CRED
WORKING
PAPERS

HIV, Risky Behavior and Ethno-linguistic Heterogeneity

CRED WP 2010/04

Miron Tequame

Center for Research in the Economics of Development

University of Namur

HIV, Risky Behavior and Ethno-linguistic Heterogeneity*

Miron Tequame†

March 3, 2010

Abstract

This paper studies the role of ethno-linguistic diversity on risky sexual behavior and HIV status of individuals. We compute three different measures of ethno-linguistic heterogeneity and show that ethno-linguistic heterogeneity is positively related to risky sexual behavior and HIV status of individuals. This effect, both on risky behavior and HIV status, is stronger for women than for men. Accordingly, we also find a positive association between ethno-linguistic heterogeneity and being in a discordant couple where the wife is HIV positive. Our findings have an implication for prevention policies as it identifies community level characteristics for implementing anti-HIV/AIDS policies.

JEL Classification: I18, J14, O12, O20, Z13

Keywords: HIV/AIDS, Risky Behavior, Ethno-Linguistic Diversity

*I would like to thank Jean-Marie Baland and Gani Aldashev for their valuable comments and suggestions. I am also grateful to participants of CRED Workshop-University of Namur, Doctoral Workshop-Louvain School of Economics, ECRU Seminar-UCL, Portugues National Health Conference(Porto) and 4th Annual AIID Workshop on AIDs (Amsterdam).

†CRED, Department of economics, University of Namur, Rempart de la Vierge, 8 B-5000 Namur, Belgium. Contact: miron.tequame@fundp.ac.be.

1 Introduction

In 2007 UNAIDS has estimated that over 33 million people are living with HIV/AIDS and 70% of this amount are concentrated in sub-Saharan Africa. Many studies have tried to explain why HIV has spread extensively and rapidely in sub-Saharan Africa. There is a general consensus that primary channel of HIV/AIDS transmission in Africa is heterosexual sex (Schmid et al., 2004). Indeed, the DHS Comparative Report of February 2009 states that HIV infection is associated with an increasing number of lifetime sex-partners, earlier age of sexual debut among women, sex with non-marital or non-cohabiting partners and alcohol use during last sex.

Based on raw data we learn that the spread of HIV in sub-Saharan countries is heterogeneous among regions. One reason might be the presence of other untreated and sexually transmitted diseases that favor HIV infection[1]. Another reason might be circumcision of men that reduces sexually transmitted diseases (Weiss, Quigley, and Hayes, 2000). There is also a role played by economic activity, for example through exports, on HIV infection, as emphasized in Oster (2009). However, in contrast to large number of studies examining individual-level determinants, few studies examine community characteristics that affect HIV epidemics.

In sub-Saharan Africa the literature states that there is little response in risky sexual behavior due to HIV epidemics(Stoneburner and Low-Beer, 2004; Bloom et al., 2000; Williams et al., 2003). However, response in risky behavior is heterogeneous among different individuals based on their socio-economic status like education of women(De Walque, 2007; Dupas, 2009), future life expectancies and future expected revenues(Oster, 2007a). To this regard, the most prevalent prevention policy, in Africa, is the anti-HIV education labeled ABC: Abstain, Be faithful and use Condom. ABC policy has proven to be effective in Uganda((Green et al., 2006)[2] and

[1]The high rate of HIV transmission in Africa si due to other untreated sexually transmitted diseases. The difference in transmission rates is large enough to explain the observed difference in prevalence between the United States and Sub-Saharan Africa (Oster 2005).

[2]However, Oster 2009 have given alternative possible reasons why ABC was claimed to be effective in Uganda. She explains that routes of exports have an important role rather than

it's being extended to other sub-Saharan countries without, up to now, any clear evidence of success. Sexual behavior has been one of the major focuses of HIV prevention efforts and understanding changes or its characteristic should be an important base for predicting the future path of the epidemic.

The aim of our paper is to analyze the diverse rate of HIV prevalence and the behavioral factors which underpin the disease to give important insights for understanding and designing effective HIV prevention policies. We study the relationship between HIV, risky sexual behavior and ethno-linguistic heterogeneity within regions, a community based characteristic. The role of heterogeneity, in terms of ethnicity, race and religion, on social capital and consequently on economic variables has been discussed in the literature[3]. It is associated with lower trust among communities because members of the community have different tastes and also because it is harder to enforce a system of social sanctions. Hence, ethno-linguistic heterogeneity has an implication on risky sexual behavior, taken as a deviating behavior from social custom, and therefore on HIV infection. Another possible route stems from information asymmetry that arises in such communities, due to different languages and cultures, thus implying a lower probability of detecting risky sexual behavior. For our analysis we compute three measures of ethno-linguistic heterogeneity widely used in the literature: ELF based on the herfindahl index, GELF based on a similarity matrix of ethnicity among couples and the Entropy index.

We use the Demographic and Health Survey datasets that give us the possibility to link the HIV status of an individual with her ethnic characteristic. We choose four sub-Saharan countries (Malawi, Cameroon, Kenya and Ethiopia, in order of HIV prevalence) based on data availability and relevant HIV prevalence. The period of the survey ranges between 2003 and 2005. There are several advantages of the DHS: it provides the most accurate estimate of HIV prevalence in the population;

population based educational campaign.

[3]For example, Alesina and La Ferrara (2000) claim that ethnic, racial and economic heterogeneity have a negative impact on social capital and consequently on economic development in the USA

individuals are not aware of their status allowing for analysis of the determinants of HIV; each individual's HIV status is linked to socio-economic characteristics of the individuals including their partner.

Recently, Pongou (2009)developped a theory of how community-level ethnic heterogeneity determines the formation of sexual networks, and how this, in turn, affects the spread of HIV/AIDS. In his paper agents derive utility from sexual relationships but, similarly to our hypothesis, sexual infidelity is punished if detected by own partner. When information circulates more easily within ethnic groups than across, agents tend to choose their partners from different groups to hide their infidelity, due to cross-group anonymity. Hence, in his paper, he emphasizes the role of information asymmetry on risky sexual behavior and provides an empirical evidence of his theorethical model. In line with this strand of research, our paper focuses more on the gender related effect of ethnolinguistic heterogeneity on HIV status of individuals, without stressing the exact theoretical mechanisms that lie behind this association.

To preview our result, first, we find a positive association between extramarital sex and ethno-linguistic fractionalization. This positive association holds also on HIV status of individuals, especially that of women. Second we disentangle the direct and indirect effect of ethno-linguistic heterogeneity on HIV status, through own behavior and that of partner, by further analysing the siero-positivity of individuals in a couple. We find a positive relationship between our indicators of ethno-linguistic heterogeneity and the probability of being in a discordant couple, to which we refer as a proxy for an objective measurement of effective extramarital sex. Third, we find a positive impact on discordant couples where the wife is HIV positive and the husband not. This implies that the effect of ethno-linguistic heterogeneity is gender related.

Our paper contributes in understanding community level characteristics which might hinder or favor the epidemics. Understanding communities that might be at poten-

tial risk of HIV reduces costs in terms of prevention policies.

The paper is organized as follows: in Section 2 we try to give an outline of the theory that lies behind our story; Section 3 show our empirical strategy and results. A sensitivity analysis is conducted in section 4. Finally, some conclusions are laid off in section 5.

2 Conceptual Framework

Ethnic fractionalization is believed to have an important role in the political economy of many countries as it might lead to political instability, poor quality of institutions, badly designed economic policy and poor economic performance. Several studies have analyzed the negative relationship between racial or ethnic heterogeneity and the provision of public goods (Alesina et al., 1999; Miguel and Gugerty, 2005a) while others have emphasized the impact on productivity(Bandiera et al., 2005; Alesina and La Ferrara, 2005). In a comparison across US counties, it is found that higher ethnic fractionalisation is associated with a lower rate of collective action in the communityVigdor (2004). There are also studies that have linked ethnic fractionalisation and trust and pariticipation in communitiesAlesina and La Ferrara (2000).

In this study, we analyse the association between ethno-linguistic fractionalization and risky sexual behavior. It goes beyond the objective of this paper in explaining the exact mechanisms that lie behind this relationship. However, in this section we discuss some of the possible mechanisms that might link sexual (risky) behavior and ethno-linguistic fractionalization.

First, an important role might be played by collective social sanctions. We assume that an individual's risky sexual behavior is subject to social sanctions, specially from own spouse and independently from specific cultural characteristics of any community. By risky sexual behavior we mean any kind of sexual encounter that

is outside own cohabiting partner or husband. The importance of collective punishment or reputational mechanisms in developing countries has been well argued in economics since the 70s[4], especially where there is lack of formal institutions(Greif, 1993). Social sanctions or social exclusions are costly in developing countries and they are part of constraints people take into account when deviating from social custom. Moreover, social sanctions based on social exclusion are costly to individuals especially if the proportion of individuals accepting the code of sanction and willing to exclude is higher. There are already studies that have linked ethnicity and collective social sanctions. In rural western Kenya, ethnic diversity has been associated with *fewer recorded community social sanctions*Miguel and Gugerty (2005b). Hence, one mechanism might be that social sanction due to risky behavior is less costly in heterogeneous societies rather than homogeneous ones.

Second, the role of information and interaction in a community is crucial for risky behavior. To be subject to social sanction individuals should be detected as having risky sexual behavior. The probability of being detected, having risky sexual behavior like extramarital relation, is higher in homogeneous societies rather than in heterogeneous societies. This is because homogeneous societies are more likely to have the same language, culture and networks making flow of information on people's characteristic and behavior easier. Individuals, who want to keep risky sexual behavior secret, might find heterogeneous societies more favorable. Furthermore social interaction provides information about the level of HIV/AIDS at community level including infectious status and risky behavior of partners(Fiske and Taylor, 1991; Rabin, 1998; Watkins and Schatz, 2001; Entwisle et al., 1996; Kohler et al., 2001; Behrman et al., 2002)[5] . Besides, one of HIV prevention campaigns is: *avoid*

[4]Akerlof (1979) has introduced reputation as a constraint where he shows that social customs, which are disadvantageous to the individual, may nevertheless persist without erosion, if individuals are sanctioned by loss of reputation for disobedience of the custom.

[5]Experimental studies have revealed strong influences of peers on risk assessments, expectations and beliefs (Fiske and Taylor 1991; Rabin 1998; Watkins and Schatz 2001). Social networks play an important role in shaping the diffusion of innovations in developing countries including the use of contraceptive (Behrman et al 2002). Moreover, qualitative data from Thailand and Kenya provide evidence that women chat with each other about family planning and AIDS (Entwisle et

partners who have multiple partners. This plays an important role on infectious status of individuals.

Third, economic development might act as a third mechanism. Economic development attracts different ethnicities from nearby regions through migration and employment. It also affects the demand and supply of risky sexual behavior (as sex is a normal good) for different reasons, ranging from commercial sex, leisure, consumption of alcohol etc..

On the other side, to better understand which of these mecchanisms have a role on HIV and risky sexual behavior we need to define ethno-linguistic heterogeneity. Measuring, and thus defining, ethno-linguistic heterogeneity in sub-Saharan countries is challenging. The traditional measure based on the Herfindahl index, the so called Ethno-Linguistic Fractionalization (ELF), is given by the probability that two randomly drawn individuals from the population belong to two different groups. If we consider a community composed of more than two different ethnic groups, with s_i being the share of group i over the total population at community level and n being the total number of ethnicities, then the ELF index would be:

$$ELF = 1 - \sum_{i=1}^{n} s_i^2$$

Nevertheless, continuous regional migrations and inter-mixing with other groups suggest ethno-linguistic fractionalization is a function of migration, colonial policy and intera-group mixing(Alesina et al., 2003).

In some societies there are different similarities among different ethnic groups which compose the community. The ELF index gives the same weight to all ethnicities, while in reality some might have similare cultural values which should be taken into account without necessarily being equal or completely different among each ethnicities. The ELF index attributes a 0 or 1 value to all individuals in the community

al 1996, Kohler et al 2001).

whether they belong or not to an ethnicity or race. To this purpose, a generalized index of fractionalization has been proposed in Bossert et al. (2008). With this indicator, the authors propose a matrix of similarity where each individual is a vector containing similarity values with all other individuals in the community. Similarity measurment can be various ranging from income, education, language and etc.

As we are studing the effect of social interactions through ethno-linguistic heterogeneity, we find this indicator appropriate to our analysis. To compute our indicator of similarity, we use a matching system of ethnicities on the dataset of couples. In our hypothesis, if a high proportion of people from ethnicity «x» gets married to another ethnicity «y» then similarity between these two ethnicities is non zero in terms of cultural values and networks. To compute similarities between two ethnicities, we take the average between four frequencies: frequency of wife from ethnicity «x» getting married to husband from ethnicity «y»; frequency of wife from ethnicity «y» getting married to husband from ethnicity «x»; and similarly for the husband. To give you some illustration on the indicator we construct, let us consider a comunity with 3 ethnicities. The similarity matrix, S, of couples' matching based on the ethnicity of the husband and wife is as follows:

$$S = \begin{pmatrix} 1 & s_{x,y} & s_{x,z} \\ s_{y,x} & 1 & s_{y,z} \\ s_{z,x} & s_{z,y} & 1 \end{pmatrix}$$

where $s_{x,y} = s_{y,x} = avg\left[f^h_{x,y}, f^h_{y,x}, f^w_{x,y}, f^w_{y,x}\right]$ and $f^h_{x,y}$ denotes the proportion of husbands «x» getting married to women of «y» (subscript h stands for husband while w for wife). The matrix is symmetric and equal to 1 on the diagonals as similarity with own ethnicity is 1. We construct a similarity matrix as above for each enumeration area in our dataset. Based on the above matrix, S, the generalized ethno-linguistic fractionalisation indicator, as proposed in Bossart et al (2008), is as follows:

$$GELF = 1 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{i,j}$$

where $i, j \in \{x, y, z\}$ and $s_{i,j}$ is an element of the matrix S.

Another index of diversity which we compute is the Entropy Index (also called the Shannon index) used for species to measure biodiversity. The ELF index gives higher weights to higher proportion of ethnicities as it has a quadratic form while the Entropy Index (EI) gives a decreasing weight as the proportion of a specific ethnicity increases in a community. It takes a maximum value when all the ethnicities are equally represented in the community. We compute a normalized EI, between [0,1] , in each enumeration area:

$$EI = \frac{-\sum_{i=1}^{n} s_i ln(s_i) - \left[\frac{(n-1)}{2N}\right]}{ln(n)}$$

where s_i is the proportion of ethnic i in the community, n is the number of ethnicities in the community and N is the total number of individuals in the community.

We computed ethno-linguistic in each enumeration areas of the population census by exploiting data at individual level. We compute the ELF, GELF and EI on ethnicities in each enumeration area. There are a total of 50 ethnicities in Cameroon, 67 in Ethiopia, 15 in Kenya and 9 in Malawi. The correlation between the three indicators we computed is 0.79 between ELF and GELF, 0.94 between ELF and EI and 0.75 between GELF and E.

We consider religious fractionalization on a total of 8 different religions in Cameroon, 6 in Ethiopia, 5 in Kenya and 8 in Malawi (including those without religion and those classified as other). We compute religious fractionalization, based only on the Herfindahl Index to capture any unobserved heterogeneity due to the concentration of specific religions that are more likely to be correlated with ethnic fractionalization. For example, we believe that being a Muslim in the Far North Region of Cameroon,

where 75% of the population are Muslim, is probably different from being a Muslim in any other region where presence of the religion is much lower.

3 Empirical Strategy and Results

We use DHS household surveys: Cameroon 2004, Ethiopia 2005, Kenya 2003, Malawi 2004. These are surveys conducted in a number of countries beginning from the late 1980s and focusing on fertility, contraception and child health. Lately, the AIDS Indicator Survey (AIS) has been conducted together with the DHS and questions about sexual behaviors like extramarital sex, premarital sex, sex within marriage and their HIV status has been included. There are 9 recent waves of surveys between 2003-2006 that include both ethnicity and AIS. However, we focus on 4 countries where HIV prevalence is at least 2% at national level. Due to spatial correlation in HIV prevalence we focus on countries that are broadly comparable.
The number of observations in our sample is 35,933 individuals where women represent 52.5%. Age of eligible indivuals, for sexual behavior and HIV, ranges between 15-49 for women and 15-59 for men.

In all specifications we use different controls to capture unobserved heterogeneity at country/regional/urban/rural/enumeration area level. Other controls based on socio-economic characteristics at individual level are also included: education, wealth, age, marital status, religion and proportion of people in the richest and richer wealth quintile in each enumeration area to capture local economic development. We provide summary statistics of these variables in Table 1.

3.1 HIV and Risky Behavior

In our sample, a total of 17% have declared at least one extramarital sex in the last 12 months with relevant heterogeneity among countries: 34% in Cameroon and

5% in Ethiopia. Out of the total number of men sampled, 25% declared at least one extramarital sex versus 11% of women, showing that women reported less risky behavior than men.

There are numerous factors that might influence individual's risky sexual behavior: general individual characteristics (age, education, wealth, marital status, health, preferences for sexual partners and susceptibility to HIV), social characteristics (communities' ethnic or religious identities), knowledge and exposure to HIV/AIDS and cost of HIV infection and prevention. Our basic model evaluate the impact of ethno-linguistic heterogeneity on risky behavior:

$$s_{i,c} = \alpha + \beta H_c + \gamma X_{i,c} + \delta D_r + \varepsilon_{i,c}$$

where $s_{i,c}$ is the number of «*declared*» extramarital relations of individual i in community c, H_c is ethno-linguistic and religious heterogeneity in the community, $X_{i,c}$ is a set of individual control variables in community c and D_r is a set of dummies at regional level. There are a total of 34 regional dummies based on the definition of the DHS survey: 12 in Cameroon, 11 in Ethiopia, 8 in Kenya and 3 in Malawi. These dummies capture all common cultural and economic factors including regional HIV prevalence, perception and social cost of infection like stigmatization. The individual control variables are education, wealth, age, marital status, religion.

Table 2 reports results from our first Poisson regression of «*declared*» extramarital relation on heterogeneity and other control variables. The database includes all eligible men and women restricted to those enumeration areas with at least five[6] individuals sampled and 70% of them residents for at least 10 years. This gives us the possibility to clean the effect of ethno-linguistic heterogeneity due to the presence of

[6]We restrict to 5 the number of individuals in the enumeration area such that our indicators of ethnolinguistic heterogeneity are more appropriate. The total number of individuals sampled in the enumeration area range between 1 and 60. Results are not conditional on the minimum number of people sampled. However, increasing this number leads to ruling out many rural areas from our analysis.

migrants. We find a positive significant impact of ethno-linguistic heterogeneity on the number of «*declared*» extramarital relation. It is worth to note that if we further distinguish by gender, the effect of ELF on declared risky behavior is driven by the sample of women and specifically married women. Married women report higher number of extramarital sex in heterogenous societies. The association is stronger if we relax our above restrictions on the sample. Particularly, as you can observe in Column 1 of Table 2, significance clearly emerges if we run our regression on the whole sample and include a dummy for being a migrant or not, where by migrant we mean individuals who are resident for less than 10 years.

In Panel B of Table 2, we replicate our analysis on the sample of individuals who have declared a positive number of extramarital sex. This allows us to rule out potential endogeneity due to under-reporting. It is quite reasonable that declaring one or superior extramarital sex is less correlated with under-reporting; given that ethno-linguistic fractionalisation could be a source for under-reporting, we may look at Panel B as a further robustness check. Summing up results, the gender related impact of ethno-linguistic heterogeneity is reinforced. However, a strong limitation of these regressions is due to the number of oberservation in the restricted sample. We do not consider individual HIV risk perception explicitly but regional dummy should capture regional level HIV prevalence. These regional dummies capture also differences in ethno-linguistic and religious fractionalization between different regions, while our community level heterogeneity captures differences within regions. For the sake of knowledge, once we controlled for ethno-linguistic heterogeneity we found no significant impact of religious heterogeneity on «*declared*» extramarital relations.

The regressions on declared risky behavior are to be taken with caution. Many studies have underlined that self-reported variables regarding private information like sexual behavior are not truthfully reported[7](Gersovitz et al., 1998; Glynn et al.,

[7]Glynn et al. (2001) showed that 12% of women who reported being virgins were HIV positive and some had other sexually transmitted infections.

2001).

To give you more illustration on reported sexual behavior we analyze our data and show that the number of extramarital sex, declared by both married or cohabiting men and women, is inconsistent with their recent sexual activity in the last four weeks. To this purpose, we use the data on couples where information on their sexual activity in the last 4 weeks and number of extramarital sex in the last 12 months is available both for the husband and wife. In our dataset, we have a total of interviewed 7,698 couples. Restricting the sample to monogamous couples, in Table 3 we construct a two-way table for being sexually active in the last 4 weeks. We show that there are over 1,046 cases where the declaration of sexual activity is discordant. Indeed, in a monogamous couple, if one declares not to be active in the last 4 weeks while the partner declares so, then the partner should be more likely to declare at least one extramarital sex in the last 12 months. In the Table 4, we compare the above statistics with their extramarital sex in the last 12 months. We learn that out of 513 cases, where the wife was sexually active and the husband not, 99% of these women declared 0 extramarital sex in the last 12 months. Extramarital sex is under-reported for men as well: only 17.5% of them declare at least one extramarital sex. It is interesting to note that the discordances in recent sexual activity is, approximately, *equally* distributed among men and women (533 versus 513). These means that they were both negligent in reporting recent sexual activity or extramarital sex in the last 12 months, leading to inconsistency. Another explanation is that both husband and wife under-report extramarital sex.

However, to disentangle this problem, we consider extramarital sex in the last 12 months as *«declared»* extramarital sex rather than effective extramarital sex. As we have mentionned above, our regressions suffer from endogeneity due to reporting biases. Individuals tend to under-report extramarital sex in homogeneous communities and this creates an upward bias on our variable of interest.

To tackle the problem of endogeneity that may arise, from next section on, we test

our hypothesis on HIV status of individuals; being HIV a non self-reported variable and a more objective measurement of risky sexual behavior.

3.2 HIV status and Ethno-linguistic Fractionalisation

Information on the HIV status of individuals is based on voluntary testing and respondents are not aware of the result. HIV status is reliable information compared to self-reported variables. However, some individuals have refused to be tested. Average response rate is higher for women rather than for men. It is also higher in rural areas rather than urban areas. Overall response rate is more than 70%. By comparing mean differences of observable characteristics of women who accepted to test and those who refused, it seems that women who refused are more likely to be HIV positive(Juhn et al., 2009)[8]. This means data on HIV are probably underestimated. HIV prevalence rate ranges between 1.3% in Ethiopia and 12.3% in Malawi[9]. It is higher among women (1.7% in Ethiopia and 13.9% in Malawi) rather than men and in urban areas rather than in rural areas. We consider HIV status of both men and women as a function of ethno-linguistic heterogeneity. In this case, we consider the direct and indirect effect of ethno-linguistic heterogeneity, through own risky behavior or that of partner's.

We replicate the above analysis on the probability of being HIV positive. Our model can be rewritten as follows:

$$HIV_{i,c} = \alpha + \beta H_c + \gamma X_{i,c} + \delta D_r + \varepsilon_{i,c}$$

We are aware that most of the individual characteristics as control variables might not be completely exogenous. Location in urban or rural areas, education and

[8]Juhn, Kalemli-Ozcan and Turan (2009) find that refusers are more likely to be educated, less likely to live in rural areas and more likely to be in the wealthiest quintile. This is a similar pattern which separates HIV positive and HIV negative women implying that HIV prevalence is underestimated.

[9]All HIV prevalence rates in this paper are calculated by using the DHS sample weights.

wealth might be, to some extent, choice variables for the individual or her family. Fore example, Antiretroviral treatment is not observed and it might create bias if it is correlated to education, wealth and if living in urban areas. The amount of bias depends on the proportion of people who were aware of their HIV status and had the treatment while being surveyed. Endogeneity arises if individuals migrate, due to their HIV status or ARV treatment, in more heterogeneous areas. We run our regressions on HIV status on a sample of people who have been living in their place of residence at least for 10 years and them being at least 70% of the community.

Table 5 report results on the effect of ELF on HIV status and on HIV prevalence at enumeration area level. For ease of interpretation, we report OLS regression coefficients. An increase of 1 percent in ethno-linguistic heterogeneity increases the probability of being HIV positive by 0.05 percent. Alternatively, it increases HIV prevalence rate at enumeration area level by 0.058. The impact is reduced if we add control variables, but significance always holds.

In Table 6, we replicate probit regressions of HIV status on ethno-linguistic heterogneity and we differentiate by gender as previously for declared extramarital sex. In Column 1, we run the regression for all the sample, obtaining strong significance for ELF and EI indicators. In Column 2, by restricting the sample to enumeration areas in which the number of people sampled is superior to five and the proportion of migrants is less than 30 percent, only the ELF indicator remains significant. From Column (3) to (5), the sample is restricted to women, married and not married while from column (6) to (8) the sample is restricted to that of men. We obtain that ethno-linguistic heterogeneity has a positive impact on HIV status of women, especially if married. The same results cannot be drawn for men.

In Table 7 we report probit regression on all the sample by considering the interaction terms between our indicators of ethno-linguistic heterogeneity and a dummy for women; gender related effect becomes more evident, confirming the above results in the previous sections.

To better understand the difference between the siero-status of men and women, i.e. due to own behavior or that of partner's, we are going to concentrate on couples by analyzing concordant and discordant couples from the next section on.

3.3 Discordant HIV status of Couples and Ethno-linguistic Fractionalization

In the dataset of couples we consider cases where HIV is discordant, i.e. either the wife or the husband is siero-positive. Table 8 shows the number of cases where HIV status is discordant among couples. Out of 7,171 couples interviewed with information on their siero-status, 5.5% of them are in a discordant couple. There are a total of 424 women and 411 men who are HIV positive with married or cohabiting marital status. A total of 191 women and 171 men are in a discordant status. Discordance among couples might occur due to risky behavior before or after marriage. It also depends if the couple is polygamous. It is interesting to note that, approximately, all HIV positive married women or men are equally divided among concordant and discordant status. If sample is restricted to monogamous couples, married for a determinated period of time, HIV discordant cases are good proxies for effective extramarital sex.

In Table 9 we report results of our analysis on discordancy. In a first moment, we consider the impact of ethno-linguistic heterogeneity on the probability of being in a disconcordant couple. The probability of being in a discordant couple measures approximately risky behavior of both the husband and wife provided that we control for the period of time the couple is married. In a second moment, we separate the effect on discordancy, whether it is the wife or husband who is HIV positive. We find that there is a positive effect of ethnolinguistic heterogeneity on the probability of being in a discordant couple and this effect is stronger for dicordancy when the wife is HIV positive.

4 Sensitivity Analysis

Ethno-linguistic heterogeneity and public goods provision has been discussed in the literature through its role on conflicts, different preferences and tastes(Alesina et al., 1999; Alesina and La Ferrara, 2000; Alesina et al., 2003; Alesina and La Ferrara, 2005)(Alesina et al., 1999; Alesina and La Ferrara, 2000; Alesina et al., 2003; Alesina and La Ferrara, 2005). In our analysis a possible mechanism, through which ethno-linguistic heterogeneity affects HIV status, might be due to public policies or public good provision at community or regional level. For example, the effect of ethno-linguistic heterogeneity on HIV status of an individual might be driven by bad public health provision or bad prevention campaign due to lack of collective actions. In light of it, the regional and rural/urban dummies should capture any difference in the provision of public policies related to HIV/AIDS epidemics at regional and urban/rural level.

Moreover, a common characteristic of African countries is the role played by policy makers in enhancing regional and ethnic favoritism in the allocation of national government funds or public good provision. Again, the dummies should capture any potential bias due to ethnic favoritism in addressing goverment funds against HIV at regional level. As a further control, we replicate our result by controlling for ethnic fixed effects rather than regional fixed effects. Results hold except for the GELF indicator of ethno-linguistic heterogeneity.

We also check if our result is not lead by migration in a way that the presence of migrants in an area might affect ethno-linguistic heterogneity: at the same time migrants might practice high risk sexual activity with respect to non migrants. To this purpose we restrict our sample to residents for at least 10 years. Ten years is the maximum number which allows us to identify those who have contracted the virus in their actual place of residence.

Finally, we also test if our results are robust compared to urban and rural areas.

DHS has a comparative definition of rural and urban areas. This allows us to ensure that our results are not triggered by differences between urban and rural areas. Even population density might be conducive to our results: a proper measure of population density could be represented by the number of people sampled in the enumeration area.

We do not report results for these robustness check. However, in any case all these controls support our findings[10].

5 Conclusion

This paper is based on the most recent nationally representative survey of four sub-Saharan countries, which are spatially comparable to determine the effect of ethno-linguistic fractionalization on HIV epidemics. We find a positive association between heterogeneity and risky sexual behavior of individuals. Our result holds both for «*declared*» extramarital sex and HIV status.

Another important finding of our paper is the gender related effect of heterogeneity on HIV. It seems that women's behavior is more shaped by social heterogeneity rather than men's. This holds for «*declared*» extramarital sex of women, HIV status and probability of being in a discordant couple where she is HIV positive. The effect on «*declared*» extramarital sex of women could be intuitive to the extent that social sanction due to declaring risky sexual behavior is more costly for women rather than men. Conversely, the effect on HIV status is less intuitive as ethno-linguistic heterogeneity seems not to affect men's siero-status. As the probability of being detected increases, when engaging in risky sexual behavior due to social homogeneity, women are less probable to be infected with HIV and finally less probable to be in a discordant couple where she is HIV positive. Ethno-linguistic heterogeneity also affects risky sexual behavior and HIV status of men; however, men are biologically

[10] All the tables containing the results are available upon request.

less exposed to HIV infection rather than women; moreover, higher proportion of men have refused to be tested and this might under-estimate HIV prevalence among men. A further explanation of our finding might be the opportunity that ethno-linguistic heterogeneity gives to HIV positive women to keep confidentiality on their status and use strategic behavior with their husband in order not to infect him. Notwithstanding, it also implies that she is aware of her status.

Pongou (2009) stressed out the role that ethno-linguistic heterogeneity plays on fidelity and social network formation among couples. We put forward an original result by emphasizing the difference between women and men. It is more likely that, not only an information asymmetry, but also a reputational mechanism plays a role on women's behavior. We also disentangle the global effect of ethno-linguistic heterogeneity on own status by differentiating among discordant couples where either she or he is HIV positive.

Our paper takes account of some characteristics at community level, which are helpful to address some specific prevention policies when, for example, Antiretroviral therapy is provided.

References

Akerlof, G., 1980. A theory of social custom, of which unemployment may be one consequence. The Quarterly Journal of Economics, 749–775.

Alesina, A., Baqir, R., Easterly, W., 1999. Public Goods and Ethnic Divisions*. Quarterly Journal of Economics 114 (4), 1243–1284.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. Journal of Economic Growth 8 (2), 155–194.

Alesina, A., La Ferrara, E., 2000. Participation in heterogeneous communities. Quarterly Journal of Economics 115 (3), 847–904.

Alesina, A., La Ferrara, E., 2005. Ethnic diversity and economic performance. Journal of Economic Literature 43 (3), 762–800.

Anarfi, J., 2004. Sexuality, migration and aids in ghana.

Bandiera, O., Barankay, I., Rasul, I., Street, H., Street, G., 2005. Cooperation in collective action. Economics of transition 13 (3), 473–498.

Behrman, J., Kohler, H., Watkins, S., 2002. Social networks and changes in contraceptive use over time: Evidence from a longitudinal study in rural kenya. Demography, 713–738.

Bloom, S., Banda, C., Songolo, G., Mulendema, S., Cunningham, A., Boerma, J., 2000. Looking for change in response to the aids epidemic: trends in aids knowledge and sexual behavior in zambia, 1990 through 1998. JAIDS Journal of Acquired Immune Deficiency Syndromes 25 (1), 77.

BOSSERT, W., D'AMBROSIO, C., LA FERRARA, E., 2008. A generalized index of fractionalization (2008-01).
URL `http://ideas.repec.org/p/mtl/montde/2008-01.html`

Brockerhoff, M., Biddlecom, A., 1999. Migration, sexual behavior and the risk of hiv in kenya. International Migration Review, 833–856.

De Walque, D., 2007. How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. Journal of Development Economics 84 (2), 686–714.

Dupas, P., 2009. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. NBER Working Paper.

Entwisle, B., Rindfuss, R., Guilkey, D., Chamratrithirong, A., Curran, S., Sawangdee, Y., 1996. Community and contraceptive choice in rural thailand: a case study of nang rong. Demography, 1–11.

Fiske, S., Taylor, S., 1991. Social cognition.

Gersovitz, M., Jacoby, H., Dedy, F., Tape, A., 1998. The balance of self-reported heterosexual activity in kap surveys and the aids epidemic in africa. Journal of the American Statistical Association, 875–883.

Glynn, J., Carael, M., Auvert, B., Kahindo, M., Chege, J., Musonda, R., Kaona, F., Buvé, A., 2001. Why do young women have a much higher prevalence of hiv than young men? a study in kisumu, kenya and ndola, zambia. Aids 15, S51.

Green, E., Halperin, D., Nantulya, V., Hogle, J., 2006. Uganda's hiv prevention success: the role of sexual behavior change and the national response. AIDS and Behavior 10 (4), 335–346.

Greif, A., 1993. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. The American Economic Review, 525–548.

Juhn, C., Sebnem, K.-O., Turan, B., 2009. Hiv and fertility in africa: First evidence from population based surveys (4473).
URL http://econpapers.repec.org/RePEc:iza:izadps:dp4473

Kohler, H., Behrman, J., Watkins, S., 2001. The density of social networks and fertility decisions: Evidence from south nyanza district, kenya. Demography, 43–58.

Miguel, E., Gugerty, M., 2005a. Ethnic diversity, social sanctions, and public goods in Kenya. Journal of Public Economics 89 (11-12), 2325–2368.

Miguel, E., Gugerty, M. K., December 2005b. Ethnic diversity, social sanctions, and public goods in kenya. Journal of Public Economics 89 (11-12), 2325–2368. URL http://ideas.repec.org/a/eee/pubeco/v89y2005i11-12p2325-2368.html

Oster, E., 2005. Sexually transmitted infections, sexual behavior, and the hiv/aids epidemic. Quarterly Journal of Economics 120 (2), 467–515.

Oster, E., 2007a. Hiv and sexual behavior change: Why not africa? NBER Working Paper.

Oster, E., 2007b. Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa.

Pongou, R., 2009. Anonymity and Infidelity: Ethnic Identity, Strategic Cross-Ethnic Sexual Network formation, and HIV/AIDS in Africa. Mimeo. Department of Economics, Brown University.

Rabin, M., 1998. Psychology and economics. Journal of Economic Literature, 11–46.

Schmid, G., Buvé, A., Mugyenyi, P., Garnett, G., Hayes, R., Williams, B., Calleja, J., De Cock, K., Whitworth, J., Kapiga, S., et al., 2004. Transmission of hiv-1 infection in sub-saharan africa and effect of elimination of unsafe injections. The Lancet 363 (9407), 482–488.

Stoneburner, R., Low-Beer, D., 2004. Population-level hiv declines and behavioral risk avoidance in uganda. Science 304 (5671), 714–718.

USAID, 2009. Levels and spread of hiv seroprevalence and associated facors: Evidence from national household surveys. Tech. rep., USAID: DHS COMPARATIVE REPORTS 22.

Vigdor, J., 2004. Community composition and collective action: Analyzing initial mail response to the 2000 census. The Review of Economics and Statistics 86 (1), 303–312.

Watkins, S., Schatz, E., 2001. Avoiding aids: global advice and local strategies, 28–31.

Weiss, H., Quigley, M., Hayes, R., 2000. Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. Aids 14 (15), 2361.

Williams, B., Taljaard, D., Campbell, C., Gouws, E., Ndhlovu, L., Van Dam, J., Carael, M., Auvert, B., 2003. Changing patterns of knowledge, reported behaviour and sexually transmitted infections in a south african gold mining community. AIDS 17 (14), 2099.

Table 1: Descriptive Statistics by Country

	HIV	RISKY BEHAVIOR	MIGRANT	MUSLIM	CHIRISTIAN	ANIMIST	NO SCHOOLING	PRIMARY EDUC	SECONDARY	HIGHER	POOREST	POORER	MIDDLE	RICHER	RICHEST
CAMEROON															
mean	.054123	.6236115	.4432788	.1680275	.7267093	.0270536	.1502108	.3969997	.4171978	.0355917	.1552113	.1806059	.2242377	.2170801	.222865
sd	.2262713	1.640.396	.4967966	.3739091	.4456708	.1622477	.3572954	.4892999	.4931203	.185279	.3621237	.3847102	.4170997	.4122778	.4161889
N	10199	10173	10199	10165	10165	10165	10199	10199	10199	10199	10199	10199	10199	10199	10199,00
ETHIOPIA															
mean	.0191855	.0632655	.1770339	.3337227	.6436191	0	.5069677	.2674192	.1912028	.0344103	.2031141	.1536785	.1451148	.1444142	.3536785
sd	.1371829	.4360931	.3817122	.4715603	.4789482	0	.4999709	.4426301	.3932637	.1822877	.4023324	.3606543	.3522303	.3515229	.4781295
N	11050	12819	12845	12843	12843	12843	12845	12845	12845	12845	12845	12845	12845	12845	12845,00
KENYA															
mean	.0659128	.2431935	.4322267	.1104399	.848063	0	.1221624	.5357565	.2542973	.0877838	.1541792	.1586406	.1724183	.2016796	.3130823
sd	.2481496	.7171129	.495418	.3134579	.3589834	0	.3274948	.4987526	.4354941	.2829987	.3611442	.3653646	.3777684	.4012805	.4637779
N	6190	7603	7621	7615	7615	7615	7621	7621	7621	7621	7621	7621	7621	7621	7621,00
MALAWI															
mean	.126044	.151377	.4337509	.1409307	.6679962	0	.175019	.6328778	.1828018	.0093014	.1516705	.2034928	.2327259	.2285497	.1835611
sd	.3319305	.5402409	.4956387	.3479831	.4709771	0	.3800194	.482066	.3865407	.0960035	.3587352	.4026341	.4226091	.4199384	.3871626
N	5268	5265	5268	5265	5265	5265	5268	5268	5268	5268	5268	5268	5268	5268	5268

Table2: Effect of Ethno-linguistic Heterogeneity on Number of Extramarital Sex in the last 12 months

A. Dependent Variable: Number of Extramarital Sex in the last 12 months							
POISSON			*WOMEN*			*MEN*	
SAMPLING	**TOTAL (1)**	**TOTAL (2)**	**MARRIED (3)**	**NON MARRIED (4)**	**MALE (5)**	**MARRIED (6)**	**NON MARRIED (7)**
ELF	0.00617**	0.0129***	0.0148***	0.00620	0.00416	0.00712*	0.00181
	(0.00311)	(0.00461)	(0.00467)	(0.00404)	(0.00339)	(0.00385)	(0.00425)
Observations	17190	9099	6065	3034	8091	4548	3543
B. Dependent Variable: Number of Extramarital Sex in the last 12 months>0							
POISSON			*WOMEN*			*MEN*	
SAMPLING	**TOTAL**	**TOTAL**	**MARRIED**	**NON MARRIED**	**MALE**	**MARRIED**	**NON MARRIED**
ELF	0.00124	0.00560***	6.27e-05	0.00642***	-0.000861	0.00127	-0.00300
	(0.00217)	(0.00191)	(0.00248)	(0.00237)	(0.00253)	(0.00288)	(0.00315)
Observations	1651	515	*138*	377	1136	299	837

Notes: We run two different set of regressions: on all declared extramarital and only on those with positive number of extramarital. All regressions include 34 regional fixed effects, religious heterogeneity, religion, wealth, education, age and if rural or urban area. Results hold even if we include ethnic fixed effects and other socio-economic characteristics. Sample is restricted to those areas where the number of people sampled, in the enumeration area is superior to 5, and the proportion of migrants in the enumeration area is less than 30%. Results are stronger if we relax this restriction. Even though we show results from Poisson regression, all results hold with OLS and Negative Binomial. Column 1 is based on the restricted sample while columns 2-4 are further restrictions based on the sample of women, married and non-married, and the last three on that of men, married and non-married.

In all regressions, samples weights are used and robust standard errors clustered at the enumeration area level.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Two way table of sexual activity among monogamous couples

		wife		
		Not Active	*Active*	*Total*
Husband	*Not Active*	753	**513**	1,266
	Active	**533**	4,724	5,257
	Total	1,286	5,237	6,523

Table 4: Extramarital Sex among couples with discordant recent sexual activity

Wife's Extramarital Sex among couples with discordant sexual activity

Extramarital Sex	Frequency	Percent
0	508	**99.03**
1	5	0.97
Total	513	100

Husband's Extramarital Sex among couples with discordant sexual activity

Extramarital Sex	Frequency	Percent
0	440	**82.55**
>0	93	17.45
Total	533	100

Table 5: IMPACT OF ETHNO-LINGUISTIC HETEROGENEITY ON HIV STATUS /PREVALENCE

	Dependent Variables			
OLS regression	*HIV STATUS*	*HIV PREVALENCE*	*HIV STATUS*	*HIV PREVALENCE*
ELF	*0.000534****	*0.0576****	*0.000207***	*0.0246****
	(7.83e-05)	*(0.00759)*	*(8.47e-05)*	*(0.00806)*
Religious Fractionalization			*-3.97e-05*	*-0.00292*
			(9.54e-05)	*(0.00922)*
REGIONAL FIXED EFFECTS	*YES*	*YES*	*YES*	*YES*
Urban	*NO*	*NO*	*YES*	*YES*
Socio-Economic Controls	*NO*	*NO*	*YES*	*YES*
Observations	*32667*	*35885*	*32628*	*35845*
R-squared	*0.062*	*0.361*	*0.082*	*0.384*

Notes: The dependent variables are HIV status and HIV prevalence calculated at enumeration area level. All regressions include 34 regional fixed effects and alternatively 135 ethnic fixed effects. Socio-economic controls include religion, wealth, education, age . Even though we show results from OLS regression, all results hold with non linear models.
In all regressions, samples weights are used and robust standard errors clustered at the enumeration area level.
**** p<0.01, ** p<0.05, * p<0.1*

Table 6: Effect of Ethno-linguistic Heterogeneity on own HIV status

Dependent Variable: HIV STATUS								
Probit	*WOMEN*					*MEN*		
	TOTAL (1)	**RESTRICTED (2)**	**TOTAL (3)**	**MARRIED (4)**	**NON MARRIED (5)**	**TOTAL (6)**	**MARRIED (7)**	**NON MARRIED (8)**
ELF	.0029***	0.00558***	0.00355**	0.00630***	0.00483**	0.00350**	0.00118	-0.00151
	(.00076)	(0.00144)	(0.00149)	(0.00184)	(0.00194)	(0.00174)	(0.00190)	(0.00431)
Observations	31854	6384	6379	5947	5943	5923	5918	*2010*
EI	.2423***	0.139	0.271**	0.386**	-0.367	-0.153	-0.0811	-0.246
	(.06972)	(0.122)	(0.134)	(0.150)	(0.285)	(0.204)	(0.242)	(0.369)
Observations	*31854*	*14739*	*7844*	*5185*	*2185*	*6416*	*3454*	*1909*
GELF	.00083	0.00156	0.00338**	0.00465**	-0.00320	-0.00123	-0.00201	0.00345
	(.00067)	(0.00136)	(0.00158)	(0.00195)	(0.00278)	(0.00207)	(0.00260)	(0.00350)
Observations	*31854*	*14739*	*7844*	*5185*	*2185*	*6416*	*3454*	*1909*

Notes: All regressions include 34 regional fixed effects, religious heterogeneity, religion, wealth, education, age and if rural or urban area. Results hold even if we include ethnic fixed effects and other socio-economic characteristics. From Column 2, the Sample is restricted to those areas where the number of people sampled, in the enumeration area is superior to 5, and the proportion of migrants in the enumeration area is less than 30%. Even though we show results from Probit regressions, all results hold with other specifications. Column 1 is based on the un-restricted sample, Column 2 on the restricted while columns 3-5 are further restrictions based on the sample of women, married and non-married, and the last three on that of men, married and non-married. In all regressions, sample weights are used and robust standard errors clustered at the enumeration area level.
**** p<0.01, ** p<0.05, * p<0.1*

Table 7: Effecto of Ethno-Linguistic Heterogeneity on HIV status by Gender

Dependent Variable: HIV Status			
ELF	-0.00239		
	(0.00220)		
ELF*FEMALE	0.00772***		
	(0.00219)		
EI		-0.227	
		(0.180)	
EI*FEMALE		0.564***	
		(0.175)	
GELF			-0.00228
			(0.00200)
GELF*FEMALE			0.00617***
			(0.00206)
FEMALE	0.142*	0.138*	0.172**
	(0.0753)	(0.0800)	(0.0759)
Religeous Fractionalisation	-0.00271	-0.00254	-0.00269
	(0.00410)	(0.00410)	(0.00412)
Migrant	0.285***	0.286***	0.282***
	(0.0653)	(0.0651)	(0.0655)
Married	-0.267***	-0.268***	-0.277***
	(0.0759)	(0.0759)	(0.0758)
Urban	0.0551	0.0779	0.0770
	(0.120)	(0.125)	(0.131)
Age	0.184***	0.183***	0.184***
	(0.0213)	(0.0213)	(0.0213)
Age Squared	-0.00247***	-0.00247***	-0.00247***
	(0.000316)	(0.000314)	(0.000314)
Constant	-5.262***	-5.264***	-5.288***
	(0.501)	(0.512)	(0.499)
Observations	14739	14739	14739

All regressions include 34 regional fixed effects,Dummy for Christian or Muslim, wealth, education. Sample is restricted to those areas where the number of people sampled, in the enumeration area is superior to 5, and the proportion of migrants in the enumeration area is less than 30%. Results are stronger if we relax this restriction. In all regressions, samples weights are used and robust standard errors clustered at the enumeration area level.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: HIV Discordance among couples

		Wife		
		HIV-	HIV+	Total
Husband	HIV-	6,311	**191**	6,502
	HIV+	**174**	197	371
	Total	6,485	388	6,873

Table 9: The Effect of Ethno-Linguistic Heterogeneity on Discordant Couples

Probit	*Dependent Variable: Discordent Couple*					
Sample	She or He POSITIVE		SHE POSITIVE -HE NEGATIVE		HE POSITIVE-SHE NEGATIVE	
	Column I	Column II	Column III	Column IV	Column V	Column VI
ELF	0.00558***	0.00355**	0.00630***	0.00483**	0.00350**	0.00118
	(0.00144)	(0.00149)	(0.00184)	(0.00194)	(0.00174)	(0.00190)
Observations	6384	6379	5947	5943	5923	5918
EI	0.386***	0.210	0.461***	0.319*	0.201	0.00219
	(0.133)	(0.131)	(0.166)	(0.166)	(0.159)	(0.167)
Observations	6641	6636	6180	6176	6170	6165
GELF	0.00397***	0.00230*	0.00453***	0.00295*	0.00240	0.000501
	(0.00139)	(0.00140)	(0.00167)	(0.00173)	(0.00186)	(0.00183)
Observations	6641	6636	6180	6176	6170	6165
Regional Fixed Effects		YES		YES		YES
Urban		YES		YES		YES
Religious Fractionalization		YES		YES		YES
Migrant		YES		YES		YES
Marital Duration		YES		YES		YES
Polygamy		YES		YES		YES
Socio-Economic Controls		YES		YES		YES

Notes: The dependent variable is a binary variable where it takes the value 1 in column I and II if the couple is discordant: either him or her HIV positive . In column III and IV it takes the value 1 whenever she is positive and he is negative. Finanlly in column V and VI it takes value 1 whenever he is positive and she is negative. In each column, it takes the value 0 whenever both are positive or negative. Socio-economic controls include religion, wealth, education, age . By migrant we mean a dummy with value one whenever the respondent is resident for less than 10 years.

In all regressions, samples weights are used and robust standard errors clustered at the enumeration area level.

*** p<0.01, ** p<0.05, * p<0.1