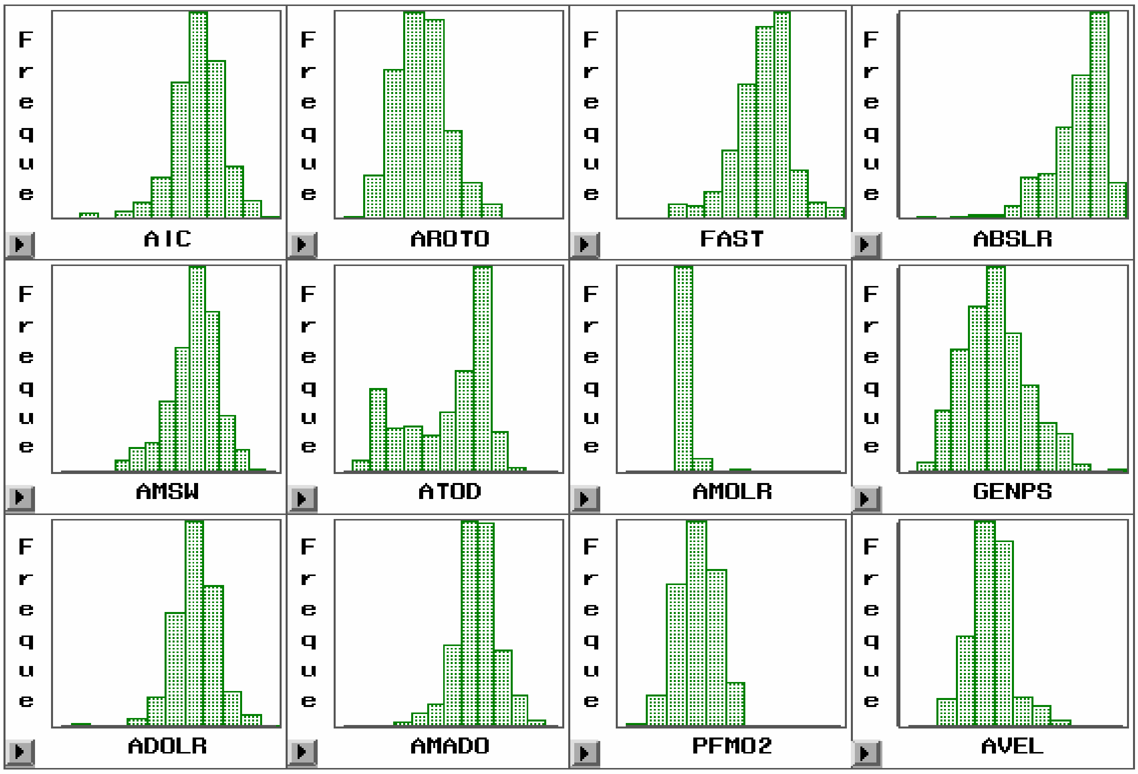
# APPENDIX

## Bar graphs for 46 gait patterns of CP patients.

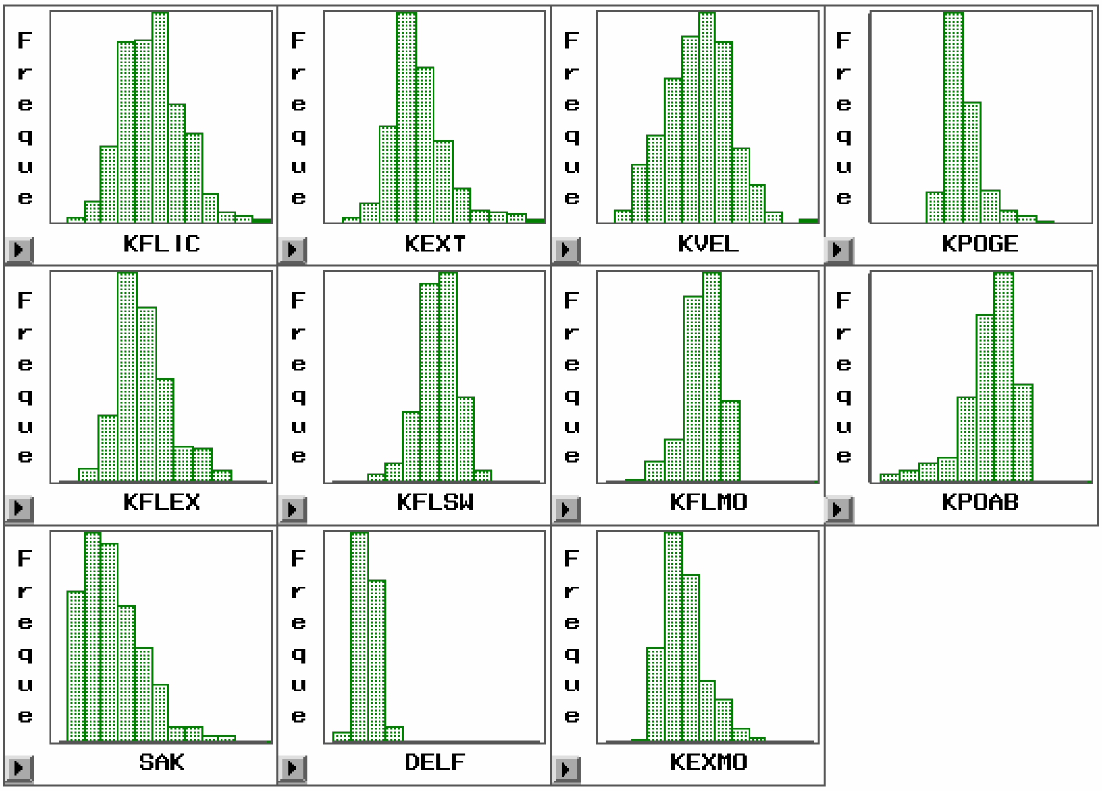
Time and Distance Parameters



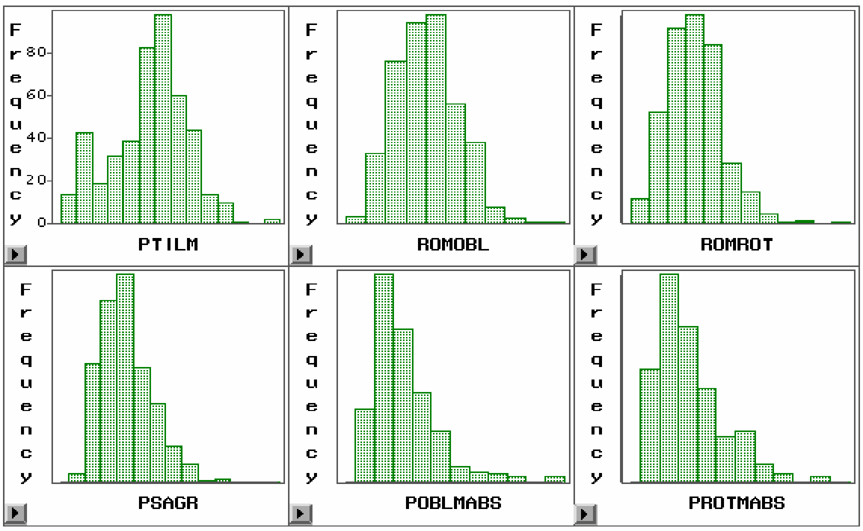

### Ankle and Foot Measurements



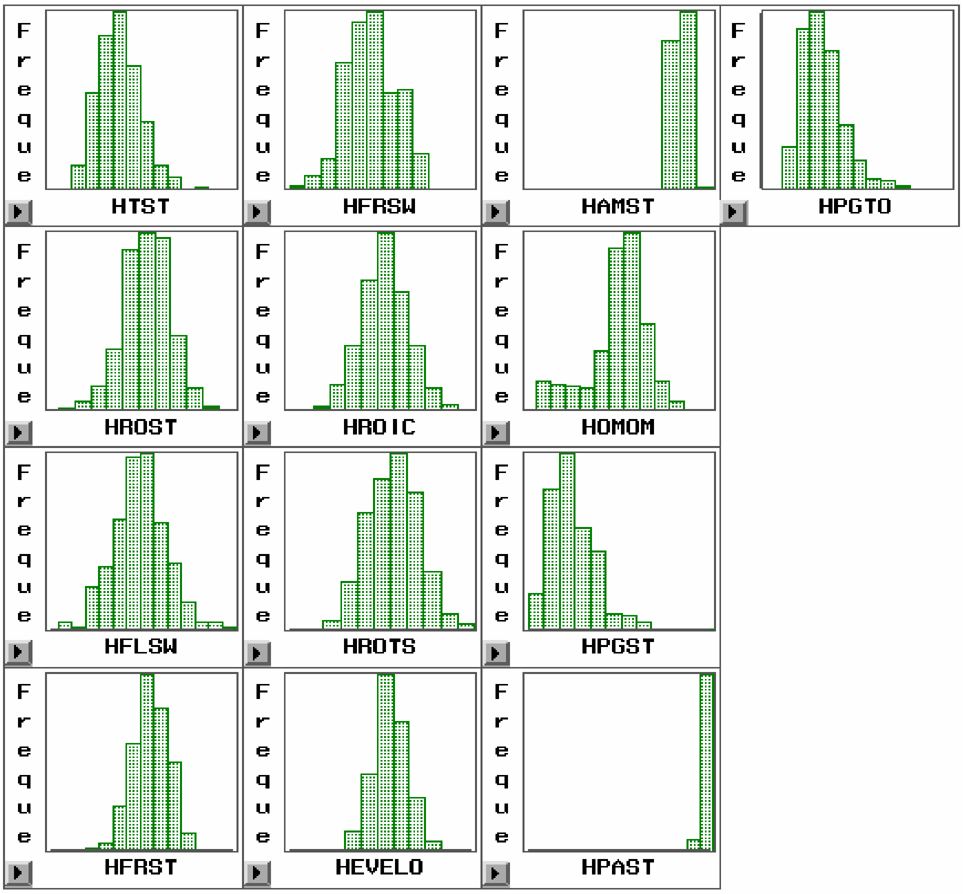
Knee Measurements




**Pelvic Measurements**



## Hip Measurements



# Auteursrechterlijke overeenkomst

*Opdat de Universiteit Hasselt uw eindverhandeling wereldwijd kan reproduceren, vertalen en distribueren is uw akkoord voor deze overeenkomst noodzakelijk. Gelieve de tijd te nemen om deze overeenkomst door te nemen, de gevraagde informatie in te vullen (en de overeenkomst te ondertekenen en af te geven).*

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
**Objective Classification of Gait Patterns in Children with Cerebral Palsy**
Richting: **Master of Science in Biostatistics** Jaar: **2007**
in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Ik ga akkoord,

**Mervic Azupardo**

Datum: **26.08.2007**