

# Socio-Economic and Demographic Classification of HIV/AIDS Patients: Using Cluster Analysis

R.H. R. Bakari<sup>1</sup> B. Bata<sup>2</sup> E. Sambo<sup>3</sup>

1 Dept of Math and Statistics University of Maiduguri Borno State Nigeria

2 NIMC MAIDUGURI (National identity management commission Borno state Nigeria)

3 RAMAT POLYTECHNIC MAIDUGURI Borno state Nigeria

## Abstract

HIV/AIDS has been spreading in an alarming rate since the beginning of the pandemic. It is estimated that at least half of the affected ones are between 15-24 years. These have called for a concerted effort to save this segment of the population for many reasons. Against this background the study was aimed at classifying socio economic and demographic several of 1500 patients, registered with the university of Maiduguri teaching hospital from 2009 to 2013. Cluster analysis was employed on the set data. The distance which shows similarity between the variable using the coefficient distance was calculated. Dendogram were formed for agglomerative schedule. It was observed that the dendogram formed by the various methods differs according to their coefficient. Even though the single, complete and ward follow similar pattern of cluster except the centroid method, which differ and gave different pattern.

**Keywords:** Agglomerative, Cluster Analysis, Dendogram, and Hierarchical

## Introduction

Clustering techniques (one of the multivariate approach), have been applied to a wide variety of research problems. For examples in the field of medicine, clustering of disease, cures for diseases, system of disease have lead to very useful taxonomies. In archeology, researchers have attempted to establish taxonomy of stone tool, general objects by applying clustering techniques. Cluster analysis seeks to partition a set of individual into some form of natural grouping. If any, it is one tool of explanatory data analysis that attempt to assess the interaction among patterns by organizing the patterns into groups or cluster. Such that patterns within cluster are more similar to each other than are pattern belonging to different clusters. The application of multivariate techniques has been in the behavioral and biological sciences. However, interest in multivariate methods has now spread to numerous other fields of investigation. For example, in education, chemistry, physics, geology, engineering, law, business, literature, religion, public broadcasting, nursing, biology, and many other fields, Morrisson (1990). Multivariate statistical analysis is concerned with data collected on several dimensions of the same individual, such observation in the social behavioral life and medical science. It therefore, helps the researcher to summarize the data and reduce the numbers of variables necessary to describe it.

HIV/AIDS (Human immunodeficiency virus/acquired immunodeficiency syndrome) is a disease of human immune-system caused by infection with HIV. The terms HIV/AIDS represent the entire range of disease caused by the HIV from early infection to late stage symptoms. During the initial infection, a person may experiences a brief period of influenza like illness. As the illness progress it interferes more and more with the immune system, making the person much likely to get infection. Harigan (1974) As put on by Oppong (1993) until the discovery in 1981, very little was known about HIV/ AIDS, however, retrospective studies have shown its existence much earlier than 1974, although, first case were reported in U.S.A in 1981, in five homosexual who were undergoing treatment for rare lung disease called pneumocystis carini pneumonitis, the isolation of the offending organism in 1980 land down the concept that is a new disease which is quite distinct from other types of immune deficiency disorder known in medicine.

America and Europe left no trace of the specific origin of HIV/AIDS also Green monkey theory, GMT, which indicated Africa was mentioned by scientist in connection with the spreads of AIDS to America and Europe because of the presence of virus (which have similar identity with the HIV in the African green monkeys. The greatest human fear of the health problem today is HIV/AIDS epidemic. It is a global pandemic. As of 2012 approximately 35.5 million people have HIV worldwide, it resulted in about 1.6 million deaths in 2012, down from a peak of 2.2 million in 2005.

HIV/AIDS is transmitted primarily via unprotected sexual intercourse (including anal and oral sex), contaminated blood transfusion, hypodermic needles, and from mother to child during pregnancy, delivery, or breastfeeding.

The term HIV/AIDS represents the entire range of disease caused by the human immunodeficiency virus from early infection to late stage symptoms. During the initial infections, a person may experience a brief period of influenza- like illness. This is typically followed by a prolonged period without symptoms. As the illness progresses, it interfere more and more with the immune system, making the person much more likely to get infections, including opportunistic infection and tumors that do not usually affect people who have been

working with the immune systems. Unaided (2004)

Thus the following clinical features considerable lose of weight and muscle wasting, frequent fever, general malaise, chronic diarrhea, aggressive Kaposi's sarcoma and different level of tuberculosis. Bennet (2004)

Public enlightenment by organization and government to mount responses to the HIV/AIDS patient particularly in the area of information, education and communication community, mobilization and capacity building programming, such as being faithful to your partner absent sex outside marriage, use condoms lumps.(Andeyi 2006) Multivariate have been applied to wide variety of HIV/AIDS research problem such as logistic multivariate regression model to examine the influence of economic and socio-cultural intervention on adolescents sexually and implications on HIV/AIDS. Pre exposure and post exposure –The pre exposure is treating people with HIV to reduce transmission risk. The post exposure –A course of antiretroviral administered within two –there day after exposure to HIV positive. As of 2013 the prevention regimen recommended in US reduces the risk further Vaccination –As of 2012 Kerrigan and Deamna (2012) ,There is no effective vaccine for HIV or AIDS .A single trial of the vaccine RV144 Published in 2009 found a partial Reduction in the risk of transmission of roughly 30% (Anderson,2012;Yu,M;2101).

AIDS Stigma exist around the world in a variety of way, including

Ostracism, rejection, discrimination and avoidance of HIV infected people. As people perceived to be infected with HIV (FMOH2010).AIDs stigma has been further divided into the following three categories:

- (1) Instrumental AIDS stigma- A reflection of fear and apprehension that are likely to be associated with any deadly and transmutable illness.
- (2) Symbolic Aids stigma: the uses of Hiv or Aids to express attitudes towards the social group or life style perceived to be associated with the disease.
- (3) Courtesy Aids stigma: stigmatization of people connected to the issue of Hiv/Aids or Hiv/positive people.

A lot has been done to solve the problem of HIV/AIDS by medical personal, individual, government, and nongovernmental organization. HIV/AIDS has been like a serious natural disaster that hit the world. Evidence available showed that HIV/AIDS rate is high worldwide. According to the latest report released from WHO as at 2012, 35.5 million people are currently living with HIV/AIDS. The report added that Nigeria rank the third in the number of HIV infection in the world, with 3.4 million people living with it and 2<sup>nd</sup> in Africa country' as from 2012. HIV/AIDS is considered a pandemic a disease outbreak which is present over a large area and is actively spreading.

However, a little has done by the statisticians; there is a need for empirical information on the proper classification of the variables with regards to HIV/AIDS infection. Such data is useful for enhancing the fight against the disease. Therefore, the work is intended to provide the empirical information with a view of classifying the variables on the basis of their similarity

The main aim of this research work is to separate the set of HIV/AIDS patients' variables into groups and study the interrelationship amongst the variables to make a preliminary assessment of their structure. The specific objectives of the research work is to;

- i. Classify the set of variables according to their similarities using different clustering technique.
- ii. Observe the inter correlation among the variables and
- iii. Determine the consistency of the various clustering method used in this research.

This research is intended to use cluster analysis to partition the set of individual variables into some form of "natural grouping" if any. The producer has been used in medical research to classify diseases, cure of disease and symptoms of the diseases that have to lead be very useful method

The data is a secondary data obtained from university of Maiduguri teaching hospital on 1500 patients from 2009-2013( 5years) Due to the sensitive nature of the data, variable like, size of family, religion.

Level of monthly income, parental background and tribe cannot be obtain. The research work is limited only the data obtained as recorded by the university of Maiduguri

The techniques of multivariate analysis are very much an inter disciplinary subjects, covering development in the area of statistics engineering, medicine, artificial intelligence, computer science, psychology and physiology ones in data mining also such as credit scoring, consumer sales analysis, credit card analysis, have attracted considerable research effort, with many method developed and advanced made. (Anderson, 1996)

Cluster analysis is a statistical procedure used to group's variables that share similar characteristics. In biology, cluster analysis is used to classify animals and plants. In medicine, cluster analysis is used to identify **diseases**. In social sciences, this procedure is popular when one wants to understand and discover homogeneous groups. Both cluster analysis and discriminant analysis classify objects, or cases into categories. One difference between these two analyses is 'group membership'. In a discriminant analysis, we already know group membership for the cases; whereas, with a cluster analysis group membership for all cases is unknown. Moreover in a cluster anlysis, we do not know how many groups there are. Thus the goal of cluster analysis is to identify the "clusters" or groups that exist in classifying our cases among similar properties. It follows the age

old saying;”birds of the same feathers tend to ‘cluster’ “together”.

Harding(2014)in his study on symptom clusters in people living with HIV/AIDS attending five palliative care facilities in two sub-saharan african countries: a hierarchical cluster analysis summarize the basic objective of cluster analysis is to discover “natural grouping” one must first developed a quantitative scale and which to measure the association (similarity) between objects. This is the first study of symptom clusters in HIV/AIDS patients receiving palliative care using a cluster analytical approach. Symptom clusters in this patient group could be identified and these are interpretable. These results also indicate that HIV/ AIDS patients who are on ART still show a high prevalence of symptoms causing a high level of distress. Given the high prevalence of psychosocial problems and stigma reported alongside ART, these data also offer important understanding of the co-existence of physical and psychological problems within symptom clusters, and therefore the patient must be assessed holistically. These clusters demonstrate that assessment and treatment planning for symptom clusters should take account of their physical, psychological and social burden. From clinical and public health perspectives, assessment and control for people living with HIV are important because symptom burden is associated with sexual risk taking, poor adherence to ART, treatment switching, viral rebound, poorer quality of life, and suicidal ideation. These findings are useful for directing assessment and care in clinical practice. The data demonstrate that patients in symptom cluster one (i.e. the ‘dermatological-related’ symptom cluster) require the greatest attention to their symptom burden, and that as function declines their burden is likely to increase. Future research should investigate symptom clusters in people living with HIV in different stages of the disease trajectory, and longitudinal study designs may detect changes over time in symptom cluster constructs. Importantly, the distress associated with symptom burden should be managed in routine

Yangi (2014) conduct his research on “The aids epidemic and economic input impact factors in chongqing, china, from 2006 to 2012: a spatial-temporal cluster analysis”and concluded that The Chongqing HIV/AIDS epidemic showed temporal-spatial clustering and was mainly clustered in the mid-western and south-western counties, showing an upward trend over time. The amount of special funds dedicated to AIDS and to the public awareness unit showed positive and negative relationships with HIV/AIDS

Webb (1999) said that the hierarchical clustering procedures are the most economy used means for summarizing data structure. The techniques essentially transform a proximity matrix into dendrogram. These techniques expected that data are available in form of a proximity matrix. Webb (1999) point out that a hierarchical tree is a nested set of partitions represented by a tree diagram or dendrogram, sectioning into a disjoint groups, if two groups are chosen from different partitions (the result of partitioning at different levels) then either the group are disjoint or one group wholly contains the others.

In general, Jain (1999) observed that, agglomerative hierarchical techniques consist of the following steps.

1. Assign each pattern for a unique cluster
2. Find the smallest entry in the dissimilarity matrix and merge the corresponding two cluster
3. Update the dissimilarity between the new cluster and others clusters
4. Return to steps 2 until all pattern fall into one cluster.

Webb (1999) and Jain (1999) gave about six hierarchical clustering algorithms properties as follows:-

- i. Single – link method
- ii. Complete-link method (furthest neighbors method)
- iii. Centroid clustering method
- iv. Group average clustering
- v. Median clustering method
- vi. Minimum variance clustering or works clustering method

On the other hand, Everitt (1974) observed that, with division method of hierarchical clustering techniques the first task is to split the initial set of individual into two. Now a set of  $n$  individual can be divided into two subset in  $2^{n-1} - 1$  ways and although Everitt (1974) pointed out that, this is obviously only possible for very small set, even with a large computer. Two families of divisive technique are briefly described below. Monothetic, which are based on the possession or otherwise of a single specified and

## Materials and methodology

The data used in this research were obtained from university of Maiduguri teaching hospital on 1500 patient’s from 2009-2013 (5years). Representation of the variables were recorded against six variables these are;

Variables	Code
Sex	1
Age	2
Occupation	3
Educational level	4
Year and date of reporting	5
Marital status	6

The objective of this research work is to separate the set of variables into groups and study the interrelationship among the variables to make a preliminary assessment of their structure. The method of analysis used in this research work is agglomerative hierarchical clustering techniques where emphasis is on single, complete, centroid, and word's linkage methods. The agglomerative hierarchical clustering technique is chosen because they are generally suitable for searching for natural cluster. The distance between variables  $i$  and  $j$  will be calculated using the coefficient distance given as:

$$d(X_i X_j) = \sqrt{1 - r_{ij}} \dots \dots \dots (1)$$

Where  $r_{ij}$  is the correlation between variable  $i$  and  $j$ . The concepts of DISTANCE and SIMILARITY are key ingredients in this statistical procedure. Distance is the measure of how far apart cases are; whereas, similarity measures the closeness of cases. Within a specific group or cluster, distance measures are SMALL while similarity measures are LARGE. Within cluster analysis, cases are grouped together on the basis of their "nearness." Cluster analysis usually employs the DISTANCE measure (how far apart cases are) in defining the clusters or groups.

Single linkage method also known as nearest neighbor defines the distance between two clusters as the minimum distance between them, that is, the distance between the closest point (entries) if  $A$  and  $B$ , is the distance between their closest variable  $d_{AB} = \text{Mindy}$  (minimum distance)

In a matrix, the element in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column gives the distance,  $d_{ij}$  between individual  $i^{\text{th}}$  and  $j^{\text{th}}$  variables.

Complete linkage method or furthest neighbor linkage method proceed in much the same manner with simple linkage method with only one important exception. At each stage, distance between cluster in determined by the distance between the two variables one from each cluster that are further away distance. Thus, complete linkage method ensures that all variables in a cluster are within some maximum distance to each other.

If  $A$  and  $B$  is the distance between their furthest variables, then

$$d_{AB} = \text{max}_d$$

Where  $d_{AB}$  is the maximum distance between the paired variable  $AB$  with another cluster

The centroid linkage method of hierarchical cluster analysis is based on the distance between their centroid clusters that is the distance between groups is defined as the distance between the group centroids.

The procedures is then to fuse group according to distance between their centroids the group with the smallest distance being fused first. The centroid method can be used with both similarity and distance measures.

This method of hierarchical cluster analysis is based on sum of square between cluster summed over all variables that is wards method attempts to minimize the sum of square (SS) of any two (hypothetical) cluster that can be formed at each stage. The distance is the increase in total sum of square which would result if the two clusters were combined and this is the deviation of observation point from their cluster means.

So that a distance matrix is form by  $D = d_{ij}$  where  $X_i$  and  $X_j$  are the observation point of the variables  $X_i$ . The aim of wards method is to minimize the total within group sum of squares.

We wish to test weather the clustering methods and the cluster formed are consistent in relation to the variables clustered using correlation matrices. A quantity closely related to the covariance matrix is the correlation matrix, the matrix of Pearson product-moment correlation coefficients between each of the random variables in the random vector  $\mathbf{X}$ , which can be written

$$\text{corr}(\mathbf{X}) = (\text{diag}(\Sigma))^{-\frac{1}{2}} \Sigma (\text{diag}(\Sigma))^{-\frac{1}{2}} \dots \dots \dots (2)$$

where  $\text{diag}(\Sigma)$  is the matrix of the diagonal elements of  $\Sigma$  (i.e., a diagonal matrix of the variances of  $X_i$  for  $i = 1, \dots, n$ ).

Equivalently, the correlation matrix can be seen as the covariance matrix of the standardized random variables  $X_i / \sigma(X_i)$  for  $i = 1, \dots, n$ . Each element on the principal diagonal of a correlation matrix is the correlation of a random variable with itself, which always equals 1. Each off-diagonal element is between 1 and -

**Results and discussions**

Table 1 Coefficient distance level and cluster joined for single linkage method

**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	6	49.061	0	0	2
2	1	4	55.776	1	0	4
3	2	3	86.971	0	0	5
4	1	5	95.624	2	0	5
5	1	2	95.713	4	3	0

Using the agglomeration schedule to determine the cluster classification from table 1, stage 1 with coefficient 49.061 has the minimum distance, which group or cluster age and marital status followed by next stage 2 with coefficient 55.776 has second distance age and educational level. The stage 2 and 3 has the maximum distance of 31.195

Table 2 Coefficient distance level and cluster joined for complete linkage method

**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	6	49.061	0	0	2
2	1	4	63.388	1	0	4
3	2	3	86.971	0	0	5
4	1	5	100.593	2	0	5
5	1	2	111.104	4	3	0

From the table .2, agglomeration schedule for complete linkage method Coefficient distance level.

Stage 1, age and marital status with coefficient 49.06, and Stage 2, age and educational level with coefficient 63.388 have max distance of 23.883 at stage 2 and 3.

Table 3 Coefficient distance level and cluster joined for centriod linkage method

**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	6	49.061	0	0	2
2	1	4	47.317	1	0	3
3	1	5	79.862	2	0	4
4	1	3	73.684	3	0	5
5	1	2	68.737	4	0	0

From the table 3, centriod linkage method Stage 1, age and marital status with coefficient 49.6, stage 2 age and education level with coefficient 47.317, stage 3 age, year and date of reporting with coefficient 79.867. Max distance of 32.55 at stage 3 and 4.

Table 4 Coefficient distance level and cluster joined for ward linkage method

**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	6	24.531	0	0	2
2	1	4	56.075	1	0	5
3	2	3	99.561	0	0	4
4	2	5	157.603	3	0	5
5	1	2	232.200	2	4	0

From the table 4, agglomeration schedule for ward linkage method Coefficient distance level. Stage 1 age and marital status with coefficient distance of 24.531 stage 2 age and education with coefficient distance of 56.075, stage 3 has 99.561 stage 4 has 157.603, stage 5 has 232.200, it has a max distance 74.597 at stage 4 and 5.

Table 5 Clustering of HIV/AIDS variable by the various methods

Linkage method	Cluster I	Cluster II	Cluster III	Cluster IV
Single method	Sex & marital status	Sex & educational level	Age & occupa	Sex & year
Complete method	Sex & marital status	Sex & educational level	Age & occupa	Sex & year
Centriod method	Sex & marital status	Sex & educational level	Age & year	Sex & occu
Ward's method	Sex & marital status	Sex & educational level	Age & occupa	Sex & year

In table 5, the number of variables clusters by the various methods. These shows the coefficient distance of each variables to another as distance of each variable to another as considered 1 by the method. Comparing the various methods, all the method grouped sex and marital status in the first cluster their educational level, occupation join the cluster at a further distance.

The variables that are common to all the linkage method in cluster I are sex and marital status. This shows the higher ratio of female to male infection with highest number of married people infected. This agrees with Orubuloye et al (1992) Opined that different timing of marriage by sex the practice of polygamy and the age gap between spouses within polygamous unions contributed to the majority of marital known infection. It also agreed with Bakane (2002) in the study of influence of economic and socio-cultural intervention on adolescent sexuality implication on HIV/AIDS that 74.6% of adolescent were sexually active and majority of the sexually active adolescent were not knowledge of sex health education.

Table 6 correlation matrices

**Correlations**

		SEX OF HJV/AIDS	AGE OF PATIENTS	OCCUPATION OF PATIENTS	EDUCATION LEVEL	MARRITAL STATUS	DATE OF REPORTING
SEX OF HJV/AIDS	Pearson Correlation	1	-.251**	.010	-.103**	.141**	-.050
	Sig. (2-tailed)		.000	.707	.000	.000	.052
	N	1500	1500	1500	1500	1500	1490
AGE OF PATIENTS	Pearson Correlation	-.251**	1	-.006	-.045	-.119**	-.021
	Sig. (2-tailed)	.000		.814	.082	.000	.417
	N	1500	1500	1500	1500	1500	1490
OCCUPATION OF PATIENTS	Pearson Correlation	.010	-.006	1	-.080**	.200**	.014
	Sig. (2-tailed)	.707	.814		.002	.000	.595
	N	1500	1500	1500	1500	1500	1490
EDUCATION LEVEL	Pearson Correlation	-.103**	-.045	-.080**	1	.118**	-.012
	Sig. (2-tailed)	.000	.082	.002		.000	.638
	N	1500	1500	1500	1500	1500	1490
MARRITAL STATUS	Pearson Correlation	.141**	-.119**	.200**	.118**	1	.023
	Sig. (2-tailed)	.000	.000	.000	.000		.385
	N	1500	1500	1500	1500	1500	1490
DATE OF REPORTING	Pearson Correlation	-.050	-.021	.014	-.012	.023	1
	Sig. (2-tailed)	.052	.417	.595	.638	.385	
	N	1490	1490	1490	1490	1490	1490

Correlation is significant at the 0.01 level (2-tailed).

The correlation matrices show that the various linkage are significant.

### Summary, Conclusion and Recommendation

A socio-economic and demographic data of HIV/AIDS patients cases were obtained from university of Maiduguri teaching hospital from 2009 to 2013 record on 1500 patients.

The distance which shows similarity between the variables using coefficient distance was calculated, and dendrogram were formed for agglomerative schedule. It was observed that the dendrogram formed by the various methods differs according to their coefficient. Even though the single, complete and ward follow similar pattern of cluster except centriod method.

The variables are grouped according to their similarity, but sex and marital status have dominated all in the first cluster. In the frequency table of sex, higher ratio of female to male infection was observer and the analysis revealed that prevalent rate is higher in married women. That is the disease affects married couples especially female. This could be to the fact that pregnant women attending antenatal clinic when tested and identified with the disease they were recorded.

Considering the variable grouped at second cluster sex and educational level were found to be similar, because must of the people were not inform and lack awareness. In the third cluster age and occupation in all the three linkage, this indicate that the disease effect mostly the working class aged from 20-45. This could be due to the fact that the major mode of transmission of HIV/AIDS is by having sex with infected person, this result agreed with the national HIV survey conducted by the Federal government of Nigeria 2003. A good cluster solution sees a sudden jump (gap) in the distance coefficient. The solution before the gap indicates the good solution of similarity between the variables.

The largest gaps in the coefficients column occur between the stages indicating a cluster solution. The agglomeration schedule is a numerical summary of the cluster solution, and the dendrogram is the graphical summary of the cluster solution.

We have been able to show the natural grouping of HIV/AIDS patients variable using cluster analysis techniques. We have indentified the variables sex and marital status is similar in pattern based on the analysis. The comparison made on the four linkage method showed that the methods are efficient because they are almost the same, even though they gouged variable at different coefficient distance. The study has shown that clustering analysis is another tool for assessing the level of classifies patients into their similarity. it is also shown that in treatment or educating HIV/AIDS patients can be grouped according to sex and marital status sex and educational level age occupation as their appear in similar pattern.

Based on the result, it is recommended that male should follow their spouse to attend antenatal clinics to be tested of the disease. Public enlighten program to educate the public's, so that adolescents are inform about sexual transmitted disease. It was observed that sex and marital status are similar in pattern; they should group and treated together. Also age and occupation are similar; the ages from 20-45 years were being mainly the victims of the disease which is the labour force, which causes less productivity. It should be fought by individual, government and nongovernmental organizations

### References

- Adeyi O, Kanki, P, Odutolu O. A. Idoko. A. (2006) "*AIDS in Nigeria a Nation on the Threshold*" Harman University Press.
- Anderson T.W. (1996) "R.A Fisher and Multivariate Analysis" Statistical Science Journal.
- Anderson, J. (February 2012) "Women and HIV Motherhood and More
- Bakare,O.E.(2002)."Influence of Economic and Socio-cultural Invention in Adolescent Sexuality, an d Implication on HIV/AIDS".Int conf on HIV
- Bennet R. (2004) HIV/AIDS. "The Challenges of this Pandemic Disease on our (Public) Health Care Delivery sy
- Everitt B.S 1974 "*Cluster Analysis*" Heinmann Education Book Ltd U.K.
- Federal Ministry of Health 2010 "A Technical Report on the 2009 National HIV Survey Abuja" (FMOH)
- Harding.R,(2014) Symptom Clusters in People Living with HIV/AIDS attending five Palliative care Facilities in two sub-Saharan African Countries: a Hierarchical Cluster Analysis Pain Symptom.
- Hartigan J.A. (1974) "*Direct Clustering of Data Matrix* of American Statistical Association V. 67, 123-129.
- Jain A.K. (1999) "*Application of Cluster Analysis* Michigan State University Press U.S.A
- Kerrigan, A. Deamna, B. (2012). The Global HIV Epidemics Among Sex Workers.
- Morison, D.F. (1990) Multivariate Statistical Method: A Printer "Chapman and Hall, London
- Opong, C. (1973) *Occupational and Conjugal Inequalities and Insecurity Effect or Family Organization and size* Oxford England Press.
- Unaid 2004 Report on the Global Aids Epidemic Geneva. UNAIDs 2004
- Webb. A. (1999) "Statistical Pattern Recognition" Oxford Uni. Press INC New York.
- Yangi,Z.(2014) The AIDS Epidemic and Economic input Impact Factors in Chongqing, China, From 2006 to 2012: a spatial-Temporal Analysis. <http://creativecommons.org/licenses/by-nc/4.0/>
- World health organization 2007 "WHO Case Definition of HIV for Surveillance and Revised Clinical Staging

---

and Immunological Classification of HIV Related Disease in Adult and Children  
([http://www.who.int/hiv/pub/guideline/HIV\\_staging\\_150\\_307.pdf](http://www.who.int/hiv/pub/guideline/HIV_staging_150_307.pdf)) (PDF) Geneva.  
Yu, M; Vajdy, M. (August 2010) “Mucosal HIV Transmission and Vaccination Strategies Through Oral  
Compared with Vaginal and Rectal Routes.