What is the Impact of Socio-Economic Environment on HIV Prevalence Rate in Africa?

Jingwei (Chrisbell) Ni, Kebing Li, Linh Nguyen, Rayne Wang

Final Draft

Introduction

HIV is a virus that attacks the human immune system and a person's immune system "will eventually be completely destroyed" without treatment (Avert, 2018). HIV/AIDS is a major public health problem and it has killed more than 35 million people so far. According to WHO's data, one million people died from HIV-related diseases worldwide in 2016.

Although the HIV prevalence among adults (aged 15-49) in Africa has steadily decreased since its peak in 2000 of 5.876 % to 4.264% in 2016, Africa is still battling with this disease because its rate is much higher than the global rate, including Africa, of 0.8% (WHO, 2017). We choose to study HIV prevalence in Africa since it is the most affected region, "with 25.6 million people living with HIV in 2016" (WHO, 2017). Behind respiratory tract infections, HIV/AIDS ranked the second leading cause of death for adults in Africa, according to WHO's African regional health report from 2014. In addition, as of 2016, the HIV prevalence rate varied within different regions of Africa. For example, Comoros' HIV prevalence rate is around 0.1% while Botswana's HIV prevalence rate is around 23%. It is important to understand factors that correlate with these phenomenon.

In this study, we will investigate whether socio-economic factors, including level of education, economic growth, inward foreign direct investment (FDI), unemployment rate, inflation rate, and gender ratio are associated with the HIV prevalence rate across African countries. If we could show the association between poor socio-economic environment and high

HIV prevalence rate, then improving the socio-economic situation by expanding education programs and health benefits may help control HIV in Africa.

Literature Review

Topics relating to HIV prevalence in African countries have been broadly discussed in recent years. In one study, Luboobi and Mugisha (2005) discussed the trends and challenges of HIV/AIDS pandemic in Africa. The authors used data from UNAIDS and employed the mathematical models developed by Mugisha (1999) to model the spread of HIV/AIDS pandemic in age-structured populations. Luboobi and Mugisha (2005) point out that HIV/AIDS pandemic in Africa is unique in comparison to all other low-income countries around the world. The prevalence of HIV/AIDS relates to the unique traditions and culture among African countries, such as sharing of wives, widow inheritance, sexual cleansing, and early marriages. Hence, the government should take more responsibility and may require help from international partnership to fight against HIV/AIDS in the future.

In another study, Steenkamp et al. (2013) applied a cross-sectional study to determine HIV status and demographic and socio-economics information in two Eastern Cape districts. The authors obtained their data by conducting interviews using a questionnaire that captured demographic and socio-economics variables. From this experiment, no significant difference can be shown for HIV prevalence in the two districts. However, there is a statistical significance between gender and HIV-status with 41% of females being HIV-positive and 21% of males being HIV-positive. There is also a strong correlation between education and HIV prevalence and unemployment and HIV prevalence.

Different from previous researches, we are focusing on Africa as a whole by looking at each African country and using more recent data. We also have different variables such as GDP, health expenditure, gender ratio, and inflation. By conducting this study, we hope to build on previous researches and expand the findings to a larger population under a more recent setting.

Data Summary

The ideal dataset for us should include socio-economic indicators from different angles across all African countries in recent years. Therefore, we can analyze the similarities and differences between these countries.

Our data is gathered from the World Bank's World Development Indicators. We have country-year level of statistics on 10 selected socio-economic indicators, covering 46 countries in Africa. We also chose 2005-2014 as our focusing years to do panel analysis.

We do not include Libya, Mauritius, Seychelles or Sao Tome and Principe because their HIV prevalence rates are not available. We also exclude South Sudan, Eritrea, Somalia, and Niger due to their missing inward foreign direct investment (FDI) and GDP per capita.

Summary Statistics for All Variables:

Variable	Observations	Mean	Standard Deviation	Min	Max
HIV prevalence rate (adult 15-49 years old), %	460	5.37	6.83	0.10	28.30
Net FDI inflows, % of GDP	459	5.02	7.44	-5.98	84.95
GDP growth, annual %	460	5.02	4.38	-36.70	22.59
GDP per capita, US$	459	2092.55	3089.46	150.51	22742.38
Unemployment rate, %	460	10.10	7.59	0.70	36.00
Health expenditure, % of GDP	460	5.93	2.25	1.63	14.39
Inflation (consumer prices), annual %/100	457	0.64	11.39	-0.36	244.11
Duration of secondary education, years	460	6.23	0.74	5.00	7.00
GDP per capita (lag 1 year), US$	459	1968.95	2959.60	127.43	22742.38
Gender ratio (male/female)	460	0.99	0.04	0.93	1.22

The average HIV prevalence rate from 2005 to 2014 in 46 countries under study is 5.37 percent. The high standard deviation of 6.83 percentage points implies a highly deviated rate across countries. The average HIV prevalence rate among young population from 15 to 24 years old differs between gender. Particularly, the average rate and standard deviation for young women population, 3.02 and 4.25 percent respectively, are much higher relative to those for young men population. We use 10 indicators to examine the socio-economic impact on the HIV prevalence rate in Africa. All the data in this study are extracted from the World Bank Data so that we have little concerns about attenuation bias on economic indicators. As for the HIV prevalence rate, the attenuation bias is more concerned because of the reliability of the government source on claiming this rate. Governments are likely underreport the HIV prevalence rate to mask the problem. In addition, we do have concerns about omitted variable bias. While some indicators like GDP per capita, unemployment rate, or health expenditure might have a clearer correlation with the HIV prevalence rate, FDI is not. We predict the correlation between FDI and the HIV prevalence rate to be negative. We add FDI to the regression model because omitting FDI can create a downward bias on GDP per capita (corr(FDI, HIV rate)*corr(FDI, GDP per capita) < 0) or an upward bias on employment rate (corr(FDI, HIV rate)*corr(FDI, unemployment rate >0).

Empirical Work

Our research studies both time fixed effect and country fixed effect. Adding time fixed effect leaves the estimated coefficients nearly the same with the baseline regression in Table 1, Column 4. However, adding country fixed effect leaves the estimated coefficients dramatically changed, shifting the signs of some variables. This is probably due to the lack of variation within each country over years. Thus, our conclusion is based on the time fixed effect regression (Column 7).

Final Model with Time Fixed Effect:

$$HIVrate_{ct} = \beta_0 + \beta_1 \text{ logGDPPC}_{ct} + \beta_2 \text{ Health\_Exp}_{ct} + \beta_3 \text{ Sec\_Education}_{ct} + \beta_4 \text{ Gender\_ratio}_{ct} + \beta_5 \text{ Unemployment}_{ct} + \beta_6 \text{ Inflation\_CPI}_{ct} + \beta_7 \text{ FDI}_{ct} + \beta_8 \text{ Lagged\_GDPPC}_{ct} + \beta_9 \text{ Unemployment*Sec\_Education}_{ct} + \beta_{10} \text{ Gender\_ratio*Sec\_Education}_{ct} + \alpha_t + \varepsilon_{ct}$$

Table 1:

	Dependent variable: HIV Prevelance Rate							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	OLS	OLS	OLS	OLS	LAD	OLS	OLS	OLS
logGDPPC	1.186***	-0.850	1.204***	0.938***	0.643	-0.874***	1.029***	-0.085
	(0.224)	(1.331)	(0.288)	(0.224)	(0.448)	(0.183)	(0.080)	(0.290)
Health Expenditure	0.459***	0.347***	0.436***	0.259***	0.233**	0.009	0.284***	0.048
	(0.115)	(0.094)	(0.094)	(0.083)	(0.107)	(0.040)	(0.037)	(0.035)
Secondary Education	-5.354***	-3.922***	-3.820***	-63.790***	-32.870*	3.456	-62.836***	2.704
	(0.369)	(0.362)	(0.365)	(10.480)	(18.529)	(4.184)	(5.259)	(3.915)
Gender Ratio		-36.784***	-32.493***	-411.636***	-198.886	27.547	-406.423***	28.676
		(8.135)	(7.993)	(67.225)	(124.535)	(19.912)	(28.297)	(17.919)
Unemployment		0.195***	0.214***	2.362***	2.456***	-0.196	2.355***	-0.241
		(0.043)	(0.043)	(0.197)	(0.393)	(0.223)	(0.089)	(0.234)
Inflation Consumer Price			0.053***	0.057***	0.060	0.003***	0.057***	0.004***
			(0.004)	(0.004)	(7.096)	(0.001)	(0.004)	(0.000)
UnemployementXEducation				-0.380***	-0.379***	0.023	-0.380***	0.030
				(0.031)	(0.056)	(0.034)	(0.014)	(0.036)
GenderRatioXsecondary				64.966***	33.896*	-3.915	64.032***	-3.082
				(10.613)	(18.035)	(4.063)	(5.215)	(3.836)
FDI			-0.063***	-0.061***	-0.060***	-0.007***	-0.059***	-0.007***
			(0.018)	(0.017)	(0.012)	(0.003)	(0.015)	(0.002)
logGDPPC(t-1)		2.069						
		(1.312)						
Constant	27.665***	53.757***	48.351***	400.171***	189.554	-12.548	394.064***	-19.394
	(3.088)	(7.519)	(7.153)	(66.537)	(131.129)	(20.711)	(29.215)	(18.502)
Observations	460	460	460	460	460	460	460	460
R-squared	0.473	0.556	0.573	0.699		0.304	0.700	0.396
Country Fixed Effect						Yes	No	Yes
Time Fixed Effect						No	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1 shows the result of all of our models. Overall, we make a few adjustments to our data. In order to interpret GDP using percent change, we transform the variables into log(GDP per capita) and log(GDP per capita last year). We also fill in the 6 missing values in our data set by using the average of the other values for the variable. However, we still tag the observation with a variable to account for any possible differences it has with the rest of the data.

Our baseline regression model is shown in Table 1 Column 4. To come up with this model, we first start with a simple model and add more variables to it. The first model we have is shown in Table 1 Column 1. In this model, we choose three variables (GDP per capita, health expenditure, and secondary education duration) that we think would be correlated with total HIV prevalence and it turns out that all three variables are significant. In the second model (Table 1 Column 2), we decide to add gender ratio, unemployment, and GDP per capita in the previous year. It turns out that GDP per capita in the previous year is not important and is taken out from future models. Model 3 adds inflation and foreign direct investment (FDI) but only inflation is statistically significant. For our final model, we want to add interaction terms. We decide to add unemployment * secondary education duration because we believe that secondary education has a stronger relationship with HIV prevalence rate when the unemployment rate is lower than a certain threshold. We also add gender ratio * secondary education duration because we believe that secondary education is only significant when the majority of population is male (high gender ratio) since less females go to school. Also, to prevent having issues due to outliers, we use both OLS and LAD to run our final regressions, where the OLS regression result is on column 4 and the LAD regression result is on column 5.

By adding more variables, R-squared from 4 regressions steadily increases. By using 8 variables in regression (4), we can conclude that 69.88% of the variation in the HIV prevalence rate is explained by the model.

From regression (4), we notice the following: 1) a 1 % increase in GDP per capita relates to a 0.0093783 percentage points increase in the predicted HIV prevalence rate, holding all other variables in the model fixed, 2) a one percentage point increase in health expenditure relates to a 0.25882 percentage points increase in the predicted HIV prevalence rate, holding all other variables in the model fixed, 3) a one percentage point increase in inflation relates to a 0.00057 percentage points increase in the predicted HIV prevalence rate, holding all other variables in the model fixed, 4) a one percentage point increase in the foreign direct investment relates to a 0.06122 percentage points decrease in the predicted HIV prevalence rate, holding all other variables in the model fixed.

By adding the interaction variable "unemploymentXeducation" into our model, we want to see if there's difference between the relationship of HIV rate and unemployment rate when secondary education is at different level. By taking partial derivative with respect to unemployment rate, we have $\partial HIV\, rate/\partial Unemployment \ = \ 2.36185 - 0.38011 * secondary\ education.$ From our data for secondary education, we see that possible secondary education years are 5, 6 or 7. Therefore, we have a one percentage point increase in unemployment rate relates to a 0.4613 percentage points increase in the predicted HIV prevalence rate when secondary education duration is 5, or a 0.29892 percentage points decrease in the predicted HIV prevalence rate when secondary education duration is 7, holding all other variables in model fixed. Therefore, we can conclude that different levels of secondary education duration indeed affect the relationship between HIV prevalence rate and unemployment rate. If the secondary education duration is relatively high, the unemployment rate and HIV prevalence rate are negatively related while the unemployment rate and HIV prevalence rate are positively related when secondary education duration is relatively low. This is probably because in a country where the secondary education duration is at a high level, the unemployed people are more educated than those living in a country where the secondary education duration is at a low level. Therefore, they care more about their health conditions, which relates to a lower HIV prevalence rate.

We also add another interaction term “genderRatioXeducation” in our model since we want to see if the education effects on HIV prevalence rate varies on different levels of gender ratio. This time, we want to take the partial derivative with respect to secondary education duration. Doing so, we have $\partial HIV\ rate/\partial secondary\ = -63.79033 + 64.96604 * gender\ ratio$. Therefore, we can see that high gender ratio relates to a higher predicted HIV rate for one more year of secondary education duration, holding other variables in the model fixed. It is consistent with our prediction that secondary education becomes more significant as the gender ratio goes up due to the fact that mainly males go to school. Therefore, when gender ratio is low, it means there are more females relative to males, and since females rarely receive an education, secondary education duration is not very significant.

To take into account the fact that there will be some outliers in our data set, we also conduct a LAD regression for our final model and the result is shown on column 5. From the result, we can see that the signs of the coefficients all stay the same as before, which means that the outliers do not affect the overall trends of the best-fit lines. Also, we notice that only foreign direct investment, inflation consumer price and unemploymentXeducation are statistically significant in the LAD regression, and their coefficients are roughly the same as the OLS regression results, so we can conclude that the outliers in our dataset don't actually affect our model.

Our final regression model is regression (7), we add the time fixed effect and look at the correlation between the HIV prevalence rate and the socio-economic indicators under study, getting rid of potential bias related to time trends. After controlling for time trends, the estimated coefficients on the independent variables are similar to these in the regression (4). For example, the magnitude of the coefficient on GDP per capita increases slightly. This means that a 1% increase in GDP per capita is related to a 0.0103 percentage point increase in the predicted HIV prevalence rate, instead of 0.0094 percentage point increase, holding all other variables in the model constant.

In estimation (6) and (8), we include country fixed effect. Adding country fixed effect leaves the estimated coefficients dramatically changed, shifting the signs of some variables. The negative relationship between GDP per capita and HIV prevalence rate is, however, consistent with our hypothesis. On the other hand, the positive correlation between education and HIV prevalence rate and the positive correlation between gender ratio and HIV prevalence rate are counterintuitive. One possible reason for this inconsistency is the lack of variation in these two variables within the same country. For each country, we have at most ten observations, which provide limited predictive power. Therefore, a small fluctuation will be magnified into a significant result.

Conclusion

Our main research goal is to estimate the relationship between HIV prevalence and GDP per capita and secondary education duration. Inconsistent with our assumption, there exists a positive relationship between GDP per capita and HIV prevalence. This implies that a higher GDP per capita relates to a higher HIV prevalence, which is opposite to our assumption of higher GDP per capita relates to a lower HIV prevalence. However, with country fixed effect, we do see a negative correlation between GDP per capita and HIV prevalence rate, which is what we expected. Secondary education duration is consistent with our assumption. Secondary education duration has a negative relationship with HIV prevalence, which implies that longer secondary education duration is related to a lower HIV prevalence.

From our model, HIV prevalence is also correlated with health expenditure, gender ratio, unemployment, inflation consumer prices, foreign direct investment, unemployment*secondary education duration, and gender ratio*secondary education duration. We also observed that secondary education duration, gender ratio, and unemployment rate have the highest coefficient; thus, they have the highest magnitude of influence on HIV prevalence.

Besides the common measurements of economy such as GDP per capita, health expenditure, and foreign direct investment, our research provides a new angle towards the HIV problem by focusing more on gender ratio and secondary education duration. As a result, we would suggest that policy makers should devote more to balancing gender inequality and improving overall education level. This is because according to do our model, more educated males is related to a lower HIV prevalence rate; therefore, by encouraging females to receive more education, it will relate to a lower predicted HIV prevalence rate as well.

Some of the results from the regressions are not consistent with our prediction. For example, while we expect higher health expenditure relate to lower predicted HIV prevalence rate, the results suggest other way around. Hence, the possible lingering question can be to reevaluate the intuition behind the relations between the HIV prevalence rate and the socio-economic indicators in the African countries under study. Further research can also expand the research by adding more years, so there will be enough variations within each country to ensure country fixed effect make sense.

Works Cited

"COUNTRY COMPARISON : HIV/AIDS - ADULT PREVALENCE RATE." Central Intelligence Agency, Central Intelligence Agency. N.p., 2014. Web. 25 Feb. 2018. <https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2155rank.html#sl>

"HIV/AIDS Fact Sheet." World Health Organization. World Health Organization, Nov. 2017. Web. 24 Feb. 2018. <http://www.who.int/mediacentre/factsheets/fs360/en/>.

"HIV/AIDS." World Health Organization. World Health Organization, n.d. Web. 25 Feb. 2018. <http://www.who.int/gho/hiv/en/>.

Luboobi, Livingston S., and Joseph Y.t. Mugisha. "HIV/AIDS Pandemic in Africa: Trends and Challenges." *SSRN Electronic Journal*, 2005. doi:10.2139/ssrn.812847.

Steenkamp, Liana, Danie Venter, Corinna Walsh, and Pelisa Dana. "Socio-economic and Demographic Factors Related to HIV Status in Urban Informal Settlements in the Eastern Cape, South Africa." *African Journal of AIDS Research* 13, no. 3 (2014): 271-79. doi:10.2989/16085906.2014.952651.

"The African Regional Health Report 2014 - The Health of the People: What Works." African Health Observatory. N.p., 2014. Web. 25 Feb. 2018. <http://www.aho.afro.who.int/en/publication/1786/african-regional-health-report-2014-health-people-what-works>.

"What Are HIV and AIDS?" AVERT, 2 Jan. 2018. Web. 24 Feb. 2018.

< www.avert.org/about-hiv-aids/what-hiv-aids>.