HIV/AIDS - Research and Palliative Care

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Clustering of HIV Patients in Ethiopia

Wondimu Biressaw[1]
Habtamu Tilaye[2]
Dessie Melese[2]

[1]Benishangul-Gumuz, Wombera Sineor Secondary and Preparatory School, Benishangul, Ethiopia; [2]University of Gondar, College of Natural and Computational Science, Departmentof Statistics, Gondar, Ethiopia

Background: Among the many worldwide health problems, HIV/AIDS has caused severe health problems in several countries. The problem is also widely seen in Ethiopia. The general objective of the study is to cluster HIV patients and to find out the factors that mostly affect the prevalence of HIV within a group (cluster) and between groups (clusters) of HIV patients.

Methods: The study is made based on the 2016 Ethiopian Demographic Health Survey (EDHS) which was collected by the Central Statistical Agency (CSA) of Ethiopia, and the survey collected a total of 26,753 samples, of which 14,785 were women and 11,968 were men and the age group was between 15 and 49 years for both. Binary logistic regression, principal component analysis, cluster analysis, and ANOVA were applied to analyze the data.

Results: The result from binary logistic regression reveals that 15 factors such as ever heard of AIDS, region, water not available for at least a day in the last 2 weeks, has a radio, family members wash their hands, location of the source of water, everything completed to water to make it harmless to drink, food cooked in the house/separate house/outside, has a mobile telephone, has a table, type of place of residence, highest education level attained, current marital status, sex of household members, and age of household members are all significant factors that affect HIV status.

Conclusion: Using these significant variables, 12 principal components are identified which describe 78% of the variation in the data. The result of HIV patients are clustered into 3 clusters and determine the status of HIV levels. Mainly, cluster 2 accounts for 50% of HIV patients whereas cluster 3 and 1 accounts for 40% and 10%, respectively.

Keywords: Ethiopian Demographic Health Survey; EDHS, cluster analysis, principal component analysis, HIV patients

Background

HIV/AIDS is a worldwide public health problem.[1] Globally, around 37.9 million people were living with HIV at the end of 2018 with 2.1 million people newly diagnosed. The sub-Saharan region is the most affected place in the world with 25.6 million people living with HIV.[2]

Ethiopia is one of the majorly affected countries in sub-Saharan Africa, with a huge number of people that are living with HIV/AIDS.[3] HIV is a major public health problem in developing countries.[4]

The spread of HIV shows remarkable differences across the population, sub-groups, regions, and countries at the sub-national level and within sub-districts.[5–9]

HIV/AIDS in Ethiopia is regularly categorized as "generalized" among the adult population with heterogeneity among regions and population groups. The rural spread appears to be comparatively epidemic but heterogeneous, with the majority of rural areas having a comparatively low prevalence of HIV-infected people.[10]

In Ethiopia around 613,000 people are living with HIV. The different prevalence rates are significant when looking at the total number of PLHIV per region as

Correspondence: Dessie Melese
Department of Statistics, College Natural and Computational Science, University of Gondar, P.O. Box 196, Gondar, Ethiopia
Email dessie.melese1@uog.edu.et

Received: 4 February 2021
Accepted: 28 April 2021
Published: 25 May 2021

population size varies from one region to another. Seventy-four percent of PLHIV are from Amhara, Oromia and Addis Ababa.[11]

In Ethiopia, HIV is considered to be concentrated in nine regions and two administrative towns. Eighty-six percent of PLHIV use antiretroviral treatment.[12]

The national prevalence rate of HIV/AIDS in Ethiopia is 0.9%. This research has used clustering of HIV patients in Ethiopia to explore relationships between HIV patients within and between clusters. Therefore, the general objective of the study is to cluster HIV patients and to find out the factors that affect HIV within a group (cluster) and between groups (clusters) of HIV patients.

Method

Data Source and Study Design

The source of the data was obtained from the Ethiopian Demographic Health Survey (EDHS) conducted in 2016. It is a cross-sectional study design conducted from January 18, 2016 to June 27, 2016.

Statistical Analysis

Statistical analysis was performed using the R statistical software.

Variables

Response variable: The response variable for this study is the HIV status of the respondents whose result is positive or negative, which can be recorded as binary (1 = positive, 0 = negative).

Explanatory variables/factors: The explanatory variables or independent variables for this study are the demographic and socioeconomic, cultural, and lifestyle conditions of people that might be vulnerable to HIV infection.

Statistical Analysis

In this study, the authors used a multiple binary logistic regression model to determine significant variables.[13]

Principal Component Analysis

Principal component analysis is a method for dropping the dimensionality of such data sets, increasing interpretability but at the same time reducing information loss.[14] It describes the correlation or variance–covariance structure between the set of variables through a few uncorrelated latent/hidden or new variables, each of which is a linear combination of the original variables which can maximize the variance accounted for.[15]

Cluster analysis is a technique of grouping variables based on similarity or distance by considering the nature of the variable or scale of measurements and the subject matter knowledge. This is in order to make objects in a group similar, and objects in different groups be relatively different.

Results

Table 1 shows that in 414 HIV cases 291 are females and 123 are males. This indicates that the problem is severe for both females than males in Ethiopia. From 414 HIV cases, 373 had enough information about AIDS and 41 of them did not have enough information about AIDS. Most of the HIV patients (350) do not make water safe to drink. And also, of 414 HIV cases, 243 of them had a table and 171 of them did not have a table. Of 414 HIV cases, 225 were married, 84 were divorced, 60 were widowed, and 45 were never married. The highest number of patients was found in cluster two, which was 208 (50%), followed by cluster three which was 165 (40%), and the least number of patients was found in cluster one which was 41 (10%) (Table 1).

Table 2 showed that region, blood test results, cluster numbers, source of drinking water, water not available for at least a day in the previous two weeks, source of water, toilet facilities, had electricity, had a radio, had a television, had a refrigerator, material used on the floor, material used in the walls, material used in the roof, relationship structure, had a telephone (landline). Shared a toilet with other households, type of cooking fuel, a place where household members wash their hands, location of the source for water, person fetching water, anything is done to water to make it safe to drink, food cooked in the house/separate building/outdoors, had a mobile telephone. Owned land usable for agriculture, hectares of agricultural land (1 decimal), owned livestock, herds of farm animals. Wealth index combined, table, chair, bed with cotton/spring mattress, electric mitad, type of residence, highest education level attained. Current marital status, sex of a household member, age of the household member, current, formerly, never married. Eligibility for the female interview, eligibility for the male interview, interviewer that took blood for HIV testing, ever heard of AIDS and number of sexual partners, including a spouse, in the last 12 months, are candidates for multiple binary logistic regression analysis with ($p < 0.1$) (Table 2).

Table 1 Frequency Distribution of HIV Patients in Ethiopian Demographic Health Survey 2016

Variable Name with Category		HIV Test Result			Percent (100%)
		Negative	Positive	Total	
Sex of house hold member	Male	11,845	123	11,968	30
	Female	14,494	291	18,579	70
Region	Tigray	2941	33	2974	7.97
	Afar	1766	23	1789	5.56
	Amhara	3479	43	3522	10.39
	Oromia	3437	23	3460	5.56
	Somali	2172	2	2174	0.48
	Benishangul	1959	16	1975	3.86
	SNNPR	3333	12	3345	2.90
	Gambela	1774	86	1860	20.77
	Harari	1235	35	1270	8.45
	Addis Abeba	2603	96	2699	23.19
	Dire Dawa	1640	45	1685	10.87
Place of residence	Urban	7880	294	8174	71
	Rural	18,459	120	18,579	29
Current marital status	Never married	8477	45	8522	11
	Married	15,947	225	16,172	54
	Widowed	475	60	535	15
	Divorced	1440	84	1524	20
Highest education level	No education/preschool	9687	92	9779	22.22
	Primary	10,180	176	10,356	42.53
	Secondary	4038	103	4141	24.87
	Higher	2434	43	2477	11.14
Anything done to water to make safe to drink	No	23,749	350	24,105	85.54
	Yes	2557	56	2613	13.52
	Do not know	33	2	35	0.48
Household members washed their hands	Observed fixed place	1582	25	1607	6.04
	Observed mobile place	13,495	275	13,770	66.42
	Not observed: not in dwelling	10,667	102	10,769	24.64
	Not observed not permeation to see	103	0	103	0.00
	Not observed other reason	492	12	504	3.00
Had mobile telephone	No	8492	60	8552	14.49
	Yes	17,847	354	18,201	85.50
Ever heard of AIDS	No	1595	41	1636	9.90
	Yes	27,744	373	25,117	90.10
Cluster	1	10,108	41	10,149	10
	2	10,190	208	10,398	50
	3	6041	165	6206	40
Total		26,339	414	26,753	100

Table 2 Chi-Square Test Results of HIV Patients in Ethiopian Demographic Health Survey 2016

Variables (X's)	Chi-Square	p-value
Region	311.02	0.000
Blood test result	53,506	0.000
Cluster number	2366.1	0.000
Source of drinking water	274.61	0.000
Water not available for at least a day two weeks	9.07	0.060
Time to get to water source (minutes)	263.05	0.000
Type of toilet facilities	146.62	0.000
Had electricity	237.21	0.000
Had radio	51.81	0.000
Had television	171	0.000
Had refrigerator	55.28	0.000
Had bicycle	0.31	0.860
Had motorcycle/scooter	0.41	0.810
Had car/truck	3.49	0.170
Material used on floor	189.41	0.000
Material used in wall	60.86	0.000
Material used on roof	56.17	0.000
Number of rooms used for sleeping	23.81	0.360
Relationship structure	176.83	0.000
Has telephone (land-line	28.91	0.000
Share toilet with other households	135.3	0.000
Type of cooking fuel	202.92	0.000
Place where household members washed their hands	48.74	0.000
Location of source for water	8.28	0.080
Person fetching water	29.73	0.000
Anything done to water to make it safe to drink	33.79	0.000
Food cooked in the house/separate building/ outdoors	51.87	0.000
Had mobile telephone	60.03	0.000
Had watch	2.14	0.340
Had animal-drawn cart	0.88	0.640
Had boat with a motor	1.07	0.590
Had a computer	2.84	0.240
Owned land usable for agriculture	159.49	0.000
Hectares of agricultural land (1 decimal)	196.74	0.000
Owned livestock, herds, or farm animals	221.88	0.000
Frequency of household members smoking inside the house	3.8	0.870
Wealth index combined	242.79	0.000
Table	65.62	0.000
Chair	51.99	0.000
Bed with cotton/spring mattress	93.49	0.000
Electric mitad	74.52	0.000
Kerosene lamp/pleasure lamp	0.97	0.610
Bagag	0.32	0.850
Type of residence	324.87	0.000
Highest education level attained	52.86	0.000
Current marital status	553.63	0.000
Sex of household member	40.57	0.000
Age of household member	275.96	0.000
Current, formerly, never married	465.16	0.000
Eligibility for female interview	37.95	0.000
Eligibility for male interview	39.56	0.000
Consent for additional test	0.6	0.960
Interviewer that took blood for HIV testing	373.32	0.000
Religion	5.95	0.920
Age of first sex	86.19	0.200
Ever heard of AIDS	11.59	0.020
Number of sexual partners, including spouse, in the last 12 months	70.07	0.000

Table 3 Result of the Multiple Binary Logistic Regressions of HIV Patients in Ethiopian Demographic Health Survey 2016

Coefficient	Estimate	Std.Error	Z.Value	P-value
Intercept	−18.64	1123.79	−0.02	0.99
Ever heard of AIDS				
Ever heard of AIDS (yes)	−0.32	0.18	−1.81	0.07
No (ref)				
Region				
Afar	0.20	0.29	0.68	0.50
Amhara	0.61	0.25	2.46	0.01
Oromia	−0.07	0.29	−0.25	0.80
Somali	−2.04	0.74	−2.77	0.01
BenishangulGumuz	0.30	0.32	0.94	0.35
SNNPR	−0.74	0.35	−2.09	0.04
Gambela	1.43	0.22	6.36	0.00
Harari	0.43	0.27	1.60	0.11
Addis Ababa	0.17	0.23	0.74	0.46
Dire Dawa	0.31	0.25	1.25	0.21
Tigray (ref)				
Water not available for at least a day in the previous two weeks				
Yes, interrupted for a full day or more	0.17	0.12	1.42	0.16
Do not know	1.08	0.45	2.38	0.02
No, not interrupted for a full day(ref)				
Had electricity				
Yes	0.36	0.22	1.59	0.11
No (ref)				
HHad radio				
Yes	0.39	0.12	3.35	0.00
No(ref)				
Number of rooms used for sleeping	0.00	0.00	0.40	0.69
One adult	14.08	1123.79	0.01	0.99
Two adults, opposite sex	13.61	1123.79	0.01	0.99
Two adults, same sex	13.85	1123.79	0.01	0.99
Three+ related adults	13.23	1123.79	0.01	0.99
Unrelated adults	13.37	1123.79	0.01	0.99
No adults (ref)				

(*Continued*)

Table 3 (Continued).

Coefficient	Estimate	Std.Error	Z.Value	P-value
Place where household members washed their hands				
Observed, mobile place	0.64	0.23	2.81	0.01
Not observed; not in dwelling	0.39	0.25	1.55	0.12
Not observed, no permission to see	−12.74	340.29	−0.04	0.97
Not observed other reason	1.02	0.39	2.59	0.01
Observed fixed place(ref)				
Location of source of water				
In one yard/plot	−3.37	1.06	−3.19	0.00
Elswhere	−2.54	0.77	−3.32	0.00
In one dwelling(ref)				0
Anything done to water to make it safe to drink				
Yes	0.39	0.14	2.83	0.00
Do not know	−12.30	742.42	−0.02	0.99
No(ref)				
Food cooked in the separate house/outside				
In separate house	−0.40	0.13	−3.07	0.00
Outside	0.07	0.15	0.47	0.64
Others	−12.96	732.42	−0.02	0.99
In the house(ref)				
Had mobile telephone				
Yes	0.54	0.18	3.06	0.00
No(ref)				
Owned livestock, herds, or farm animals				
Yes	−0.23	0.15	−1.49	0.14
No (ref)				
Wealth index combined	−0.11	0.08	−1.43	0.15
Table				
Yes	0.27	0.13	2.06	0.04
No (ref)				
Place of residence				
Rural	−1.02	0.22	−4.61	0.00
Urban (ref)				
Highest education level attained				

(*Continued*)

Table 3 (Continued).

Coefficient	Estimate	Std.Error	Z.Value	P-value
Primary	0.60	0.15	3.99	0.00
Secondary	0.60	0.18	3.32	0.00
Higher	−0.24	0.22	−1.09	0.28
Do not know	0.65	1.05	0.62	0.54
No education, preschool				
Current marital status				
Married	0.76	0.20	3.85	0.00
Widowed	2.33	0.25	9.19	0.00
Divorced	1.60	0.22	7.36	0.00
Never married (ref)				
Living together (ref)				
Not living together (ref)				
Sex of household members				
Female	0.55	0.13	4.27	0.00
Male (ref)				
Age of household members	0.05	0.01	7.64	0.00

Multiple Binary Logistic Regression Results

Table 3 showed that the odds of an individuals who had heard about AIDS are 0.73 times than those individuals who had not heard about AIDS.

The number of HIV patients in the Gambella region is 4.17 times the number of HIV positive cases in the Tigray region and Amhara which is 1.84 times the reference region. The problem is less in SNNPR than the reference region by 48% and less in Somali which is 13% than the Tigray reference region.

Binary logistic regression analysis also shows that the odds of individuals who had a radio is 1.5 times individuals who had no radio. The source of drinking water is in one yard/plot area is 3.4% less likely to be infected with HIV and for those whose source of drinking water is elsewhere it is 7.9% less likely to be infected with HIV than those whose source of drinking water is in one dwelling. The problem for those who made water safe to drink were 0.39 units lower than those who did not make water safe to drink.

The result for food cooked in separate house/outside indicates the chance of those who cooked their food in a separate building is 0.40 units lower than those who cooked their food in the house (ref). The odds of individuals who had mobile telephone are 1.7 times those who had no mobile telephone. The result for the place of residence indicated individuals wh lived in rural areas had 36% fewer HIV infections than those whose place of residence was in an urban area.

The highest education level indicates that the odds of being an HIV positive individual with both primary and secondary education level was 1.8 times that of those who had no education. Current marital status indicated that the category married, widowed, and divorced was found to be 2.1, 10.27, and 5 times that of those who never married (ref), respectively.

The prevalence of HIV infection varies by sex. The results indicate that the chance of females being HIV positive was 1.7 times higher than males (Table 3).

Amount of explained variance: The first seven components are taken (57% of the variation would be explained), the first nine components are taken (66% of the variation

would be explained), and the first twelve components would be taken (78% of the variation would be explained).

Subject matter consideration: From this aspect, we observed that the results from the analysis of six principals, the proportion of variation to explain the variable is 52%. This shows that there is around 48% of information loss to explain the variables and 12 principal components are more direct to interpret and easy to relate variables (Table 4).

Principal factor one is related to place of residence. The correlation between key variables in principal factor one showed that there was a strong positive association (0.81) with individuals who had electricity. There was a good indirect correlation (−0.70) with individuals who had their own livestock, herds, or farm animals. There was a good positive correlation (0.58) with the wealth index combined, and a positive correlation (0.33) with the region. There was a positive correlation (0.40) with the highest education level attained.

Principal factor two was related to the age of the household. The correlations between key variables of principal factor two suggested that there was a direct correlation (0.25) with current marital status.

Principal factor three was related to the region. This component primarily measured the regional state of HIV patients. Principal factor four was related to the sex of household members. There was a negative correlation (−0.22) with the age of first sex. Principal factor five was associated with the place where food was cooked. Principal factor six was associated with the relationship structure. The correlation between the key variables of this principal factor was a good positive correlation with the number of rooms used for sleeping. Principal factor seven was related to individuals who had a mobile telephone. Principal factor eight was associated to everything done to water to make it safeto drink.

Principal factor nine was associated to the wealth index combined. The correlation between the key variables of principal factor nine suggests the following:

There is a good positive correlation (0.61) with individuals who had a table and also there was a positive correlation (0.34) with highest education level attained and individuals who had electricity. There was a positive correlation with individuals who had a radio and individuals who had a mobile telephone, 0.37 and 0.31 respectively. There was a negative correlation (−0.36) with the place where household members washed their hands.

Table 4 The Standard Deviation of Principal Components

Principal components	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21
Standard deviation	2.14	1.28	1.15	1.13	1.04	1.03	0.99	0.98	0.96	0.95	0.93	0.92	0.87	0.84	0.80	0.78	0.76	0.68	0.64	0.49	0.44
Proportion of variance	0.22	0.08	0.06	0.06	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.01	0.01
Cumulative proportion	0.22	0.30	0.36	0.42	0.47	0.52	0.57	0.61	0.66	0.70	0.74	0.78	0.82	0.85	0.88	0.91	0.94	0.96	0.98	0.99	1.00

Table 5 Principal Value and Significant Variables from Binary Logistic Regression

Variables (X's)	MR1	MR9	MR2	MR7	MR3	MR5	MR6	MR4	MR11	MR8	MR10	MR12	Fc	Fu
Region	0.33	0.17	0.00	0.02	0.89	0.00	0.03	0.02	−0.03	0.01	0.05	−0.11	0.94	0.06
Water not available for at least a day in the previous two weeks	0.02	−0.02	0.00	0.00	0.02	−0.02	−0.01	0.01	−0.01	0.05	0.55	−.09	0.31	0.69
Had electricity	0.81	0.34	0.00	0.07	0.11	0.02	0.03	0.02	−0.03	0.01	−0.02	0.11	0.79	0.21
Had radio	0.15	0.37	0.02	0.11	0.08	0.00	0.11	0.01	−0.02	0.04	−0.06	0.05	0.20	0.80
Number of rooms used for sleeping	0.02	0.28	0.00	0.02	0.01	0.05	0.54	−0.01	0.03	0.06	0.01	−0.01	0.38	0.62
Relationship structure	0.03	0.09	−0.05	0.06	0.02	0.00	0.57	−0.02	−0.19	0.00	−0.03	0.08	0.39	0.61
Place where household members washed their hands	−0.19	−0.36	0.00	−0.03	0.00	0.00	−0.12	−0.02	0.01	−0.06	0.03	0.13	0.20	0.80
Location of source for water	0.03	0.01	0.00	−0.02	0.02	−0.02	−0.02	0.02	0.00	−0.02	−0.03	−0.12	0.02	0.98
Anything done to water to make safe to drink	0.05	0.07	0.00	0.02	0.01	0.01	0.05	0.01	0.00	0.63	0.05	0.06	0.41	0.59
Food cooked in the house/separate building/outdoors	0.05	0.02	−0.01	0.03	0.00	0.87	0.04	−0.03	0.01	0.02	−0.04	0.10	0.77	0.23
Had mobile telephone	0.25	0.31	−0.03	0.90	0.03	0.04	0.12	−0.02	−0.04	0.03	0.00	0.11	1.00	0.00
Owned livestock, herds, or farm animals	−0.70	−0.16	0.04	−0.06	−0.13	−0.02	0.03	−0.04	−0.04	−0.06	−0.04	0.22	0.59	0.41
Wealth index combined	0.58	0.69	−0.01	0.11	0.06	0.07	0.07	0.00	−0.03	−0.02	−0.04	0.14	0.86	0.14
Had table	0.18	0.61	0.00	0.06	0.05	0.00	0.14	0.03	−0.02	0.01	0.06	−0.08	0.45	0.55
Type of place of residence	−0.87	−0.24	0.00	−0.09	−0.11	−0.02	−0.08	−0.04	0.01	−0.03	−0.04	0.08	0.84	0.16
Highest education level attained	0.40	0.34	−0.18	0.13	0.06	0.02	0.05	−0.19	−0.17	0.07	0.00	−0.19	0.44	0.55
Current marital status	−0.02	−0.05	0.25	−0.03	−0.02	0.01	−0.20	0.13	0.57	0.00	−0.02	0.02	0.45	0.37
Sex of household member	0.03	−0.03	−0.11	−0.03	0.02	0.03	0.02	0.76	0.18	0.00	−0.03	0.03	0.63	0.01
Age of household member	−0.06	0.02	0.96	−0.02	0.00	−0.01	−0.05	−0.09	0.22	0.01	−0.01	−0.02	0.99	0.98
Ever heard of AIDS	0.00	0.00	0.00	0.00	−0.01	0.01	0.01	−0.13	0.00	0.00	0.00	0.01	0.02	0.95
Age of first sex	0.01	−0.02	0.00	0.00	0.01	0.00	0.00	−0.22	0.01	0.00	−0.02	0.02	0.05	0.95

Principal factor 10 primarily measured water being unavailable for at least a day in the previous two weeks. Principal factor 11 primarily measured current marital status. The correlation between this key variable suggests there was a positive correlation (0.22) with the age of the household member. Principal factor 12 primarily measured owns livestock, herds, or farm animals (Table 5).

Cluster Analysis

Agglomerative Clustering of Variables

Start with the individual variables. Thus, there are initially as many clusters as objects. The most similar variables are first grouped, and these initial groups are merged according to their similarities. Then those groups with low similarity are taken as clusters. Eventually, as the similarity decreases, all sub-groups are fused in to a single cluster. From the above result the suggestion would be six clusters, where two variables (had radio and age of household member) are each forming an individual cluster. Where the more the shorter distance of joining implies the more clusters is similar. Most of the variables are grouped in clusters 1 and 2, whereas variable sex of household is removed from cluster 1 and added to cluster 5 and the variable 'had radio' is removed from cluster 5 and added to cluster 2 by k mean clustering.

K-Mean Clustering of Variables

It is one of the non-hierarchical cluster analyses with a purpose of assigning elements to pre-determined clusters, in a way that each item is assigned to a cluster with the nearest mean. Based on the results, clustering by this method almost agrees with agglomerative method, with some exceptions, such as this method merges variable 'had radio' and splits 'sex of household' as one cluster, but the agglomerative method merges the variable 'sex of household' and splits the variable 'had radio' as one cluster.

Bootstrap Clustering of Variables

Bootstrap clustering suggests that the cluster with a large p-value is highly supported by the data. Hence, the number of cluster and element selection had to be done based on the desired p-value. It gives a statistically significant number of clusters for the desired level of confidence. If the number of times items are assigned together is at least at a desired level of confidence, then this group is considered as one cluster with the desired level of confidence, e.g., If some groups of items are assigned together, with the number of times being greater than or equal to 0.95 then these groups of items are considered to be one cluster with a 95% confidence level. The result assures the existence of the first three clusters and the remaining three clusters (4, 5, and 6) are rejected because their confidence levels are 93%, 57%, and 85%, respectively and are less than the 95% confidence level (Table 6). To understand Table 6 see supplementary material (Table S1).

Discussion

The analysis results showed that the odds of individuals who had heard about AIDS are 0.73 times likely than individuals who had not heard about AIDS. This does not coincides with the finding of.[16]

In the Amhara and Gambella regions, there are 1.84 and 4.17 times the HIV cases as compared to the Tigray region, respectively. In Somali and SNNPR there are 0.13 and 0.48 times the HIV cases thatn in the Tigray region (ref), respectively. The prevalence of HIV in Ethiopia is estimated at 1.55%. This finding shows the prevalence is reduced from the finding of.[17] This indicates that the prevalence of HIV infection varies from region to region.

The people who had a radio were 1.5 times more than those who had no radio.[18] This was the main tool used to address people, creating awareness programs and ensuring the people had enough comprehensive knowledge about HIV.

Where sources of drinking water is in one yard/plot area, 3.4% of people areless likely to be infected with HIV. A source of drinking water found elsewhere means 7.9% of people are less likely to be infected with HIV, than

Table 6 List of Variables in Each Cluster

Clusters	Agglomerative Clustering Method	K-Mean Clustering Method	Bootstrap Clustering Method
Cluster 1	HV104,HV201A, HV241,HV235, HV237,HV230A, HV115	HV115, HV230A, HV241,HV237, HV201A,HV235	HV104, HV201A, HV241,HV235, HV237, HV230A, HV115
Cluster 2	Hv206,HV270, HV024,HV106, HV243A, SH121G;	HV206,HV270, SH121G, HV207, HV243A, HV024,HV106;	Hv206,HV270, HV024,HV106, HV243A, SH121G;
Cluster 3	HV246,HV025;	HV246,HV025;	HV246,HV025;

e source of drinking water is in one dwelling. Those who make water safe to drink is 0.39 units lower than those who do not make water safe to drink.

The odds of those who cooked their food in a separate house are 67% less likely than those who cooked their food in their house. The odds of individuals who had a mobile telephone are 1.7 times more than those who had no mobile telephone.[19] This shows that the problem is higher for those who had a mobile telephone than those who did not.

Adults who live in rural areas were 36% less likely to be HIV positive than adults who lived in urban areas. This indicates the problem is more severe in urban areas.[17]

Regarding the educational levels of individuals who had primary, secondary, and higher education levels were 1.82, 1.82, and 0.78 times than no education or only pre-school level, respectively.[20,21] This indicates individuals whose education level is primary, secondary, and higher education were most likely to be infected with HIV than those who have no education or preschool level only.

Also the result of current marital status indicates that, the categories married, widowed, and divorced was found to be 2.1, 10.27, and 5 times that of those who never married, respectively.[22] This result indicates that individuals who were married, widowed, or divorced are most likely to be infected with HIV than those individuals who never married.

Conclusions

A binary logistic regression reveals that 15 factors, such as: ever heard of AIDS, region, water not available for at least a day in the previous two weeks, had a radio, place where household members washed their hands, location of source of water, anything done to water to make it safe to drink, food cooked in separate house/outside, had a mobile telephone, had a table, type of residence, highest education level attained, current marital status, sex of household members, and age of household members are significant factors which affect HIV status. Using these significant variables, 12 principal components are identified which describe 78% of the variation in the data. As a result HIV patents are clustered into three clusters to determine HIV status. Mainly cluster two accounts for 50% of HIV patients, whereas clusters one and three account for 10% and 40%, respectively.

Abbreviations

AIDS, Acquired Immunodeficiency Syndrome; HIV, Human Immunodeficiency Virus; WHO, World Health Organization; PLWH, people living with HIV; HAPCO, HIV/AIDS Prevention and Control Office; CSA, Central Statistical Agency; EDHS, Ethiopian Demographic Health Survey; UNAIDS, United Nations Program on HIV/AIDS; USAID, United States Agency for International Development; FMOH, Federal Ministry of Health.

Data Sharing Statement

The data sets used and analyzed during the current study are available from the Ethiopian Demographic and Health Survey 2016.

Ethical Consideration

Ethical clearance was obtained from the college review board of the University of Gondar, College of Natural, and Computational Science. A formal letter of cooperation was written for Central Statistical Agency.

Acknowledgments

The authors would like to acknowledge that the Ethiopian Central Statistical Agency and the data were obtained from the EDHS 2016; they have given permission to access the data after we have prepared the proposal on the title.

Author Contributions

All authors made substantial contribution to conception and study design, acquisition of data, analysis and interpretation, took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interests in this research article.

References

1. FHAPCO, Report on progress towards implementation of the UN Declaration of Commitment on HIV/AIDS. 2010.
2. WHO, HIV/AIDS. Fact Sheet, World Health Organization. 2018 http://Www.Who.Int/Mediacentre/Factsheets/Fs360/En/2016..
3. FMOH, Report on progress towards implementation of the UN Declaration of Commitment on HIV/AIDS. 2010.
4. Agegnehu CD, et al. Determinants of comprehensive knowledge of HIV/AIDS among reproductive age (15–49 years) women in Ethiopia: further analysis of 2016 Ethiopian demographic and health survey. 2016.
5. UNAIDS. *UNAIDS Report on the Global AIDS Epidemic*. UNAIDS: Geneva; 2012.
6. Km D, Hw J, Jw C. The evolving epidemiology of HIV/AIDS. 2012.

7. Asamoah Odei E, Calleja JG, and Boerma J. HIV prevalence and trends in sub-Saharan Africa: no decline and large sub-regional differences. *The Lancet*. 2004.
8. Shalik N, Adbullah F, Lombard CJ. Masking through averages: intra-provincial heterogeneity in HIV prevalence within the Western. *South Afr Med J*. 2006. 98.
9. Wand H, Whitaker C, Ramjee G. Geoadditive models to assess spatial variation of HIV infections among women in local communities of Durban, South Africa. 2011;10(1):28. doi:10.1186/1476-072X-10-28
10. Sam D, Challenges of Containing New HIV Infections in Ethiopia (view at Google). 2013.
11. FHAPCO. HIV Prevention in EthiopiaNational Road Map 2018-2020. 2018.
12. UNAIDS. Country progress report on HIV/AIDS: (view at Google). 2011.
13. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons; 2000.
14. Joffifle JT, Cadima, I. Principal component analysis: a review and recent developments. *Philosophical Trans R Soc A*. 2016.
15. Richard AJ. Applied Multivariate Statistical Analysis. 2007.
16. WHO. Major UN study finds alarming lack of knowledge about HIV/AIDS among young people. 2002.
17. Lakew Y, Benedict S, Haile D. Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. 2015;5(11): e008669. doi:10.1136/bmjopen-2015-008669
18. Bogale S, Analysis: HIV/Aids is surging in Ethiopia,Again. 2017.
19. Tesfaw A, Jara D, Temesgen H. Dietary diversity and associated factors among HIV positive adult patients attending public health facilities in Motta Town, East Gojjam Zone, Northwest Ethiopia. 2017.
20. Serra MAAO, et al. Socio demographic and Behavioral Factors Associated with HIV vulnerability according to sexual orientation. 2016.
21. Amo D, Julia. Inequalities by educational level in response to combination antiretroviral treatment and survival in HIV-positive men and women in Europe (1996-2013): a collaborative cohort study). 2017.
22. Shisana O, Risher K, Celentano D, et al. Does marital status matter in an HIV hyper endemic country? Findings from the2012 South African National HIV Prevalence. *Incidence Behav Survey*. 2016.