Anonymity and Infidelity: Ethnic Identity, Strategic Cross-Ethnic Sexual Network Formation, and HIV/AIDS in Africa

Roland Pongou

Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, USA

E-mail: Roland_Pongou@brown.edu

Date: March 28, 2009

Abstract: We develop a theory of how community-level ethnic heterogeneity determines the formation of sexual networks, and how this in turn affects the diffusion of HIV/AIDS. Specifically, the model assumes that agents derive utility from sexual relationships, but sexual infidelity is socially prohibited and penalized if detected. It implies a mechanism wherein ethnic heterogeneity provides incentives for optimizing agents to multiply sexual partners across ethnic groups while lowering the probability of infidelity detection, and thus positively affects the spread of HIV/AIDS. We use individual level data from six representative countries of sub-Saharan Africa to test the model. In doing so, we find a direct effect of ethnic heterogeneity on both the number of sexual partners and HIV infection. Robustness checks show that this effect is not driven by a lack of public goods in ethnically diverse communities. Interestingly, ethnic heterogeneity is shown to have no effect on anemia which is largely untreated like HIV/AIDS, but which unlike HIV/AIDS, does not involve socially prohibited human interactions. This finding supports our theory that diseases which involve socially prohibited human interactions spread through hidden and anonymous networks. Our study offers a new and fresh explanation for the high concentration of the AIDS epidemic in sub-Saharan Africa.

Keywords: Ethnic identity; ethnic heterogeneity; anonymity; infidelity; strategic sexual network formation; HIV/AIDS; Africa.

I am indebted to Andrew Foster for guidance and helpful advice. My gratitude also goes to Francis Bloch, Ken Chay, Jakina Debnam, Esther Duflo, Jane Fortson, Jeffrey Greenbaum, Rachel Heath, Blessing Uchenna Mberu, Sriniketh Nagavarapu, Marie Patience Pongou, Sam Preston, Michelle Poulin, Louis Putterman, Paul Schultz, Roberto Serrano, Adam Storeygard, Olumide Taiwo, David Weil, and several participants at the Brown University Department of Economics Micro Lunch and the 2008 Northeast Universities Development Consortium at Boston University for useful discussions and feedback.

----

1 Introduction

We propose a theory of how community-level ethnic heterogeneity determines the formation of sexual networks, and how these networks affect the diffusion of HIV/AIDS. Ethnic groups in a community constitute pre-defined communication networks wherein information flows more easily within groups than across. Limited communication across groups thus fosters cross-group anonymity, providing incentives for optimizing individuals to choose their sexual partners across groups rather than within, in order to minimize the probability of being caught. We show that this optimizing behavior implies a mechanism wherein ethnic heterogeneity encourages infidelity and positively affects the spread of HIV/AIDS, a new hypothesis that we also test empirically.

There is a growing body of research on the economic causes and consequences of the AIDS epidemic, and on how to reverse its progression (see e.g, Canning 2006, Oster 2007, Fortson 2008a, 2008b, Bloom and Mahal 1997, Hogan and Salomon 2005). Most of this literature focuses on sub-Saharan Africa, where HIV/AIDS has slowed a century of substantial progress in the fight against infectious diseases, and has caused a sharp decline in life expectancy and Disability-Adjusted Life Years (UNAIDS 2008, Lopez et al 2006). This region accounts for 67 percent of the 33 million people living with HIV/AIDS today, and for 72 percent of AIDS deaths, but the epidemic is spreading rapidly in other parts of the world such as South and East Asia (UNAIDS 2008). Beyond its immediate impact on life quality and longevity, the AIDS epidemic has carried devastating social and economic impacts. By primarily striking prime-age adults, it has caused or exacerbated poverty in most households, and has increased the burden of orphanhood and widowhood (Matshalaga 2002, Monash and Boerma 2004, Heuveline 2004, UNAIDS 2008). There is also evidence that the epidemic has negatively affected human capital investment and economic growth (Fortson 2008a, Bonnel 2000). These devastating effects have made the study of the social and economic determinants of HIV/AIDS crucial for the design of policies to help curtail the growth of the epidemic. Documented determinants include the high transmission rates of the AIDS virus, poverty[1], spatial mobility, and limited behavioral change (Oster 2005, 2007, Arnafi 1993, Lurie et al 2003, Thornton 2006, Orubuloye and Oguntimehin 1999).

Little is known about the role of community attributes in fostering multiple and concurrent partnerships, the most important factor in the rapid spread of HIV/AIDS (Epstein 2007, Halperin and Epstein

[1] Note however that several studies have found higher HIV prevalence among the rich compared to the poor (see e.g., Mishra et al. 2007, Fortson 2008b)

2004, Royce et al. 1997, Morris and Kretzschmar 1997). By developing and testing a new model that relates ethnic diversity to sexual behavior, the current study increases our knowledge of the rationale and mechanism that underlie the formation of sexual networks and the diffusion of HIV/AIDS in multi-cultural environments. First, our model predicts that, within countries, communities that feature greater ethnic diversity are more likely to foster and sustain concurrent partnerships across groups, and are thus prone to generalized epidemics. Second, this study enhances our understanding of both the dynamics and high concentration of HIV/AIDS in sub-Saharan Africa, the world's most ethnically diverse region.[2] In so doing, it also informs public policies aimed at curbing the spread of this devastating epidemic in the region. In particular, our findings speak loudly to the role of complete social and ethnic integration in minimizing the effects of HIV/AIDS.

The primitives of the theoretical model are as follows. In a community, reside an equal number of men and women partitioned into different ethnic groups, each of equal size. Agents derive utility from sexual intercourse with agents of the opposite sex, with utility increasing in the number of sexual partners. However, an agent with multiple partners is committing an act of infidelity and is punished if caught. Because information flows more easily within groups than across, the probability of infidelity detection is greater if the cheated partners belong to the same group. We show that this underlying communication network structure causes optimizing agents to choose their sexual partners across ethnic groups rather than within, in order to decrease the probability of infidelity detection and thus its associated cost. As a corollary, it is shown that increasing the number of ethnic groups in the community results in each agent adding more sexual links, which in turn increases the prevalence of HIV/AIDS. These findings also introduce ethnic diversity as an underlying factor in the spatial variation of HIV/AIDS prevalence within countries.[3] We test empirically these theoretical predictions of the model.

The analysis is based on six sub-Saharan African countries: Burkina Faso, Cameroon, Ethiopia, Ghana, Kenya and Malawi. These countries represent to a large extent the geographical and cultural diversity of sub-Saharan Africa, and provide a representative mapping of the AIDS situation in the region. They are characterized by marked ethnic and linguistic diversity (Alesina et al. 2003), which is consistent with anthropological findings that most African societies are organized around distinct ethnic groups and enclaves. The commonality of language and culture facilitates communication among people of the same group. Cultural

[2]Note however that we do not examine the relationship between ethnic heterogeneity and HIV infection at the country level, but we do this at the micro-level, thus controlling for country level factors such as HIV prevention policies, and a host of other factors operating at the macro level.

[3]Cross-country differences have been severally documented, and explained by such factors as HIV prevention policies and male circumcision (see, e.g., Asamoah et al. 2004, Potts 2000, Potts et al. 2008). But our analysis controls for country-level factors and cultural sources of variation in factors like male circumcision or the acceptability of multiple sexual partnerships.

boundaries between ethnic groups however cause information transmission to be less efficient across groups. However, despite a noted persistence of ethnic identities in most societies, Anthropologist Fredrik Barth (1969) argues that these identities are not bounded and cannot be viewed as cultural isolates. To this effect, he writes:

> "[...] categorical ethnic distinctions do not depend on an absence of mobility, contact and information, but do entail social processes of exclusion and incorporation whereby discrete categories are maintained despite changing participation and membership in the course of individual life histories." (P. 9)

In post-colonial Africa, interethnic relationships within countries are further facilitated by the heritage of the colonial language, but ethnic identification has persisted in most countries (Posner 2003, Eifert, Miguel and Posner 2007), resulting in moderately fragmented societies. As implied by our theoretical model, we expect this sort of cultural diversity to enable people to engage with multiple sexual partners across groups, which sets propitious grounds for the spread of HIV/AIDS.

This hypothesis is tested using pooled nationally representative Demographic and Health Surveys (DHS) supplemented by data on population density from censuses. The DHS constitute a unique source of information on individuals' sexual behavior, HIV status, and ethnic affiliation, and have been used in recent studies to examine the economic causes and consequences of the AIDS epidemic in sub-Saharan Africa (Oster 2005, 2007; Fortson 2008a, 2008b). Information on ethnic affiliation enables us to compute a newly defined index of ethnic heterogeneity at the cluster or census zone level, and at the region/urban/rural level. We find a positive effect of ethnic heterogeneity on the number of sexual partners and on HIV infection. This effect is robust to the inclusion of a range of demographic, socioeconomic and cultural variables. Despite the fact that the cluster-level and the region/urban/rural-level indexes of ethnic heterogeneity are only moderately correlated, the effect of the first index remains unchanged while that of the second index disappears once both variables are controlled for, essentially indicating that the space of strategic sexual networking is best represented by clusters. The effect of ethnic heterogeneity is also robust to an alternative measure wherein the index of ethnic diversity is decomposed into two indexes representing respectively the distinct contribution of individuals' "own" and "opposing" ethnic groups to the overall index.[4]

We conduct several robustness checks that allow to rule out alternative explanations for the effect of ethnic heterogeneity on sexual behavior and HIV infection. First, we show that this effect is robust to separate

[4]This decomposition is reminiscent of the "Us/They" dichotomization of Duclos, Esteban and Ray (2004), but our index is different from theirs. This decomposition also allows to account for potential group and social preferences in sexual behavior.

estimation by urban and rural areas, establishing that it does not simply reflect urban-rural differences in the outcomes. Second, this effect persists and is similar when estimated separately for migrants and non-migrants. Third, this effect is not driven by a lack of public goods in more ethnically diverse communities.[5] Finally, we show that ethnic heterogeneity has no effect on anemia. Anemia is highly prevalent in most African countries, is largely untreated like HIV/AIDS; but unlike HIV/AIDS, it does not involve socially prohibited human interactions. A positive effect of ethnic heterogeneity on such a condition would largely call into question our interpretation of the positive effect of ethnic heterogeneity on sexual promiscuity and HIV infection. That we find no such effect lends credence to our theory and our argument that diseases which involve socially prohibited human interactions spread through hidden and anonymous networks.

Aside from contributing to the literature on the social and economic causes of HIV/AIDS, our study relates to other literatures. First, our theoretical model is an extension of the fidelity model in Pongou (2009a) to a multi-ethnic environment, and it shares several features of models used in recent studies of network formation (Aumann and Myerson 1988, Jackson and Wolinsky 1996, Bala and Goyal 2000).[6] These models generally postulate that economic agents derive utility from direct and/or indirect links with other agents, while incurring a maintenance cost of direct links. Equilibrium networks are then predicted based on various concepts of stability. In our fidelity model, agents derive satisfaction from direct links with opposite sex partners, and incur a penalty cost if their infidelity is detected. This leads to a single-peaked utility function, a property which realistically implies that in stable matchings, no agent can have beyond a certain number of partners. Our model also departs from existing models by the consideration of an underlying and exogenous communication network - ethnic enclaves - which by its nature affects the strategic behavior of agents when forming links. This consideration also naturally relates our study to the literature on how identity and socio-cultural distance affect social interactions as well as economic decisions such as schooling or childbearing (Akerlof 1997, Akerlof and Kranton 2000). Most of these studies show that interactions primarily take place among socio-culturally similar individuals. But we elucidate a mechanism wherein cultural distance gives rise to a sort of interaction - sexual promiscuity - which is socially prohibited, and which involves as a result culturally distant individuals, therefore demonstrating how ethnic diversity fuels cross-group anonymity, and is propitious to the spread of HIV/AIDS. Our results are also consistent with experimental studies on the effect of anonymous environments on the reporting of sensitive behaviors.[7]

[5] A recent literature shows that public goods are underprovided in more ethnically diverse communities because of a lack of trust and social capital (Easterly and Levine 1997; Alesina, Baqir and Easterly 1999, Miguel and Gugerty 2005; also see Costa and Kahn 2003 and the references therein).

[6] Also see Jackson (2003) and the references therein.

[7] It has been shown for instance that audio computer-assisted self-interviewing produces more accurate reporting of sexual activity and other sensitive behaviors like drug injections than face-to-face interviews (Mensch et al. 2003, Jaya et al. 2008, Des Jarlais et al. 1999).

Finally, given the documented effect of ethnic diversity on the spread of HIV/AIDS in our analysis, and given the detrimental impact of the AIDS epidemic on human capital investment (Fortson 2008b), our study can also be seen as contributing to the recent literature on how ethnic heterogeneity affects economic development (Easterly and Levine 1997; Alesina, Baqir and Easterly 1999, Cinyabuguma and Putterman 2006, Horowitz 1985, Costa and Kahn 2003). But we demonstrate a different mechanism through which ethnic diversity affects outcomes. By postulating and testing a new hypothesis relating cultural diversity to HIV/AIDS, we also offer a new and fresh explanation for the high concentration of this epidemic in sub-Saharan Africa.

The remainder of this paper proceeds as follows. In section 2, we devise the fidelity model and derive results for the effect of ethnic diversity on sexual behavior and HIV/AIDS. Section 3 describes the data we use to test the predictions of the model. This test is conducted in section 4. Sensitivity analyses are conducted in section 5, and robustness checks aimed at ruling out alternative hypotheses are considered in section 6. Section 7 examines the effect of ethnic heterogeneity on anemia. Finally, section 8 discusses implications of the results, an extension of the fidelity, and concludes.

2 The fidelity model

2.1 Individual decision-making

We model the individual decision-making of sexual partners choice in a community made up of e ethnic groups G_1, ..., G_e, each of equal size. Each group has a distinct language and culture, which makes social interactions easier and more intense within groups than across. We assume that each individual derives utility from sexual relationships with opposite sex partners. However, sexual infidelity and promiscuity are socially prohibited, and the cheater incurs a penalty cost if he/she is detected.

Assuming that an individual i has s sexual partners, and that c is the cost incurred by i if one of his/her sexual partners comes to discover his/her relationship with another partner, agent i's utility function is defined as:

$$u(s) = v(s) - t(s)c$$

where $v(s)$ is concave and is increasing in s, and $t(s)$ the number of detected sexual liaisons of i.[8]

[8] Note that $t(s) = 0$ if $s = 0$ or $s = 1$ because if an agent has no or one sexual partner, she is obviously not unfaithful to anybody; $t(s)$ will be more clearly defined in the sequel.

Let j and k be two sexual partners of agent i, and $D(j:(i,k))$ the event that j detects the liaison between i and k. We assume that $D(j:(i,k))$ and $D(k:(i,j))$ occur with the same probability, which is defined as:

$$\Pr ob(D(j:(i,k))) = \Pr ob(D(k:(i,j))) = \begin{cases} p \ if \ \exists h \in \{1,..e\} : j \in G_h \ and \ k \in G_h \\ q \ if \ not \\ with \ q < p. \end{cases}$$

That is, $D(j:(i,k))$ and $D(k:(i,j))$ occurs with probability p if j and k belong to the same ethnic group, and with probability q if the two individuals belong to two different ethnic groups. Assumption $q < p$ means that a cheater's infidelity is more likely to be detected by the cheated agents if the latter belong to the same group than if they belong to two different groups. This assumption stems from the fact that linguistic and cultural barriers between groups make the circulation of information easier and less costly within groups than across.

Individual i can choose sexual partners across groups, which implies that his/her decision function can be expressed as:

$$f : \mathbb{R}_+ \to \mathbb{R}_+^e$$
$$s \to (s_1, ..., s_e) \ with \ \sum s_h = s$$

where f is a mapping from the set of positive real numbers $\mathbb{R}_+$ to the $e-$dimensional cartesian product of $\mathbb{R}_+$ $\mathbb{R}_+^e$.[9] f associates with a number of sexual partners s of agent i its allocation $(s_1, ..., s_e)$ across the e ethnic groups that make up the community; s_h is the number of sexual partners chosen by agent i in ethnic group G_h, with $1 \leq h \leq e$. It is obvious that sexual partners across groups add up to s.

How should agent i optimally allocate her sexual partners across ethnic groups? Assuming that $(s_1, ..., s_e)$ is such an allocation, a partner j of agent i belonging to ethnic group G_h is expected to detect $p(s_h - 1) + q(s - s_h)$ liaisons of i with other partners. The s_h sexual partners of agent i that belong to ethnic group G_h are thus expected to detect $s_h[p(s_h - 1) + q(s - s_h)]$ liaisons of i with other partners. Summing up across all partners and ethnic groups yields $\sum_{h=1}^{e} s_h[p(s_h - 1) + q(s - s_h)]$, which is the total expected number of liaisons that the s sexual partners of agent i will detect. These detections impose a cost of $\sum_{h=1}^{e} s_h[p(s_h - 1) + q(s - s_h)]c$ on i, so that i's expected utility function is:

$$Eu(s, s_1, ..., s_e) = v(s) - \sum_{h=1}^{e} s_h[p(s_h - 1) + q(s - s_h)]c.$$

Assuming that i is an expected utility maximizer, her decision-making problem is thus:

$$\max_{(s, s_1, ..., s_e)} Eu(s, s_1, ..., s_e) = v(s) - \sum_{h=1}^{e} s_h[p(s_h - 1) + q(s - s_h)]c \quad subject \ to \ \sum s_h = s$$

[9]Strictly speaking, the analysis should be conducted in the discrete space N, the set of natural numbers; we choose to do this in the continuous space R_+ to simplify the exposition of our results, but all the results hold in the discrete space N as well.

Let s^* be the optimal number of sexual partners of i, and $(s_1^*, ..., s_e^*)$ the optimal allocation of these partners across the e ethnic groups. If we assume p, q and c to be strictly positive, the following proposition states that agent i will equally distribute her sexual partners across groups.

Proposition 1 *:* $s_1^* = s_2^* = ... = s_e^* = \frac{s^*}{e}$.

Proof : The proof is straightforward and available upon request.

We would like to stress again that we choose to work in the continuous space $\mathbb{R}_+$ rather than the discrete space $\mathbb{N}$ to facilitate the exposition of our results. In the discrete space $\mathbb{N}$, Proposition 1 would be: $\forall h$, $l \in \{1, ..., e\}$, $s_h^* = s_l^*$ or $s_h^* = s_l^* \pm 1$.[10] This would imply, for instance, that if there are three ethnic groups, and the optimal number of sexual partners for an agent is 2, then his/her optimal allocation will be (1,1,0), (1,0,1) or (0,1,1), and not $(\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$ as Proposition 1 actually implies. However, all the results that follow also hold in the discrete space $\mathbb{N}$.

2.2 Comparative Statics

We now want to assess the relationship between the number of ethnic groups e and the optimal number of sexual partners s^*. For simplicity, we can assume that $v(s)$ is linear in s, and we pose:

$$v(s) = s.$$

At the optimal allocation of partners (see Proposition 1), this linearity assumption implies the following expression of the expected utility function of agent i:

$$\begin{aligned} Eu(s) &= s - \sum_{h=1}^{e} \tfrac{s}{e}[p(\tfrac{s}{e} - 1) + q(s - \tfrac{s}{e})]c \\ &= s - [(\tfrac{p-q}{e} + q)s^2 - ps]c \end{aligned}$$

To find the optimal number of sexual partners s^*, first take the derivative of $Eu(s)$ with respect to s and equate it to zero:

$$1 - [2(\tfrac{p-q}{e} + q)s - p]c = 0$$

Solving for s^* yields:

$$s^* = \tfrac{1+pc}{2c} \tfrac{e}{qe+p-q}.$$

Note that the function $Eu(s)$ is single-peaked, increasing in the interval $[0, s^*]$ and decreasing in the interval $[s^*, +\infty)$. The following Proposition states that the optimal number of sexual partners increases with the number of ethnic groups.

Proposition 2 *:* $\frac{\partial s^*}{\partial e} > 0$.

[10] The proof is available upon request.

Proof : $\frac{\partial s^*}{\partial e} = \frac{1+pc}{2c} \frac{p-q}{(qe+p-q)^2} > 0$ because $p - q > 0$.□

The partial derivative of s^* with respect to e implies that if the level of ethnic integration in a community is sufficiently high that $p - q$ is very small, then the marginal effect of ethnic heterogeneity on the number of sexual partners will be close to zero, all other factors being held constant. Accordingly, multiple sexual partnerships in an ethnically divided community are partly driven by cultural barriers that prevent communication across groups.

The following proposition also holds.

Proposition 3 *: 1)* $\frac{\partial s^*}{\partial c} < 0$

2) $\frac{\partial^2 s^*}{\partial e \partial c} < 0$

3) Assume $0 < p < 1$ *and pose* $p = q + a$ *for any* p*,* a *being a constant. Then* $\frac{\partial s^*}{\partial p} < 0$

4) Assume $0 < p < 1$ *and pose* $p = q + a$ *for any* p*,* a *being a constant. Then* $\frac{\partial^2 s^*}{\partial e \partial p} < 0$.

Proof : 1) $\frac{\partial s^*}{\partial c} = \frac{-2}{4c^2} \frac{e}{qe+p-q} < 0$.

2) $\frac{\partial^2 s^*}{\partial e \partial c} = \frac{-2}{4c^2} \frac{p-q}{(qe+p-q)^2} < 0$.

3) $\frac{\partial s^*}{\partial p} = \frac{e}{2c} \frac{(1-e)ca-e}{[(1-e)a+pe]^2} < 0$

4) $\frac{\partial s^*}{\partial e \partial p} = \frac{a}{2c} \frac{(1-e)ca-e}{[(1-e)a+pe]^2} < 0$.□

Propositions 2 and 3 imply that although the number of sexual partners increases with ethnic heterogeneity, it increases less for individuals on whom a higher cost of infidelity is imposed or in areas in which the probability of detection is higher. If any concave, continuous and twice differentiable utility function $v(s)$ is assumed, Proposition 2, Propositions 3-1 and 3-3 still hold, but Propositions 3-2 and 3-4 hold only if the arguments of s^* lie in certain intervals.

3 Data

3.1 Data sources

We use individual and household level data from Demographic and Health Surveys conducted in recent years by Measure DHS (http://www.measuredhs.com/) in Burkina Faso (2003), Cameroon (2004), Ethiopia (2005), Ghana (2003), Kenya (2003) and Malawi (2004). These surveys are nationally representative, and are comparable across countries and over time for most questions. In recent years, the DHS have collected individual level HIV data through direct HIV testing. These data constitute a unique and reliable source of information on HIV infection based on nationally representative samples, and are especially pertinent in sub-Saharan Africa where the AIDS epidemic has had the most devastating impact. They have proven crucial to the analysis of health and sexual behavior in Africa and other developing regions of the world (see, e.g., Oster 2007, Fortson 2008a, 2008b).

Data collection follows similar standards in all countries. Each country is partitioned into a certain number of regions representing the largest administrative units at the sub-national level, and each region is partitioned into urban and rural areas.[11] In each region/urban/rural stratum, a two-stage probabilistic sampling technique is used to select clusters or census enumeration zones in the first stage[12], and households in the second. In each household, information is collected on basic demographic and socioeconomic characteristics, and on health outcomes for all or some household members. For the purpose of this study, the DHS are supplemented by data on population density from censuses provided by City Population (http://www.citypopulation.de/).

3.2 Data descriptions

3.2.1 Socioeconomics

We present the summary statistics of most variables used in the analysis in Table 1. These variables are gender, age, marital status, religion, education, occupation, urban/rural place of residence, migration status and population density. Males represent 32 percent of the sample.[13] Average age is around 29 years old, however age ranges from 15 to 49 for women and from 15 to 59 for men. Most people are married (57 percent), or are Christian (67 percent). The average level of educational attainment is about 5 years of schooling. We note a relatively high level of unemployment (31 percent), and that most people (63 percent) work in the traditional sector which comprises agriculture, manual work and sale.

Rural dwellers (69 percent) make the largest share of the population. The proportion of non-migrants (44 percent) is also relatively high. Population density is provided for each region and each major city within each country. We note that an average person lives in a region whose density is 632 people per km^2. The variables described above also serve as controls in the estimation of the effects of ethnic heterogeneity on sexual behavior and HIV infection. These latter variables are described below.

[11] Great cities (e.g., Nairobi, Yaounde, Douala, etc.) are however sampled separately from the largest sub-national regions that include them, and are obviously only urban. There are 58 regions (including great cities) and 100 region/urban/rural strata for the six countries considered in our analysis.

[12] There are 2740 census zones in the pooled sample.

[13] This is because the DHS focuses more on women, thus more women than men are selected to be interviewed. But the sample of men is representative.

3.2.2 Dependent variables: Sexual behavior and HIV infection

We analyze sexual behavior and HIV infection. Information on these variables is collected for a representative sample of men and women. Sexual behavior is measured as the number of reported sexual partners (excluding or including wife or husband) in the 12 months preceding the survey. HIV status is determined through direct tests administered by interviewers to selected household members, and is either positive if an individual is infected with the AIDS virus, or negative otherwise. In the light of numerous evidences that respondents often underreport or overreport their sexual activity in household survey (Mensch et al. 2003, Des Jarlais et al 1999, Jaya et al 2008, Hewett et al. 2004)), HIV status stands as a more reliable and objective measure of sexual behavior in the absence of a curative treatment for AIDS. The magnitude of underreporting may vary across several demographic and socioeconomic characteristics[14], as well as across levels of ethnic heterogeneity; therefore, a problem arises with using the reported number of sexual partners as the dependent variable. It is also possible that the underreporting bias increases with increasing sexual activity, implying for instance that estimates of the effect of ethnic heterogeneity on number of reported sexual partners are very likely to be biased downward. Analyzing the effect of ethnic heterogeneity on HIV status, however, is expected to yield unbiased or less biased estimates.

Table 1 shows that the average number of sexual partners per person is 0.78 when husbands and wives are included, and 0.21 when they are excluded. The proportion of individuals infected with the AIDS virus is 4.4% for the pooled sample.[15]

3.2.3 Independent variable: Ethnic heterogeneity

The number of ethnic groups reported by the DHS is 11 for Burkina Faso, 50 for Cameroon, 69 for Ethiopia, 9 for Ghana, 15 for Kenya, and 8 for Malawi, totalling 162 ethnic groups for the pooled sample. Nevertheless, some ethnic groups are aggregated in the data as a result of linguistic and cultural proximity. In Cameroon for instance, this first level aggregation has yielded 50 ethnic groupings out of the over 200 ethnic groups that exist in the country.

The theoretical model assumes equally sized ethnic groups for simplicity, although ethnic groups are rarely uniform in size. Accordingly, most of the literature accounts for this uneven distribution by drawing upon the so called Herfindahl index of ethnolinguistic fractionalization (hereafter ELF) (see e.g., Easterly and

[14] For instance, this magnitude is generally higher for females compared to males (see, e.g., Zaba et al. (2004)).

[15] The average number of reported sexual partners varies across countries, ranging for instance from 0.08 in Ethiopia to 0.52 in Cameroon when husbands and wives are excluded. The prevalence of HIV infection varies across countries as well, ranging from 2% in Burkina Faso, Ethiopia and Ghana, to 5% in Cameroon, 7% in Kenya, and 12% in Malawi.

Levine 1997). This index measures the probability that two randomly selected individuals from a community belong to two different ethnic groups, and is computed as:

$$ELF = 1 - \sum_{r=1,\ldots,e} \frac{N_r^2}{N^2}$$

where e is the number of ethnic groups in the community, N_r is the size of ethnic group r, and N the size of the community.

This expression of the ethnolinguistic fractionalization index implicitly assumes that individuals are selected in two successive random draws with replacement. The same individual can thus be drawn twice, which is not an appropriate assumption. To illustrate, consider a community of two people who belong to different ethnic groups. By construction, the probability that the two individuals are from two different ethnic groups is 1, whereas the Herfindahl index would incorrectly indicate a probability of 0.5. Accordingly, we propose a measure of ethnic heterogeneity based on the assumption that individuals are selected in successive random draws without replacement. The ethnic heterogeneity index of order k ($k \geq 2$) is the probability that k $different$ randomly selected individuals from a community belong to k $different$ ethnic groups. This index is expressed as:

$$EHI(k) = 1 - \sum_{r=1,\ldots,e} \frac{N_r(N_r-1)\ldots(N_r-k+1)}{N(N-1)\ldots(N-k+1)}$$

For our analysis, we choose $k = 2$, thereby measuring ethnic heterogeneity (hereafter EHI) as the probability that 2 $different$ randomly selected individuals from a community belong to 2 $different$ ethnic groups. This index is calculated for each cluster or census zone, and for each region/urban/rural stratum, yielding measures that vary a great deal within each country.

For our analysis, we compute the EHI for each cluster and for each region/urban/rural stratum.[16] Table 1 shows that in the pooled sample, an average person resides in a community where the cluster-level EHI is 0.32, and where the region/urban/rural stratum-level EHI is 0.55. Lastly, although the EHI and ELF are conceptually different, they are highly correlated in the DHS (the coefficient of correlation is 0.997). The EHI and the number of ethnic groups at the census zone level are also highly correlated (0.887) in the DHS, which shows that in addition to accounting for the size of each ethnic group, the EHI also accounts for the number of ethnic groups.

[16] However, results show that clusters best represent the space of strategic sexual interactions than the much larger region/urban/rural strata (see section 4.2).

4 Ethnic heterogeneity, number of partners, and HIV

In this section, we document the effect of ethnic heterogeneity on both the number of reported sexual partners and HIV infection. Figure 1- Panels A and B describe a positive relationship between ethnic heterogeneity and the reported number of partners.[17] Similarly, in Figure 1- Panel C, we find a positive relationship between ethnic heterogeneity and HIV prevalence. A similar analysis is conducted for each country and presented in Figure 2-A for non-marital partnerships and Figure 2-B for HIV infection. Residents of more ethnically diverse communities have more sexual partners and are thus more likely to be infected with the AIDS virus.

Moreover, the relationship between ethnic heterogeneity and HIV infection tends to be stronger than that between ethnic heterogeneity and number of sexual partners (Figures 2-A and 2-B)[18]; this is consistent with the idea that sexual activity may be underreported in more ethnically heterogeneous communities. This reinforces the argument that HIV infection is a more objective and better measure of sexual behavior than the reported number of sexual partners. The estimated effect of ethnic heterogeneity on HIV infection will not be biased, whereas the effect of this variable on the number of reported sexual partners will be biased downward.

4.1 OLS and fixed effect estimates of the effect of EHI

The effect of ethnic heterogeneity on number of sexual partners and HIV infection is estimated using an OLS regression model of the form:

$$Y_{ij} = \theta EHI_j + X_{ij}\pi + \in_{ij} \qquad (1)$$

where Y_{ij} is either the reported number of sexual partners (including husbands and wives) in the past 12 months preceding the survey for an individual i residing in a community j, or a binary indicator for whether i is infected with HIV[19]; EHI_j is the ethnic heterogeneity index for community j; and X_{ij} is a vector of demographic and socioeconomic variables. θ is the average change in the outcome variable caused by a move from complete ethnic homogeneity (i.e. $EHI_j = 0$) to complete ethnic heterogeneity ($EHI_j = 1$). $\in_{ij}$ is an unobserved disturbance term assumed to be uncorrelated with EHI_j given X_{ij}.

Results for the estimation of (1) are presented in Table 2-Panel A for the number of reported sexual partners and Table 2-Panel B for HIV infection. In Column (1) of each panel, EHI is included as the

[17] Individuals are grouped by EHI quintile and the average number of partners is computed in each group. Panel A includes husbands and wives among sexual partners while Panel B considers only non-marital partners.

[18] This is more visible in Malawi.

[19] We use the OLS model to estimate the effect of EHI on HIV status which is a binary variable for ease in the interpretation of results. In results not shown, we conduct this analysis using the probit model, and all our conclusions are unaltered.

only predictor. We note that a change from complete ethnic homogeneity to complete ethnic heterogeneity increases the average number of sexual partners by 0.20 and the probability of HIV infection by 6.8 percentage points. Columns (2)-(5) of Panel A show that the effect of ethnic heterogeneity on number of sexual partners is robust to the inclusion of a range of demographic and socioeconomic factors.[20] In Column (2) of Panel B, we additionally control for community sexual behavior measured as the prevalence of multiple sexual partnerships in a community.[21] This changes little to the effect of EHI on HIV infection. This effect is also robust to the addition of other demographic and socioeconomic factors in Columns (3)-(5).

Overall, these results show that the effect of EHI on sexual behavior is not driven by the socioeconomic composition of ethnically homogeneous versus ethnically heterogeneous communities. Specifically, given the documented role of rural-urban migration and other types of spatial mobility in increasing risky behavior and HIV infection (Lurie et al. 2003, Mberu 2008), it was important to show that the effect of EHI is not confounded by factors such as urban/rural place of residence or migration.[22] Also of interest is the fact that controlling for population density does not affect our estimates. A high population density is likely to lower the search cost of a sexual partner because residents are much closer to each other, which makes it easier to have more partners. That our estimates are robust to the inclusion of this variable simply suggests that its correlation with EHI is weak.

Finally, we control for ethnic affiliation (or ethnic group fixed effect) in Column (6) of each panel.[23] The inclusion of this variable is justified by the fact that HIV-related cultural practices such as male circumcision vary across ethnic groups, and ethnic enclaves generally have unequal exposure to ethnic heterogeneity as a result of historical and political roots.[24] Also, because each country has a distinct set of ethnic groups, controlling for ethnic affiliation automatically accounts for country level factors such as HIV/AIDS prevention policies, the social acceptability of concurrent partnerships, national economic variables, and a host of other macro level factors that may affect the spread of HIV.[25] Controlling for ethnic group fixed effect halves the effect of EHI on each outcome, but this effect remains economically and statistically significant.

Note however that if, a priori, there is no reason why sexual behavior should differ across ethnic groups,

[20] These factors include gender, age, education, religion, occupation, marital status, urban/rural place of residence, migration status, and population density.

[21] This prevalence is measured by the proportion of individuals who reported having had more than one sexual partner in the past 12 months. Under the assumption that sexual networking takes place within the community, this prevalence is supposed to capture the effect of an individual's sexual network. Thus, an individual's HIV status reflects her sexual behavior after the effect of her sexual network is netted out.

[22] In sensitivity analysis in section 4-3, we estimate (1) separately for urban areas and rural areas, and for migrants and non-migrants; the effect of ethnic heterogeneity on each outcome is still economically and statistically significant.

[23] This is equivalent to controlling for 161 ethnic group binary variables.

[24] In most African countries, ethnic enclaves that settled and developed near capital cities or touristic sites are more likely to be exposed to a greater degree of ethnic diversity because of the high attractiveness of these areas.

[25] See, e.g., Asamoah et al. 2004, Potts 2000, Lieberman 2007, Potts et al. 2008.

ceteris paribus, a decrease in the effect of EHI after controlling for ethnic group fixed effect may reflect that members of ethnic enclaves that have greater exposure to ethnic heterogeneity due to their historical settlement are more likely to engage in multiple sexual partnerships, and are consequently more vulnerable to HIV infection.[26] If so, the estimates of the effect of EHI should be those obtained in Column (5) of each panel. However, in any case, ethnic heterogeneity has been shown to provide incentives for people to engage in multiple sexual partnerships, which in turn increases the likelihood of HIV infection.

4.2 The space of strategic sexual networking

We estimate the effects of the region/urban/rural-level EHI on number of sexual partners and on HIV infection. Considering the effect of ethnic diversity at this more aggregate level addresses the issue that clusters might be too small to appropriately represent the space in which sexual interactions take place.[27] The results are presented in Table 3. Panel A shows the effects of EHI on number of sexual partners, and Panel B shows the effects of EHI on HIV infection. For clarity and to facilitate comparison, we report the effects of cluster-level EHI in Columns (1) and (2) of each panel. Columns (3) and (4) report the effects of the region/urban/rural-level EHI. These effects are economically and statistically significant. When controls are included, the effect of the region/urban/rural-level EHI on number of sexual partners is slightly greater than that of the cluster-level EHI, but the effect of the former index on HIV infection is slightly smaller than that of the latter index.

The cluster-level EHI and the region/urban/rural-level EHI are only moderately correlated (the correlation coefficient is 0.44), which makes it possible to control for both variables in Column (5) of each panel. We note that the effect of the cluster-level EHI on either outcome persists and is almost unaffected, while that of the region/urban/rural-level EHI loses statistical power and even vanishes for the case of HIV infection. These results are important, and essentially indicate that the space of strategic sexual networking and HIV transmission is best captured by clusters. As *repeated* sexual interactions are often necessary for HIV transmission[28], the results also reveal that sexual interactions that may occur outside of clusters are not as sustained or intense as those which occur within clusters.[29] In the subsequent analyses, only the cluster-level

[26] This interpretation would be consistent for instance with the Luo being disproportionately affected by HIV due to their proximity to Lake Victoria, a site of high attractiveness.

[27] This concern would however not apply to rural areas where clusters are bigger and sometimes cover entire villages.

[28] This is because the male-to-female and the female-to-male transmission rates per coital act are strictly lower than 1 (see, e.g., Gray et al. (2001)), implying that the probability that an infected person transmits the virus to an uninfected partner converges to 1 only over repeated sexual interactions: in fact, let ℓ, $0 < \ell < 1$, be the probability that an infected person transmits the virus to an uninfected partner during a coital act. Assuming independence of transmission over time, the probability of transmission after k coital acts is $1 - (1 - \ell)^k$, which converges to 1 as k tends to $+\infty$. Note that k can be taken to measure how a sexual relationship is sustained over time.

[29] If strategic sexual relationships outside of clusters were as sustained as those occurring within clusters, the effect of the region/urban/rural EHI on HIV infection would be stronger.

EHI is considered.

4.3 Decomposing the effect of ethnic heterogeneity

The estimation of the effect of ethnic heterogeneity on both the number of reported sexual partners and HIV infection in section 4-1 assumes that the summands of EHI have the same effect on the outcomes. This assumption might be strong, especially in the presence of ethnic preferences for sexual partners, or if the cost of infidelity is inflicted unevenly for partners of the same ethnic group and those from different groups. Although without loss of generality, these potential heterogeneities are not accounted for in the theoretical model, they might have significant empirical effects. As a result, we decompose the effect of ethnic diversity into the effects of "own" and "opposing" ethnic groups' contribution to EHI.

Let i be an individual who belongs to ethnic group $r(i)$ and resides in a community j. Let EHI_j denote the level of ethnic heterogeneity in community j. Recall that EHI_j is defined as:

$$EHI_j = 1 - \sum_{r=1,\ldots,e} \frac{N_r(N_r-1)}{N(N-1)}$$

where N_r is the population size of ethnic group r in j, and N the total population size of community j.

EHI_j can be decomposed as:

$$EHI_j = 1 - [\frac{N_{r(i)}(N_{r(i)}-1)}{N(N-1)} + \sum_{r\neq r(i)} \frac{N_r(N_r-1)}{N(N-1)}]$$

If we define $EHI\_own_{ij} = \frac{N_{r(i)}(N_{r(i)}-1)}{N(N-1)}$ and $EHI\_other_{ij} = \sum_{r\neq r(i)} \frac{N_r(N_r-1)}{N(N-1)}$, then:

$$EHI_j = 1 - [EHI\_own_{ij} + EHI\_other_{ij}].$$

From the perspective of individual i, EHI_own_{ij} and EHI_other_{ij} represent respectively what we call "own" and "opposing" ethnic groups' contribution to EHI in community j. Note however that EHI_own_{ij} and EHI_other_{ij} are perfectly negatively correlated with EHI_j by construction.

For our decomposition exercise, we estimate the following OLS regression:

$$Y_{ij} = \theta_1 EHI\_own_{ij} + \theta_2 EHI\_other_{ij} + X_{ij}\pi + \in_{ij} \quad (2)$$

Because EHI_own_{ij} and EHI_other_{ij} are negatively correlated with EHI_j, our parameters of interest θ_1 and θ_2 are expected to have the opposite sign of the coefficient θ on EHI_j in equation (1).[30] In the absence of the sort of group level heterogeneities previously mentioned, θ_1, θ_2 and θ should be equal in absolute value.

The results of our investigation are reported in Table 4-Panel A for the number of reported sexual partners and Table 4-Panel B for HIV infection. The effects of EHI_own and EHI_other on number of

[30] Note that one could regress each of the outcome variables on $1 - EHI\_own_{ij}$ and $1 - EHI\_other_{ij}$ instead, and that would have yielded coefficients equal to those that will be obtained in the current regression, but with the opposite sign.

sexual partners are very close across columns (Table 4-Panel A). In Column (1) in which only EHI_own and EHI_other are included, the coefficients on these two variables are respectively -0.194 and -0.171, but the difference in the magnitude of their effects decreases as more controls are included. In Column (6) in which all controls are included, the coefficients on the two variables are respectively -0.065 and -0.063, and we cannot reject the null hypothesis that they are equal. Furthermore, we cannot reject the null hypothesis that they are both equal in absolute value to the coefficient (0.066) on EHI obtained in Column (6) of Table 2-Panel A.

However, the controlled and non-controlled effects of EHI_own and EHI_other on HIV infection are different in magnitude (Table 4-Panel B). In Column (1), the coefficients on these two variables are -0.061 and -0.034 respectively, but the difference in these effects decreases as more controls are added. In Column (6) which includes all controls, the coefficients on these variables are respectively -0.029 and -0.011. We note that only the effect of EHI_own is closer in abosulte value to that of EHI in Table 2-B. In Column (6), we cannot reject the null hypothesis that the coefficient on EHI_own is equal in absolute value to the coefficient (0.032) on EHI in Column (6) of Table 2-Panel B, but a test for equality between the coefficient on EHI_other and -0.032 is rejected at the level 0.07.[31]

There is little evidence of significant empirical effects of the group level heterogeneities suspicion of which led to the decomposition exercise conducted here. Although ethnic preferences in long-term relationships like marriage are real in all societies, the equality of the effects of EHI_own and EHI_other suggests that such preferences do not exist in the sort of, often short-term, relationships examined here as assumed in the theoretical model. The difference in the effects of these variables on HIV infection is small, but understanding the mechanism behind this would be of interest. It might be, for instance, that sexual practices vary according to whether partners belong to the same group or are from different groups. Future research along this line promises to yield interesting findings.

5 Sensitivity analysis

5.1 Effect of EHI by urban/rural place of residence

Despite the fact that the effect of EHI on number of sexual partners or on HIV infection is robust to the control of urban/rural place of residence (Table 2), we estimate the effect of this variable on each of the two outcomes separately for rural and urban areas. Because urban areas are known to be a bit more ethnically

[31] Note however that a test for equality between the coefficient on EHI_other and -0.029 could not be rejected.

diverse than rural areas, this analysis will show the extent to which the effects of EHI do not simply reflect urban-rural differences in sexual behavior and HIV prevalence. Another advantage to conducting this analysis is that clusters are generally large in rural areas, and are thus likely to capture all the sexual interactions involving theirs residents.[32]

The results are presented in Table 5. Panel A and Panel B show the effects of EHI on number of sexual partners and on HIV infection, respectively. For each of the two panels, Columns (1)-(4) show the estimates for urban areas, and Columns (5)-(8) show the estimates for rural areas. We note that the effects of EHI on number of partners are positive, and are similar in urban and rural areas when no control is included (Panel A, Columns (1) and (2)). These effects diminish with the addition of controls, but are robust to their inclusion, and are greater in magnitude in urban areas compared to rural areas (Panel A, Columns (2)-(4) and Columns (6)-(8)). That the effects are smaller in rural areas may reflect that the cost of infidelity is greater in these areas which are known to be more conservative, and/or that information flows more easily in these areas due to their associative life style, as opposed to the individualistic life style that often characterizes urban areas (see Proposition 3). The effects of EHI on HIV infection are positive and are robust to the inclusion of controls. We also note that these effects are greater in rural areas compared to urban areas (Panel B, Columns (1)-(4) and Columns (5)-(8)).

These findings show that the effects of EHI do not simply reflect urban-rural differences in sexual behavior and HIV prevalence.

5.2 Effect of EHI by migration status

We estimate the effects of EHI on sexual behavior and HIV infection for migrants and non-migrants separately. This analysis addresses the potential bias in the effect of EHI on number of sexual partners possibly caused by the fact that sexual interactions among migrants might not have taken place in their clusters of residence at the time of the survey.[33] This bias would be plausible especially if clusters in which sexual interactions occurred and those in which these migrants resided at the time of the survey had different levels of ethnic diversity. But given that non-migrants are likely to sexually interact in their clusters (even if their partners include migrants), this bias would be minimal among them.

The results are presented in Table 5. Panel C and Panel D show the effects of EHI on number of sexual partners and on HIV infection, respectively. For each panel, Columns (1)-(4) show the estimates for

[32] This issue however has already been resolved in section 4-2 in greater details.
[33] To minimize this bias, we exclude visitors from the sample of migrants.

non-migrants, and Columns (5)-(8) show the estimates for migrants. We find that the effect of EHI on number of partners is positive, and is similar in magnitude for non-migrants and migrants whether controls are included or not (Panel C, Columns (1)-(4) and Columns (5)-(8)).[34] The effect of EHI on HIV infection is positive and is robust to the inclusion of controls, and is slightly greater for migrants (Panel B, Columns (1)-(4) and Columns (5)-(8)).

Given that the effect of EHI on each outcome is similar for non-migrants and migrants, this establishes that the effect of EHI for migrants suffers little bias.

5.3 Effect of EHI corrected for sample bias due to HIV non-response

We estimate the effect of EHI on HIV infection taking into account the fact that not all interviewees that were eligible for HIV testing allowed themselves to be tested. HIV response rates were higher in some countries, in some communities within countries, and among certain individuals. This is likely to introduce a selection bias in our estimate of the effect of EHI on HIV infection, especially if compliance and EHI are correlated. We attempt to correct for this potential bias by restricting our analysis to sub-samples with varying HIV response rates. Our results are shown in Table 6. For clarity, we present the estimates based on the entire sample in Panel A, and the other estimates in Panels B, C and D.

5.3.1 Effect of EHI by country-level HIV response rate

We estimate the effect of EHI on HIV infection separately for countries in which the HIV response rate was high and those in which it was low. The HIV response rate was high in Cameroon (95 percent), Burkina Faso (86 percent), and Ghana (82 percent), and was low in Kenya (72 percent), Malawi (72 percent), and Ethiopia (75 percent). The results for these two groups of countries are shown in Table 6, Panel B. In Column (1), we control only for EHI. We find that the coefficient on EHI is more than twice as large for low response rate countries as for high response rate countries (0.111 versus 0.048); but when controls are included (Column (2)), the difference between the two estimates vanishes (0.031 versus 0.029), and both are statistically significant at the level 0.1%. The estimates in Column (2) are also similar in magnitude to the estimated effect of EHI when all countries are pooled together (Table 6, Panel A, Column (2)). These results strongly suggest that the selection bias due to HIV non-response is entirely driven by observable characteristics, implying that unobservables are not a great cause for concern.

In results not shown, we find a positive and statistically significant effect of EHI on HIV for each

[34] The coefficient on EHI is 0.157 for non-migrants and 0.150 for migrants when no control is included (Columns (1) and (5)), and is respectively 0.072 and 0.069 when all controls are included (Columns (4) and (8)).

country.[35][36] Overall, country-level results are robust, introducing ethnic diversity as an important factor in the spatial variation of HIV prevalence *within* countries, and implying that the results based on the pooled sample are not driven by a few countries.

5.3.2 Effect of EHI by cluster-level HIV response rate

We also estimate the effect of EHI on HIV infection by cluster-level HIV response rate. Clusters are divided into three sub-groups: those with response rate lower than 80%, those with response rate weakly greater than 80% and weakly lower than 99%, and those with response rate equal to 100%.[37] The results are presented in Table 6, Panel C. We find that when controls are not included (Column (1)), the effects of EHI are similar for the three sub-groups, and are very close to the estimate (0.068) based on the entire sample (Panel A, Column (1)). When all controls are included (Column (2)), all the estimates are still similar and are close to the estimate obtained from the entire sample (0.032). These results again indicate that the potential bias caused by HIV non-response is insignificant.

5.3.3 Effect of EHI by the propensity to be tested for HIV

Finally, we estimate the effect of EHI on HIV taking into account the propensity to be tested for HIV infection. First, we estimate the effect of EHI on HIV testing controlling for all the variables in the regression of HIV infection on EHI in Table 2, Panel B, column (6). We find that the effect of EHI on testing is negative but not statistically significant (result not shown).[38] We predict the propensity score and account for this new variable in the estimate of the effect of EHI on HIV. The results are presented in Table 6-Panel D. In column (1), we estimate the effect of EHI based on the entire sample. In Column (2), we control for all variables in addition to the propensity score and its quadratic term. The effect of EHI is similar to that obtained when the propensity score variables are not controlled for (0.033 versus 0.032).

We also partitioned individuals into three groups based on their propensity score: those with propensity score lower than 80%, those with propensity score weakly greater than 80% and weakly lower than 99%, and those with propensity score greater than 99%. The coefficients on EHI in the three groups are respectively 0.109, 0.047 and 0.069 when no control is included; and are respectively 0.059, 0.024 and 0.074 when controls

[35] However, the effect for Kenya is no longer statistically significant when controls are included, although it is still positive. This might be explained by the fact that in Kenya, communication barriers between ethnic groups is in reality weaker than usually thought (that is, $p - q$ is small in Proposition 2), because Kenya, like Tanzania, has successfully promoted the use of one national language (Swahili) after independence.

[36] Also, see Pongou (2007a, 2007b).

[37] In all clusters with response rate greater than 99%, all interviewees eligible for HIV testing were tested.

[38] This might explain why in Panels B and C, the effect of EHI on HIV varies little by country or cluster HIV response rate once controls are included.

are included. All these estimates are highly statistically significant. Of particular interest is the fact that the effect of EHI for the group with the highest propensity scores (P>99%) is the greatest in magnitude (0.074), despite its small sample size (1457 observations).

Overall, these results show that the selection bias in the effect of EHI on HIV due to HIV non-response is almost entirely driven by observable factors, and it disappears when these factors are controlled for.

6 Is it public goods provision?

The documented positive effect of ethnic heterogeneity on number of sexual partners and HIV prevalence may reflect the fact that ethnic heterogeneity slows economic growth and weakens the provision of public goods. A recent literature on the impact of ethnic diversity on public goods provision argues that people of different groups may have different preferences and as a result cannot achieve a public good allocation that adequately serves each group. This outcome occurs because residents contribute less to public goods if they view their contribution as primarily benefitting people outside of their group.[39] This literature thus suggests that more ethnically diverse communities are more likely to lack public infrastructure such as health care facilities, or generally do not provide high quality health care services. As such, residents of these communities would have lower access to adequate information about methods of prevention against HIV/AIDS, or they may be infected with the AIDS virus when seeking treatment for malaria, for instance, due to a careless injection with a contaminated syringe or contaminated blood during transfusion. If so, the documented effect of ethnic heterogeneity on number of sexual partners and HIV infection is spurious.

In this section, we thus test the hypothesis that this effect is driven by the provision of better public health infrastructure in less ethnically heterogeneous communities. More precisely, we look at the effect of ethnic heterogeneity on variables reflecting supply of and demand for health care. These measures include contact with health care workers and the demand for injections.

6.1 Contact with health care workers

Contact with public health professionals is one of the most important means of relaying adequate information about methods of preventing HIV/AIDS and other sexually transmitted diseases. In all countries except Kenya, the DHS asks women whether they visited a health care facility in the 12 months preceding the

[39] Cultural diversity fosters growth-retarding policies in developing countries, and has been shown to be inversely related to various forms of civic engagement including organizational membership, support for income redistribution, support for public education, public spending on public infractructure, country-level AIDS-related expenditures, trust, voting turnout, and volunteering (Costa and Kahn 2003, Luttmer 2001, Miguel and Gugerty 2005, Lieberman 2007, Alesina, Baqir, and Easterly 1999).

survey. We estimate the effect of ethnic heterogeneity on health care facility visits. The results are presented in Table 7-Panel A. Column (1) includes only EHI as predictor. We note that it has a positive and statistically significant effect on the outcome. When controls are included in Column (2), the effect of EHI completely vanishes and is no longer statistically significant.

Although health care facility visits reflect the supply of health care, they also indicate the demand for a range of health care services. The small impact of ethnic heterogeneity on this variable signifies that controlling for it in a regression estimating the effect of EHI on sexual behavior or HIV infection would leave the size of this effect unaffected. This suggests that this effect is not driven by the demand for health care.

6.2 Demand for injections

We also estimate the effect of ethnic heterogeneity on the demand for injections. In Burkina Faso, Cameroon and Malawi, interviewees were asked to report the number of injections received in the 3 months preceding the survey. This information was collected for men and women in Cameroon and Malawi, and for women only in Burkina Faso. As a result of blood transfusion and injection-induced HIV infection, we must consider whether ethnic heterogeneity has an effect on the demand for injections. The results are reported in Table 7-Panel B. In Column (1), only EHI is included, and its effect is positive and statistically significant. In Column (2), we include controls, and the effect of EHI diminishes considerably and is no longer statistically significant. Similar results (not shown) are obtained for Cameroon when considering the number of blood transfusions in the 3 months preceding the survey as the dependent variable (this variable was not collected in other countries.)

The small effect of ethnic heterogeneity on the demand for health care implies that the consumer's price of health care or the household budget allocated to health is not affected by ethnic heterogeneity. Pollak (1969)'s theory of conditional demand functions thus suggests that the effect of ethnic heterogeneity on the demand for sex does not depend on the demand for health care. Empirically, conditioning on the demand for health care in a regression estimating the effect of EHI on number of sexual partners or on HIV infection leaves the size of this effect unchanged.

The results presented in Table 7-Panels A and B imply that higher sexual promiscuity and HIV prevalence in more ethnically heterogeneous communities are not driven by the supply of or demand for health care services in these communities.[40] These findings are not surprising as the spread of HIV in sub-Saharan Africa

[40] In results not shown, we also estimate the effect of EHI on HIV infection controlling for whether distance and/or transportation to health care center constitute a big issue to an individual, and this does not affect the size of the effect of EHI.

is mainly driven by sexual interactions (UNAIDS 2008). Our findings are also consistent with the fact that antiretroviral treatment was almost non-existent in the countries analyzed here at the time of the survey. Also, very few people really knew about their HIV status, and could not seek treatment even if it existed. Overall, these results and facts lend support to our hypothesis that ethnic heterogeneity affects HIV by providing incentives for people to engage in multiple and strategic sexual partnerships across groups.

7 How does ethnic heterogeneity affect human interactions-unrelated health conditions? The case of anemia

Evidence presented thus far supports the hypothesis that ethnic heterogeneity positively affects HIV infection by directly modifying the intensity of human interactions through increasing the number of sexual partners that each person has. This conclusion is also true because there is no curative treatment for HIV/AIDS. The hypothesized mechanism through which ethnic diversity within a community affects HIV infection would be called into question if this factor had a similar effect on a condition not directly involving prohibited human interactions and for which a curative treatment is not available or is scarce. The plausibility of our hypothesis, however, would be enhanced if such an effect were not found. In this sub-section, we examine the effect of ethnic heterogeneity on anemia, a health condition that reflects a low number of red blood cells in a person's blood. Anemia can arise for a variety of reasons including loss of blood, iron deficiency, kidney disease, liver disease, pregnancy induced water weight, poor nutrition, alcoholism, lack of vitamins and minerals, or even hereditary conditions. Causes of anemia are not related to human interactions. Like HIV, anemia also remains widely untreated in most African countries where it constitutes a major public health issue. In all countries examined in this study except Kenya, a sample of blood was taken for selected women, and was analyzed to determine the level of red blood cells and hemoglobin, the oxygen-carrying protein in the red blood cells. A hemoglobin level below 12.0 g/dl is diagnostic of anemia. The proportion of anemic women was the lowest in Ethiopia (28 percent), but was high in Malawi (45 percent), Cameroon (45 percent), Ghana (45 percent), and Burkina Faso (53 percent).

To estimate the effect of ethnic heterogeneity on anemia, we regress a binary indicator for whether a women is anemic or not on EHI. Results are presented in table 8. In Column (1), we include only EHI as predictor. We note that the coefficient on this variable is negative and not statistically different from zero. When controls are included (Column (2)), the effect of ethnic heterogeneity is even closer to zero.

We conclude that EHI does not have a net effect on health conditions that are not caused or channeled by prohibited human interactions, and which like HIV/AIDS are not driven by a lack of public goods, and for which treatment is generally not sought. This further enhances the plausibility of our hypothesis that ethnic heterogeneity affects HIV infection by directly modifying the intensity of human interactions through providing incentives for strategic cross-group sexual networking.

8 Conclusion

This research tests a theory of strategic sexual network formation in an ethnically heterogeneous community with limited communication across ethnic groups. The model is applied to understanding the effect of ethnic diversity on sexual behavior and HIV infection in sub-Saharan Africa. Built upon the assumption that individuals derive utility from sexual relationships with opposite sex partners, and that sexual infidelity is socially prohibited and penalized if detected, the model implies a mechanism wherein ethnic heterogeneity provides an incentive for optimizing agents to choose sexual partners across ethnic groups rather than within, in order to lower the probability of infidelity detection. Ethnic heterogeneity thus encourages infidelity and positively affects the spread of HIV/AIDS, a new hypothesis that we also test using individual level data from six countries that represent the geographical and cultural diversity of sub-Saharan Africa and the situation of HIV/AIDS in the region. We find a positive effect of ethnic heterogeneity on both the number of sexual partners and HIV infection. This effect is also robust to an alternative measure of ethnic heterogeneity in which the index of ethnic diversity is decomposed into two indexes representing respectively the distinct contribution of individuals' own and opposing ethnic groups to the overall index. This effect is also robust to separate estimation for urban and rural areas, and for migrants and non-migrants. Robustness checks show that this effect is not driven by a lack of health-related public goods in ethnically diverse communities. Interestingly, ethnic heterogeneity is shown to have no effect on anemia, which is very prevalent in most African countries, is largely untreated like HIV/AIDS, but which unlike HIV/AIDS does not involve socially prohibited human interactions.

Before we conclude, it would perhaps be appropriate to highlight some important features of the fidelity model and discuss some of its applications and a possible extension. The proof of Proposition 2 implies that the marginal effect of ethnic heterogeneity on multiple partnerships is negligible if the level of ethnic integration is sufficiently important that barriers between groups are minimal. African countries in which successful ethnic integration has been achieved are few. After independence, most African countries embarked on policies aimed at healing ethnic divisions introduced or exacerbated by colonial powers implementing the “divide and rule” strategy. While these policies have been relatively successful in countries like Tanzania, in most countries however, they have only provided an institutional platform on which different ethnic nationalities could interact in the interest of the state, but have not been powerful enough to completely erase cultural barriers between groups. In predominantly Muslim countries however, religion, through the prescription of strong norms and practices, has proven to be a factor of unity, and has eventually minimized or eclipsed the role of ethnic identification in dividing the society.[41] One would therefore expect the effect of ethnic diversity on multiple sexual partnerships to be small in such countries. The fidelity model could also be extended to an asymmetric setting in which an individual's marginal utility of sex depends on whether his/her partner belongs to his/her group or to an opposing group.[42] In a community featuring very strong ethnic divisions, this marginal utility can be very small if the partner belongs to an opposing group (that is, b can be very small), implying that most sexual interactions would occur within groups.[43] This can indeed be the case in countries like Rwanda, Liberia, Somalia or the Democratic Republic of Congo where years of civil wars have exacerbated ethnic divisions. In such societies, one would expect little effect of ethnic heterogeneity on strategic cross-group sexual networking, and low HIV prevalence. In most African countries however, policies of ethnic integration have been carried out only with partial success, leading to moderate ethnic divisions, and allowing different groups to interact without completely uniting. In these countries, the effect of ethnic diversity on strategic cross-group sexual networking is important as our analysis has shown.

By demonstrating a positive effect of ethnic heterogeneity on HIV prevalence *within countries*, our study contributes to a better understanding of the persistence of the global AIDS epidemic and its concentration in sub-Saharan Africa, the most ethnically fragmented region of the world. In doing so, it also helps us to understand some limitations of existing policies aimed at curbing the spread of this disease in most African societies. From a theoretical point of view, our study also implies that diseases which involve

[41]This mostly applies to sub-Saharan Africa where the Sunni and Shi'a divisions are not as salient as in some countries of the Middle-East.

[42]In this setting, the utility function of an agent i belonging to an ethnic group r can be expressed as:

$$u(s_r, s_{-r}) = as_r + bs_{-r} - c(s_r, s_{-r})$$

where a and b are real numbers, and $c(s_r, s_{-r})$ an infidelity cost function of the type studied in the fidelity model.

[43]b can be very small not because of the quality of sexual intercourse with a partner belonging to an opposing group, but because of the high maintenance cost that such a relationship could imply. In fact, each of the first two terms of the right hand side of the utility function $u(s_r, s_{-r}) = as_r + bs_{-r} - c(s_r, s_{-r})$ can be written as a linear combination of intrinsic utility derived from the act of sex and the maintenance cost of sexual partnerships, which is:

$$u(s_r, s_{-r}) = a_1 s_r + a_1 s_{-r} - a_2 s_r - b_2 s_{-r} - c(s_r, s_{-r})$$

where a_1, the marginal intrinsic utility of sex, is the same for own group's and opposing group's partners; and a_2 and b_2 the marginal maintenance cost. We have $a = a_1 - a_2$ and $b = b_1 - b_2$. If b_2 is very high due to the difficulty to maintain out-group relationships because of ethnic tensions for instance, then b will be very small and can even turn negative, thus precluding any out-group relationship.

socially prohibited human interactions spread through hidden and anonymous networks. Empirically, we have provided an instance in which anonymity, fueled by ethnic diversity, affects such deviant behaviors like infidelity. In general, our findings speak to the AIDS tragedy of incomplete and partial ethnic integration in most African countries. Public policies aimed at reversing the course of the AIDS epidemic should therefore also pursue the goal of *complete* social and ethnic cohesion.

References

Akerlof, George A. (1997): "Social distance and social decisions," *Econometrica* 65(5), 1005-1027.

Akerlof, George A., and Rachel Kranton (2000): "Economics and Identity," *The Quarterly Journal of Economics*, 115(3), 715–753.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003): "Fractionalization," *Journal of Economic Growth* 8(2), 155-194.

Alesina, Alberto., Reza Baqir, and William Easterly (1999): "Public Goods and Ethnic Divisions," *The Quarterly Journal of Economics* 114(4), 1243–1284.

Arnafi, John (1993): "Sexuality, migration and AIDS in Ghana-A socio-behavioral study," *Health Transition Review* 4, 273-295.

Asamoah-Odei, Emil, Jesus M. G. Calleja, and J Ties Boerma (2004): "HIV Prevalence and Trends in Sub-Saharan Africa: No Decline and Large Subregional Differences," *The Lancet* 364(3), 35-40.

Aumann, Robert, and Roger Myerson (1998): *Endogenous formation of links between players and coalitions: An application of the Shapley value*, in "The Shapley Value" (Alvin E. Roth, Ed.) Cambridge Univ. Press, Cambridge.

Bala, Venkatesh, and Sanjeef Goyal (2000a): "A Noncooperative Model of Network Formation," *Econometrica* 68(5), 1181-1230.

Barth, Fredrik (1969): Ethnic Groups and Boundaries. The Social Organization of Cultural Difference. Oslo: Universitetsforlaget.

Bloom, David E., and Ajay S. Mahal (1997): "Does the AIDS Epidemic Threaten Economic Growth?," *Journal of Econometrics* 77(1), 105-124.

Bonnel, René (2000): "HIV/AIDS and Economic Growth: A Global Perspective," *South African Journal of Economics* 68(5), 360-379.

Canning, David (2006): "The Economics of HIV/AIDS in Low-Income Countries: The Case for Prevention," *Journal of Economic Perspectives* 20(3), 121-142.

Cinyabuguma, Matthias, and Louis Putterman (2006): "Sub-Saharan growth surprises: Geography, Institutions and History in an All African Data Panel," Working Paper 21, Department of Economics, Brown University.

City Population. http://www.citypopulation.de/

Costa, Dora L., and Matthew E. Kahn (2003): "Civic Engagement and Community Heterogeneity: An Economist's Perspective," *Perspectives on Politics* 1(1), 103-111.

Des Jarlais, Don C., Denise Paone, Judith Milliken, Charles F. Turner, Heather Miller, James Gribble, Qiuhu Shi, Holly Hagan, Samuel R. Friedman (1999): "Audio-Computer Interviewing to Measure Risk Behaviour for HIV among Injecting Drug Users: A Quasi-Randomised Trial," *The Lancet* 353, 1657–1661.

Duclos, Jean-Yves, Joan Esteban, and Debraj Ray (2004): "Polarization: Concepts, Measurement, Estimation," *Econometrica* 72(6), 1737—1772.

Easterly, William, and Ross Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *The Quarterly Journal of Economics* 112(4), 1203-1250.

Eifert Benn, Edward Miguel, and Daniel N. Posner (2007): "Political Competition and Ethnic Identification in Africa," Afrobarometer Working Paper No. 89.

Epstein, Helen (2007): *The Invisible Cure: Africa, The West, and The Fight Against AIDS.* New York: Farrar, Straus and Giroux.

Fortson, Jane (2008a): "Mortality Risks and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa," Mimeo, University of Chicago.

Fortson, Jane (2008b): "The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS," *Demography* 45(2), 303-322.

Gray, Ronald H., Maria J. Wawer, Ron Brookmeyer, Nelson K. Sewankambo, David Serwadda, Fred Wabwire-Mangen, Tom Lutalo, Xianbin Li, Thomas vanCott, Thomas C. Quinn, and the Rakai Project Team (2001): "Probability of HIV-1 Transmission per Coital Act in Monogamous Heterosexual, HIV-1 Discordant Couples in Rakai, Uganda," *The Lancet* 357, 1149-1153.

Halperin, Daniel T., Helen Epstein (2004): "Concurrent Sexual Partnerships Help to Explain Africa's High HIV Prevalence: Implications for Prevention," *The Lancet* 364 (9428), 4-6.

Heuveline, Patrick (2004): "Impact of the HIV Epidemic on Population and Household Structure: The Dynamics and Evidence Today," *AIDS* 18(Suppl. 2), S45-S53.

Hewett, Paul C., Barbara S. Mensch, Annabel S. Erulkar (2004): "Consistency in the Reporting of Sexual Behaviour by Adolescent Girls in Kenya: A Comparison of Interviewing Methods," *Sexually Transmitted Infections* 80: ii43-ii48.

Hogan, Daniel, and Joshua A. Salomon (2005): "Prevention and treatment of human immunodeficiency virus/acquired immunodeficiency syndrome in resource limited settings," *Bulletin of the World Health Organization* 83, 135-143.

Horowitz, Donald L. (1985): Ethnic groups in conflict, Berkeley: University of California Press.

Jackson, Matthew O. (2004): A Survey of Models of Network Formation: Stability and Efficiency, Chapter 1 in Gabrielle Demange and Myrna Wooders (eds) "Group Formation in Economics; Networks, Clubs and Coalitions," Cambridge University Press, Cambridge U.K.

Jackson, Matthew O., and Asher Wolinsky (1996): "A Strategic Model of Social and Economic Network," *Journal of Economic Theory* 71, 44-74.

Jaya, Michelle J. Hindin, and Saifuddin Ahmed (2008): "Differences in Young People's Reports of Sexual Behaviors According to Interview Methodology: A Randomized Trial in India," *American Journal of Public Health* 98(1): 169 - 174.

Lieberman, Evan S. (2007): "Ethnic Politics, Risk, and Policy-Making: A Cross-National Statistical Analysis of Government Responses to HIV/AIDS" Comparative Political Studies 40(12), 1407-1432.

Lopez, Alan D., Colin D. Mathers, Majid Ezzati, Dean T. Jamison, and Christopher J. L. Murray (2006): *Global Burden of Disease and Risk Factors*, Oxford University Press and the World Bank.

Lurie, Mark N., Brian G. Williams, Khangelani Zuma, David Mkaya-Mwamburi, Geoff P. Garnett, Adriaan W. Sturm, Michael D. Sweat, Joel Gittelsohn, Salim S. Abdool Karim (2003): "The Impact of Migration on HIV-1 Transmission: A Study of Migrant and Non-Migrant Men, and Their Partners," *Sexually Transmitted Diseases* 40(2), 149-156.

Luttmer, Erzo F.P. (2001): "Group Loyalty and the Taste for Redistribution," *Journal of Political Economy* 109(3), 500–28.

Matshalaga, Neddy R. (2002): "Mass Orphanhood in the Era of HIV/AIDS" *The British Medical Journal* 324, 185-186.

Mberu, Blessing U. (2008): Internal Migration and Premarital Reproductive Behavior Among Adolescents in Nigeria. PhD Dissertation, Department of Sociology, Brown University.

Measure DHS. http://www.measuredhs.com/

Mensch Barbara S., Paul C. Hewett, Annabel S. Erulkar (2003): "The Reporting of Sensitive Behaviour by Adolescents: A Methodological Experiment in Kenya," *Demography* 40, 247–68.

Miguel, Edward, and Mary K. Gugerty (2005): "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics* 89 (11-12), 2325-2368.

Mishra, Vinod, Simona Bignami-Van Assche, Robert Greener, Martin Vaessen, Rathavuth Hong, Peter Ghys, J. Ties Boerma, Ari Van Assche, Shane Khan, Shea Rutstein (2007): "HIV Infection Does not Disproportionately Affect the Poor," *AIDS* 21(Suppl. 7), S17-S28.

Monasch, Roeland, and Ties J. Boerma (2004): "Orphanhood and Childcare Patterns in Sub-Saharan Africa: An Analysis of National Surveys from 40 countries," *AIDS* 18 (suppl. 2), S55-S65.

Morris, Martina, and Miriam Kretzschmar (1997): "Concurrent Partnerships and the Spread of HIV," *AIDS* 11(5), 641-648.

Orubuloye, I.O., and Folakemi Oguntimehin (1999): Death is Pre-ordained, It Will Come When It Is Due: Attitude of Men to Death in the Presence of AIDS in Nigeria, in Caldwell et al. (eds.): "Resistances to Behavioural Change to Reduce HIV/AIDS Infection in Predominantly Heterosexual Epidemics in Third World Countries," *Health Transition Centre*, 1999.

Oster, Emily (2005): "Sexually Transmitted Infections, Sexual Behavior and the HIV/AIDS Epidemic," *The Quarterly Journal of Economics* 120(2), 467-515.

Oster, Emily (2007): "HIV and Sexual Behavior Change: Why Not Africa?," Mimeo, University of Chicago.

Pollak, Robert A. (1969): "Conditional Demand Functions and Consumption Theory," *The Quarterly Journal of Economics* 83(1), 60-78.

Pongou, R. (2007a): "Ethnic Diversity and HIV/AIDS in Cameroon," Mimeo. Department of Economics, Brown University.

Pongou, R. (2007b): "Anonymity and Infidelity: Ethnic Identity, Strategic Cross-Ethnic Sexual Network formation, and HIV/AIDS in Africa," Mimeo. Department of Economics, Brown University.

Pongou, R. (2009a): "Fidelity and the Contagion Potential of Unimagined Networks," Mimeo. Department of Economics, Brown University.

Posner, Daniel N. (2003): "The Colonial Origins of Ethnic Cleavages: The Case of Linguistic Divisions in Zambia," *Comparative Politics* 35(2), 127-146.

Potts, Malcom (2000): "Male Circumcision and HIV Infections," *The Lancet* 355(11), 926.

Potts, Malcom, Daniel T. Halperin, Douglas Kirby, Ann Swidler, Elliot Marseille, Jeffrey D. Klausner, Norman Hearst, Richard G. Wamai, James G. Kahn, Julia Walsh (2008): "Reassessing HIV Prevention," *Science* 320, 749-750.

Royce, Rachel A., Arlene Seña, Willard Cates, and Myron S. Cohen (1997): "Sexual Transmission of HIV," *The New England Journal of Medecine* 336(15), 1072-1078.

Thornton, Rebecca (2006): "The Demand for and Impact of Learning HIV Status: Evidence from a Field Experiment," Mimeo, Harvard Unviersity.

UNAIDS, 2008. 2008 Report on the Global AIDS Epidemic. Executive summary. UNAIDS/08.27E / JC1511E.

Zaba, Basia, Elizabeth Pisani, Emma Slaymaker, and J. Ties Boerma (2004): "Age at First Sex: Understanding Recent Trends in African Demographic Surveys," *Sexually Transmitted Diseases* 80(Suppl. II), ii28-ii35.

Table 1: Summary statistics

Variables	# Observations	Mean	Std. Dev.
Is infected with HIV	49696	0.044	0.205
Number of reported sexual partners including husband/wife in past 12 months	83819	0.781	1.099
Number of reported sexual partners excluding husband/wife in past 12 months	83819	0.210	1.057
Cluster-level ethnic heterogeneity index (EHI)	83819	0.319	0.287
Region/urban/rural-level ethnic heterogeneity index (EHI)	83819	0.550	0.247
Community prevalence of multiple sexual partnership	83819	0.048	0.069
Male	83819	0.318	0.466
Age	83819	28.641	10.315
Never married	83819	0.300	0.458
Married	83819	0.568	0.495
Cohabiting	83819	0.057	0.231
Divorced	83819	0.027	0.162
Christian	83819	0.668	0.471
Muslim	83819	0.262	0.440
Other religion	83819	0.070	0.255
Education (in years)	83819	4.762	4.505
Unemployed	83819	0.305	0.460
Traditional sector employee	83819	0.628	0.483
Modern sector employee	83819	0.063	0.242
Urban	83819	0.318	0.466
Non-migrant	83819	0.437	0.496
Regional population density (# people per km^2)	83819	632.486	1660.586

Summary statistics are based on pooled Demographic and Health Surveys data from six countries including Burkina Faso, Cameroon, Ethiopia, Ghana, Kenya and Malawi. HIV testing was conducted on a representative sub-sample strictly included in the sample of all interviewees. Traditional sector employment includes agriculture, sale and manual work. Modern sector employment includes professional and managerial work and services.

Table 2: OLS estimates of the effects of ethnic heterogeneity on reported number of sexual partners in the past 12 months and on HIV infection

Panel A	**# sexual partners in the past 12 months**					
	(1)	(2)	(3)	(4)	(5)	(6)
EHI	0.199***	0.206***	0.180***	0.171***	0.163***	0.066***
	[0.023]	[0.022]	[0.022]	[0.022]	[0.022]	[0.023]
# observations	83819	83819	83819	83819	83819	83819
R-squared	0	0.09	0.09	0.09	0.09	0.11
Panel B	**Is infected with HIV (0/1)**					
	(1)	(2)	(3)	(4)	(5)	(6)
EHI	0.068***	0.060***	0.061***	0.060***	0.060***	0.032***
	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]	[0.005]
# observations	49696	49696	49696	49696	49696	49696
R-squared	0.01	0.03	0.03	0.03	0.03	0.07
Controls						
Demographics and socioeconomics	NO	YES	YES	YES	YES	YES
Urban/rural residence	NO	NO	YES	YES	YES	YES
Migration status	NO	NO	NO	YES	YES	YES
Regional population density	NO	NO	NO	NO	YES	YES
Ethnic group fixed effect	NO	NO	NO	NO	NO	YES

Demographics and socioeconomics include gender, age education, religion, occupation, and marital status for Panel A, in addition to community sexual behavior measured as the prevalence of multiple sexual partnerships within a cluster for Panel B. Controlling for ethnic group fixed effect is equivalent to controlling for 161 binary variables representing all the ethnic groups in the six countries considered in our analysis. Robust standard errors in brackets are corrected for intracluster correlation.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: OLS estimates of the effects of EHI computed at the cluster level and at the region/urban/rural level

Panel A	**# sexual partners in past 12 months**				
	(1)	(2)	(3)	(4)	(5)
EHI (cluster level)	0.199***	0.066***			0.053**
	[0.023]	[0.023]			[0.026]
EHI (region/urban/rural level)			0.315***	0.078*	0.056
			[0.071]	[0.046]	[0.045]
# Observations	83819	83819	83819	83819	83819
R-squared	0	0.11	0.01	0.11	0.11
Controls	NO	YES	NO	YES	YES
Panel B	**Is infected with HIV (0/1)**				
	(1)	(2)	(3)	(4)	(5)
EHI (cluster level)	0.068***	0.032***			0.030***
	[0.005]	[0.005]			[0.007]
EHI (region/urban/rural level)			0.066***	0.023**	0.011
			[0.023]	[0.010]	[0.009]
# Observations	49696	49696	49696	49696	49696
R-squared	0.01	0.07	0.01	0.07	0.07
Controls	NO	YES	NO	YES	YES

Controls include gender, age education, religion, occupation, marital status, urban/rural place of residence, migration status, population density and ethnic group fixed effect for Panel A, and community sexual behavior in addition for Panel B. Robust standard errors in brackets are corrected for within-cluster correlation in Columns (1)-(2) of each panel, and for within-region/urban/rural stratum correlation for Columns (3)-(5) of each panel.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: OLS estimates of the effects of "own" and "opposing" ethnic groups' contribution to EHI on number of sexual partners and on HIV infection

Panel A	**Number of sexual partners in the past 12 months**					
	(1)	(2)	(3)	(4)	(5)	(6)
EHI (own ethnic group)	-0.194***	-0.211***	-0.185***	-0.178***	-0.170***	-0.065***
	[0.024]	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]
EHI (opposing ethnic groups)	-0.171***	-0.234***	-0.206***	-0.209***	-0.202***	-0.063*
	[0.038]	[0.037]	[0.036]	[0.036]	[0.036]	[0.035]
# observations	83819	83819	83819	83819	83819	83819
R-squared	0	0.09	0.09	0.09	0.09	0.11
Panel B	**Is infected with HIV (0/1)**					
	(1)	(2)	(3)	(4)	(5)	(6)
EHI (own ethnic group)	-0.061***	-0.053***	-0.055***	-0.054***	-0.054***	-0.029***
	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
EHI (opposing ethnic groups)	-0.034***	-0.026**	-0.028**	-0.028**	-0.028***	-0.011
	[0.010]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
# observations	49696	49696	49696	49696	49696	49696
R-squared	0.01	0.03	0.03	0.03	0.03	0.07
Controls						
Demographics and socioeconomics	NO	YES	YES	YES	YES	YES
Urban/rural residence	NO	NO	YES	YES	YES	YES
Migration status	NO	NO	NO	YES	YES	YES
Regional population density	NO	NO	NO	NO	YES	YES
Ethnic group fixed effect	NO	NO	NO	NO	NO	YES

Demographics and socioeconomics include gender, age education, religion, occupation, and marital status for Panel A, in addition to community sexual behavior measured as the prevalence of multiple sexual partnerships within a cluster for Panel B. Controlling for ethnic group fixed effect is equivalent to controlling for 161 binary variables representing all the ethnic groups in the six countries considered in our analysis. Robust standard errors in brackets are corrected for intracluster correlation.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: OLS estimates of the effects of EHI on reported number of sexual partners and HIV infection by urban/rural residence and by migration status

	Urban areas				**Rural areas**			
Panel A: # sexual partners in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EHI	0.236***	0.238***	0.213***	0.109*	0.230***	0.138***	0.139***	0.044**
	[0.052]	[0.048]	[0.049]	[0.056]	[0.023]	[0.021]	[0.021]	[0.017]
# Observations	26691	26691	26691	26691	57128	57128	57128	57128
R-squared	0	0.06	0.06	0.09	0	0.13	0.13	0.16
Panel B: Is infected with HIV (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EHI	0.052***	0.057***	0.056***	0.020**	0.074***	0.064***	0.063***	0.033***
	[0.008]	[0.009]	[0.009]	[0.009]	[0.008]	[0.008]	[0.008]	[0.007]
# Observations	15639	15639	15639	15639	34057	34057	34057	34057
R-squared	0	0.03	0.03	0.07	0.01	0.03	0.03	0.07
Controls								
Demographics and socioeconomics	NO	YES	YES	YES	NO	YES	YES	YES
Migration status + Regional population density	NO	NO	YES	YES	NO	NO	YES	YES
Ethnic group fixed effect	NO	NO	NO	YES	NO	NO	NO	YES
	Nonmigrants				**Migrants**			
Panel C: # sexual partners in past 12 months	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EHI	0.157***	0.202***	0.177***	0.072***	0.150***	0.175***	0.148***	0.069**
	[0.026]	[0.026]	[0.028]	[0.026]	[0.030]	[0.029]	[0.027]	[0.030]
# Observations	36636	36636	36636	36636	45486	45486	45486	45486
R-squared	0	0.17	0.17	0.19	0	0.06	0.06	0.08
Panel D: Is infected with HIV (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EHI	0.055***	0.047***	0.055***	0.024***	0.064***	0.064***	0.064***	0.036***
	[0.006]	[0.006]	[0.007]	[0.007]	[0.006]	[0.007]	[0.008]	[0.007]
# Observations	22499	22495	22495	22495	26294	26294	26294	26294
R-squared	0.01	0.03	0.03	0.07	0.01	0.02	0.02	0.07
Controls								
Demographics and socioeconomics	NO	YES	YES	YES	NO	YES	YES	YES
Urban/rural residence + Regional population density	NO	NO	YES	YES	NO	NO	YES	YES
Ethnic group fixed effect	NO	NO	NO	YES	NO	NO	NO	YES

Demographics and socioeconomics include gender, age education, religion, occupation, and marital status for Panels A and C, and community sexual behavior in addition for Panels B and D. Controlling for ethnic group fixed effect is equivalent to controlling for 161 binary variables representing all the ethnic groups in the six countries considered in our analysis. Robust standard errors in brackets are corrected for intracluster correlation.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: OLS estimates of the effects of ethnic heterogeneity on number of partners and HIV infection by HIV response rate

Panel A: All sample	(1)	(2)	# observations	R-squared	R-squared
	0.068***	0.032***	49696	0.01	0.07
	[0.005]	[0.005]			
Panel B: by country-level HIV response rate (RR)	(1)	(2)	# observations	R-squared	R-squared
High RR countries	0.048***	0.029***	27231	0.01	0.04
	[0.005]	[0.006]			
Low RR countries	0.111***	0.031***	22465	0.02	0.09
	[0.009]	[0.011]			
Panel C: by cluster-level HIV response rate (RR)	(1)	(2)	# observations	R-squared	R-squared
RR < 80%	0.071***	0.030***	14207	0.01	0.08
	[0.009]	[0.011]			
80% ≤ RR ≤ 99%	0.064***	0.030***	24611	0.01	0.07
	[0.007]	[0.007]			
RR = 100%	0.063***	0.038***	10878	0.01	0.06
	[0.011]	[0.014]			
Panel D: by the individual propensity (P) to be tested for HIV	(1)	(2)	# observations	R-squared	R-squared
All sample	0.068***	0.033***	49696	0.01	0.07
	[0.005]	[0.006]			
P < 80%	0.109***	0.059***	8976	0.01	0.08
	[0.014]	[0.018]			
80% ≤ P ≤ 99%	0.047***	0.024***	39263	0.01	0.06
	[0.005]	[0.005]			
P > 99%	0.069***	0.074**	1457	0.01	0.08
	[0.026]	[0.031]			
Controls	NO	YES		NO	YES

High HIV response rate countries include Cameroon (95 percent), Burkina Faso (86 percent) and Ghana (82 percent), and low HIV response rate countries include Ethiopia (75 percent), Kenya (72 percent) and Malawi (72 percent). Controls include community sexual behavior, gender, age education, religion, occupation, marital status, urban/rural place of residence, migration status, population density and ethnic group fixed effect for all panels, and the propensity score and its quadratic term for the "All sample" regression of Panel D. The propensity score is predicted based on a probit regression of HIV infection on community sexual behavior, gender, age education, religion, occupation, marital status, urban/rural place of residence, migration status, population density and ethnic group fixed effect. Robust standard errors in brackets are corrected for intracluster.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Estimates of the effects of ethnic heterogeneity on supply of and demand for health care services.

Panel A	**Visited health facility in past 12 months (0/1)**	
	(1)	(2)
EHI	0.227***	0.007
	[0.014]	[0.015]
# observations	54497	54497
R-squared	0.02	0.14
Controls	NO	YES
Panel B	**# injections received in past 3 months**	
	(1)	(2)
EHI	0.325***	0.051
	[0.046]	[0.048]
# observations	43217	43217
R-squared	0	0.02
Controls	NO	YES

Controls include gender, age, education, religion, occupation, marital status, urban/rural place of residence, migration status, regional population density, ethnic group fixed effect. Robust standard errors in brackets are corrected for intracluster correlation.

* significant at 10%; ** significant at 5%; *** significant at 1%

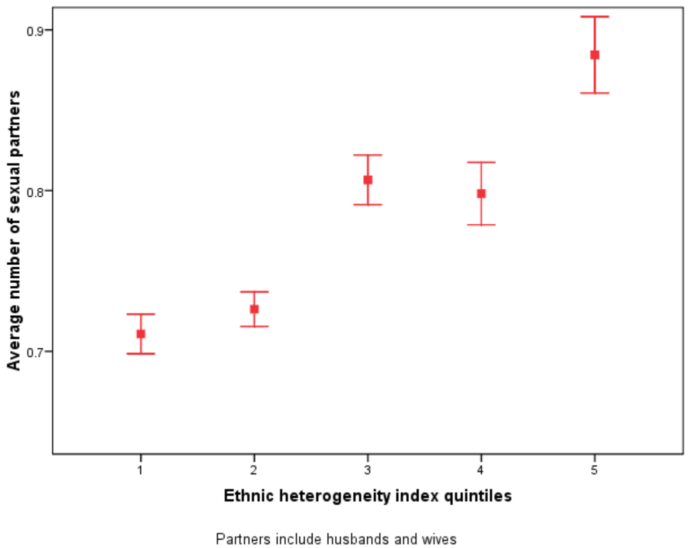
Table 8: Estimates of the effects of ethnic heterogeneity on anemia

	Individual is anemic (0/1)	
	(1)	(2)
EHI	-0.008	0.005
	[0.016]	[0.018]
# observations	23301	23301
R-squared	0	0.06
Controls	NO	YES

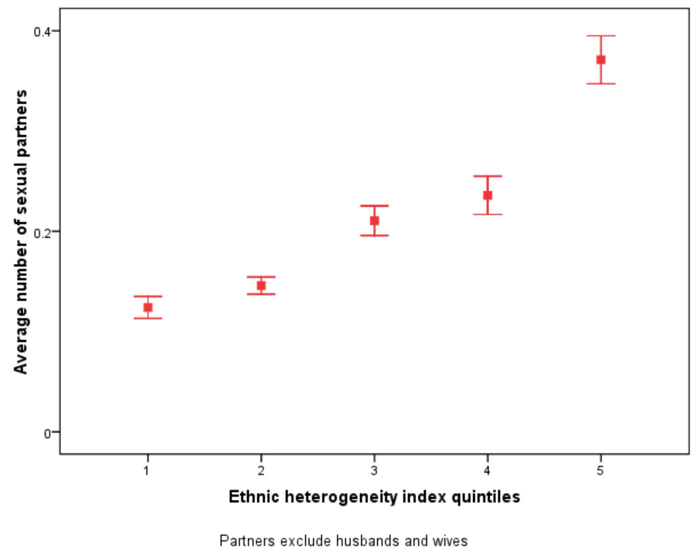
Controls include gender, age, education, religion, occupation, marital status, urban/rural place of residence, migration status, regional population density, ethnic group fixed effect. Robust standard errors in brackets are corrected for intracluster correlation.
* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1: Effect of ethnic heterogeneity on number of sexual partners and HIV prevalence

Panel A



Panel B



Panel C

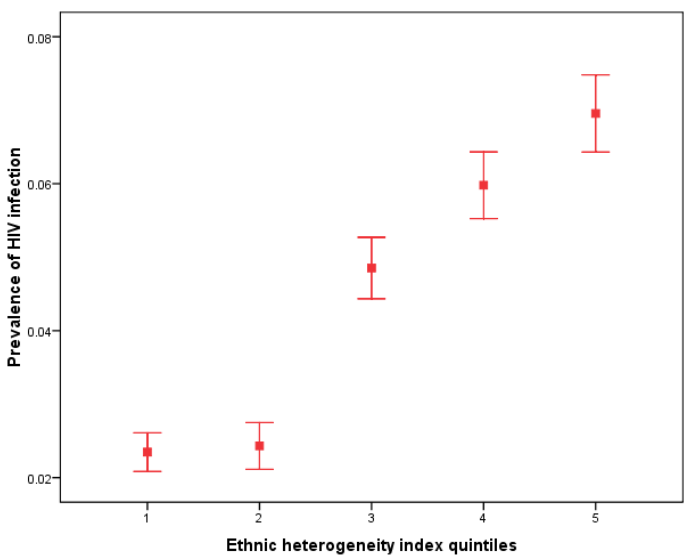


Figure 2-A: Effect of ethnic heterogeneity on number of sexual partners (partners exclude husbands and wives)

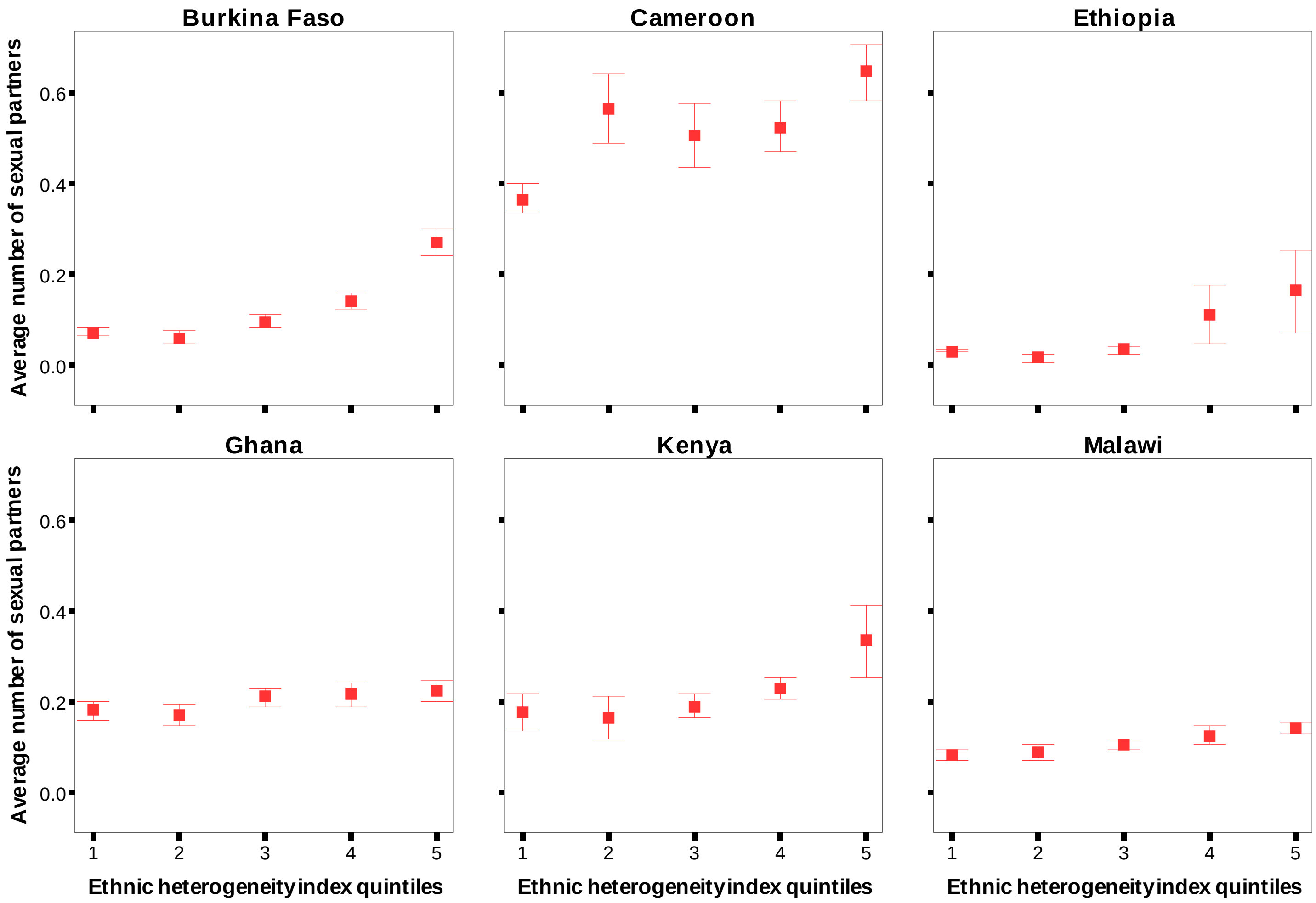


Figure 2-B: Effect of ethnic heterogeneity on HIV prevalence