TSE WP working paper series

09-120

Research Group: *Development*

November, 2009

How do Roads Spread AIDS in Africa? A Critique of the Received Policy Wisdom

ÉLODIE DJEMAI

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

How do Roads Spread AIDS in Africa? A Critique of the Received Policy Wisdom *

Elodie DJEMAI†
Toulouse School of Economics

This version: November 2009
First draft: June 2009

Abstract

This paper empirically analyzes the influence of road proximity on HIV-infection using geographical data on road infrastructure and the Demographic and Health Surveys collected in six African countries. Firstly we show that living in proximity to a major road increases the individual risk of infection. This observed relationship is found to be sensitive to the use of the road and to be robust after correcting for potential selection bias related to the non random placement of people. Secondly, our findings reveal that road infrastructure improves the level of HIV/AIDS-knowledge and facilitates access to condoms, providing no support to the hypothesis that HIV-infection is purely due to ignorance and misfortune. Thirdly, we find that the increased risk of infection is driven by a higher likelihood of engaging in casual sexual partnerships that more than offsets the effect of the increased use of condoms.

JEL Codes:I10, 012, 018, R23, C21

Keywords: HIV/AIDS epidemic, spatial inequalities, risk taking, infrastructure, geography, Sub-Saharan Africa

---

*I thank Jean-Paul Azam for his suggestions and support. This paper has also benefited from rich comments by Andrew Clark, Youenn Lohéac, Celine Nauges, Emily Oster, Paul Seabright, Stephane Straub as well as by participants at the Journées de Microéconomie Appliquée (2009) and at the Toulouse School of Economics Seminar (2009). I am grateful to Jitka Vinduskova for her helpful hints about ArcGis. The usual disclaimer applies.

†Toulouse School of Economics (ARQADE), Manufacture des Tabacs, 21 allée de Brienne, 31000 Toulouse, France, elodiedjemai@gmail.com

1 Introduction

The HIV/AIDS epidemic is among the most difficult of the many challenges facing most countries in Sub-Saharan Africa. Both researchers and policy makers are mobilized in order to find appropriate ways to reduce the propagation and to curb the epidemic. Over the last 25 years, public policies to fight AIDS in Africa have been based upon providing information and condoms. Despite this, the rate of new infections does not fall drastically and even well informed and wealthy people get HIV-infected, suggesting that HIV-contamination is not only a matter of ignorance and misfortune. Earlier papers have worked on other potential determinants that would explain HIV-related behaviors and infection; and mobility has been pointed as one of the factors contributing to the fast spread of the epidemic in Africa (e.g. Oster, 2009). Accordingly, while infrastructure is mentioned as a prerequisite for development, growth, and the improvement of health conditions in developing countries, in the context of the HIV/AIDS epidemic, road infrastructure is also highlighted as a transmitter of the epidemic from region to region. Through the displacement of people, trade along roads contributes to rapid HIV propagation. The impact of road infrastructure on HIV/AIDS outcomes is not straightforward and constitutes an empirical question. The literature on mobility and AIDS suggests that road infrastructure may have a negative externality on this particular health problem. This paper is an attempt to investigate the relationship between road infrastructure and the spread of AIDS at the individual level by estimating the effect of the distance to a road on the individual risk of being HIV-infected. In this paper, using geographical data on road infrastructures and survey data from Demographic and Health Surveys, we consider three questions. First we examine the role of road infrastructure on the spread of AIDS by estimating the effect of proximity to a road on the likelihood of HIV-infection. Second, we investigate whether this relationship between road proximity and HIV-infection is supported by the story of ignorance and misfortune. Finally, we consider how individual preferences for protection differ according the individuals' location.

To answer these questions, we use the most recent Demographic and Health Surveys collected in Cameroon, Ethiopia, Ghana, Kenya, Malawi and Zimbabwe. In this set of African countries, Demographic and Health Surveys (DHS hereafter) provide the three types of information required for our empirical analysis: a complete standardized questionnaire on socio-demographic characteristics and HIV-related knowledge, attitudes and practices; the result of HIV-testing; and geographical data that allow us to locate the sampled cluster to which each individual belongs on a country map. This latter ingredient is the key element to allow for the combination of the survey data with the geographical data on road infrastructure in order to compute the distance between the individual's location and the nearest paved road. The main result of this paper is that proximity to a road increases the individual risk of HIV-infection and that the effect is sensitive to the use of the road. This supports the idea that the effect of road infrastructure on AIDS results from increased mobility and the greater number of opportunities to have sex induced by the presence of a road. We show that the increase in the access to condoms and the improvement of the HIV/AIDS-knowledge induced by road infrastructure are not sufficient to prevent people from getting contaminated. This finding undermines the role of improved information and access to condoms emphasized in public policy-oriented circles. The mechanism driving the relationship between distance to a road and the risk of infection is shown to be the increase in the demand for casual sexual partners that more than offsets the increase in condom use. Hence, these people choose to expose themselves to the risk of infection, despite their better access to self-protective devices.

The empirical strategy we follow in this paper has four characteristics. We estimate the role of road infrastructure on HIV-infection among the general population at the individual level, using the distance from the individual's location to the nearest road as the measure of interest. The approach differs from the existing related literature in four respects. Firstly, a number of papers have stressed the beneficial effects of infrastructure and communication infrastructure, such as road and railways, in the developing world (Jacoby, 2000; Donaldson, 2008; Straub et al, 2008; Banerjee et al, 2009). These papers promote investments in road construction as a way to accelerate development and growth. The role of infrastructure in health outcomes has also been investigated, especially by Fay et al (2005) who looks at the effect of improved sanitation and housing materials in reducing child mortality and child malnutrition using data from the Demographic and Health Surveys. However here we study the effect of roads on health outcomes for the HIV/AIDS epidemic. We test whether the road infrastructure influences the risk of being HIV-infected, arguing that road infrastructure might have two competing effects on the epidemic. On the one hand, road infrastructure facilitates access to markets and hence might facilitate access to the supply of protective measures that could prevent people from being contaminated. On the other hand, road infrastructure facilitates physical communication and this might lead to a rise in the risk of infection, bringing people living close to a road in touch with a mobile population who is more at risk of infection. Indeed, the risk profile of the mobile population has been investigated by sociologists, anthropologists and economists who agree on the fact that a mobile population is more likely to be HIV-infected and to undertake HIV-related risky behavior. Long-distance truck drivers have been the focus of analysis in Oruboloye et al (1993), Huygens (2001), Gouws and Ramjee (2002) and the temporary migrant workers in Meekers (2000) and Adaji Nwokoji and Ajuwon (2004).

Secondly, this paper studies the general population while in the existing literature on HIV/AIDS, the question of mobility is mostly examined from the mobile population's perspective. In this respect, this paper is much more related to Oster (2009) who predicts the regional prevalence rate among the general population. Only Tanser et al (2000) describe the relationship between road and HIV in the rural South African setting using data from Antenatal Clinic Surveillance. They show a correlation between the location of the clinics and the ANC-prevalence rate. This type of analysis should be interpreted carefully because antenatal clinics are not uniformly distributed within countries; their location might be strongly determined by the proximity to the road network and this antenatal surveillance system provides the rate of HIV prevalence among the pregnant women who voluntarily come to these clinics to receive antenatal health care. Thirdly, one point of departure from the related literature is that we use individual-level data to predict the HIV-status and HIV-related behaviors of adults and to estimate the influence of road proximity on these outcomes. Fourth, the originality of the paper is to combine survey data with geographical data and to apply cartographic techniques in order to compute the distance between each sampled cluster and its nearest paved road.

We show that proximity to a major road has a positive and significant impact on the likelihood of being HIV-infected such that increasing the distance to a road by 10% would decrease the risk of being HIV-infected by 0.049 percentage point. When controlling for the effect of mobility by introducing a proxy for the road traffic as a function of trade flows, we find that this observed negative relationship between road distance and HIV-infection is sensitive to the use of the road. In fact, the increased probability of infection resulting from road proximity may be driven by the fact that the agents living in accessible areas are more likely to be in touch with mobile people and particularly foreigners who come from regions or countries where the prevalence rate is different and potentially higher. By contrast, people living in remote areas are somehow protected against HIV propagation since the prevalence rates are more stable than in more accessible areas.

The role of the supply and the demand for protection is examined to disentangle which mechanisms are driving the observed relationship. Access to the supply of protective measures is found to be greater in accessible areas compared to remote ones since the likelihood of having access to condoms and the quality of HIV/AIDS-knowledge decrease with the distance to a road. The observed relationships between road proximity and HIV-infection and between road proximity and access to protective measures suggest that ignorance and misfortune are not driving our results. Road infrastructure provides a better level of knowledge and a better access to condoms but this is not sufficient to prevent people from being infected. This empirical finding suggests that the observed relationship is due to deficiencies in the demand for protection instead of deficiencies in the supply. Indeed, the increase in the risk of infection due to road infrastructure is found to be driven by an increase in the demand for casual sexual partners that offsets the rise in condom use that is found in proximity to a road.

Two main policy implications can be drawn from our empirical findings. The first policy implication concerns the design of public policies. Over the last 25 years, public policies to fight AIDS in Africa have been hinged on providing information about HIV transmission and subsidizing or providing condoms for free. Our results suggest that it is necessary to inform people about the risk and to provide the self-protective measures but this is not sufficient to prevent them from being infected. This paper supports the idea that there is an additional dimension that has to be taken into account, that is the risk taking. In fact to curb the spread of the HIV/AIDS epidemic, it would be relevant to provide people more incentives to use the self-protective measures that are available to them. The second policy implication concerns the road investment. This paper suggests that road infrastructure has additional costs and benefits that were not explored beforehand in the literature. Building more roads will increase the access to condom and improve individual knowledge about the risk of infection but will also rise the prevalence rates.

The paper is organized as follows. Section 2 presents the related literature and develops to what extent road infrastructure might influence HIV/AIDS outcomes. Section 3 describes the empirical strategy, the data from the Demographic and Health Surveys and the geographical data used in this paper. Section 4 explores the role of proximity to a road in the likelihood of HIV-infection. Section 5 examines whether the HIV-infection is a result of ignorance or a deliberated risk by estimating the effect of road infrastructure on the HIV/AIDS-knowledge and the access to condoms and estimates the individual demand for self-protection. Section 6 presents a number of robustness tests and section 7 concludes.

2 Road, health outcomes and HIV/AIDS epidemic

The role of infrastructure on economic performance has been widely investigated in the literature. At the country level, infrastructures such as railroads and roads are found to be strong determinants of development. Investing in infrastructures facilitates trade as it immediately reduces the transportation cost (Jacoby, 2000). The benefits of such an investment are also found years after since disparities in colonial investments in West Africa are found to be one of the main determinant of the current differences in economic outcomes and performance, even decades after Independence (Huillery, 2009). The expansion of countries are interlinked with the access to railroads. In the existing literature, a focus have been stressed on Asian countries (Straub et al, 2008) and particularly China (Banerjee et al, 2009) and India (Donaldson, 2008). Donaldson (2008) and Banerjee et al (2009) studied the role of railroads in trade expansion and in the determination of income level and income growth, respectively. A concern in this latter type of relationship is reverse causality because infrastructures are potentially driving the growth trajectory and at the same time, wealthy countries are more able to finance public investment in infrastructures than poor countries. When controlling for endogeneity problem, the role of railroads in income level and income growth turns out to be mitigated in Banerjee et al (2009). Infrastructures have beneficial effects not only at the country level but also at the individual level as it provides people with extended access to markets and health care facilities among other things. Accordingly, improvement in health conditions and especially in reducing child mortality and child malnutrition were made possible thanks to infrastructures such as improved sanitation and housing materials (Fay et al, 2005).

Even though the direct impact of road infrastructure on HIV/AIDS outcomes has never been explored, there are a number of reasons why HIV-infection might be associated to the proximity of road infrastructure. One might argue that road infrastructure has two competing effects on the risk of infection. Firstly, road infrastructure facilitates physical communication and hence might accelerate the spread of the epidemic. Secondly, road infrastructure reduces the distance to and facilitates the access of the markets, including the market for condoms and for knowledge about the HIV/AIDS epidemic, and hence might reduce the cost of protection and prevent people from being infected.

Firstly, HIV-infection may be associated with the presence of road infrastructure via the extension of the sexual network induced by the presence of a major road and the flow of people using it and, in particular, the flow of mobile population. One may argue that the agents who are living close to a road have a higher individual risk of HIV-infection because they are potentially in touch with a high number of people. In particular, roads are used by people from other regions within a country or from other countries where the prevalence rate is potentially higher. This implies that the probability of getting HIV-infected from a unprotected sexual intercourse with such a partner is higher than with a sexual partner of the same cluster in which the epidemic is contained. On the other hand, people might be in contact with the mobile population and mobile people have been found to be more likely to be HIV-infected and to engage in risky sexual behaviors than the rest of the population. Previous works by sociologists, anthropologists and economists examine the risk profile of a mobile population in the context of the HIV/AIDS epidemic, and especially long-distance truck drivers (e.g. Oruboloye et al, 1993; Rakwar et al, 1999; Gouws and Ramjee, 2002; Ferguson and Morris, 2007) and temporary migrant workers (Meekers, 2000; Adaji Nwokoji and Ajuwon, 2004). For example, Oruboloye et al (1993) study long-distance truck drivers in Nigeria and find that they are more likely to engage in multiple sexual partnerships, including stable partnerships with women who are not commercial sex workers. The same idea applies for temporary labor migrants because they live far from home and they are living in a male environment as in the case of the mine workers in South Africa (Meekers, 2000). Adaji Nwokoji and Ajuwon (2004) study the variation in risk among the Naval personnel according to the time they spend abroad and found that mobility is a significant determinant of risky attitudes since the naval personnel who had been posted abroad are found to have a higher number of sexual partners, to be more likely to have ever had sex with a female sex worker, and to be less likely to have used condoms during their last sexual intercourse with a commercial sex worker than their colleagues who experienced a local transfer. In addition, individuals living close to roads may decide to use them and to move to other regions where the prevalence rate is higher and hence where the likelihood of becoming HIV-infected is higher. By contrast, one might imagine that the prevalence rates are contained and stable in remote areas because the rate of out-migration is low and the contact with people from elsewhere limited. However the simple contact or increase in the opportunities to have sex induced by the presence of a road is not sufficient to increase the risk of infection. The HIV/AIDS epidemic is an infectious disease and not a communicable disease like the flu or the meningitis. In the case of the communicable diseases, the virus is spread over by the simple contact or passage of someone infected. In the context of the HIV/AIDS, differences in risk taking behavior are necessary to make the epidemic enter the area.

Secondly, road infrastructure facilitates the supply of consumer goods and access to markets. The quality of individuals' knowledge about HIV/AIDS epidemic should be greater in proximity to a road as the agents living in accessible areas are reachable by sensitization groups and they have greater access to media. The chance of receiving any sensitization campaign is very low for people living one hundred kilometers from the nearest road. Moreover media circulate information about HIV/AIDS and access to media is lower in remote areas where access to magazines and newspapers is conditioned upon access to any market, and the possibility of watching TV or listening to the radio are conditioned upon access to electricity which is highly dependent on the presence of a road. Concerning access to markets, condoms might be more easily found close to a road than elsewhere.

An empirical question is to examine which of these two effects dominate in the reduced-form relationship between road distance and HIV-infection. However, as said earlier, the increase in the opportunity to have sex resulting from the presence of road infrastructures and human movement may not explain as such the increase in the risk of HIV-infection because the agents have still the possibility to self-protect against the risk of infection, all the more so as one might expect the access to these protective measures facilitated by the road infrastructure. The demand for preventive measures is also at stake. The condom use and the number of lifetime sexual partners are two major determinants of the probability of being HIV-infected. The role of road proximity on the demand for condom use and on the demand for casual sexual partners are not straightforward. Even though one might imagine that the road extends the potential sexual network, the increase in the supply may not be followed by an increase in demand for casual sex given the additional risk of infection.

In terms of public policy, it is worth knowing all the costs and benefits of road infrastructure, and in particular, the costs and benefits of building a road on HIV/AIDS outcomes were not taken into account when the governments or institutions decide to build paved roads. The spread of knowledge and protective measures are additional benefits induced by the presence of road infrastructure, while the cost appears to be a rise in the risk of HIV-infection and a spread of the epidemic from region to region. It is worth estimating the impact of road proximity in order to quantify the increase in risk induced by a road.

3 Empirical strategy

We first establish the relationship between the proximity to a road and the likelihood of being HIV-infected. Next, we examine whether the observed relationship differs according to gender, urban residence, educational attainment and wealth. We analyze the traffic scenario by testing whether the effect is sensitive to the traffic flows. Eventually, we consider whether heterogeneities in the access to protection is more relevant in explaining HIV-risk than heterogeneities in the demand for protection.

The dependent variable in the baseline analysis is the HIV-status elicited by the blood sample collected by Demographic and Health Surveys during the data collection. We regress HIV-status on distance to road and a set of regressors. The distance to the nearest major road is calculated using cartography tools and using the GIS[1] data on the location of the sampled clusters. In the data, sampled clusters aggregate the individuals who live in the same geographical area. This feature of the data motivates the application of panel data models to take into account unobserved heterogeneities across sampled clusters. Consequently, throughout the paper we employ a random effects model except when we perform a bivariate probit model to jointly estimate the choice of sexual partner and the choice of condom use.

3.1 Demographic and Health Surveys

DHS data are collected in several countries across the world using a standardized sampling design and standardized questionnaires that allows cross-country comparisons in terms of health care, maternal and child health. A module about HIV/AIDS

[1] GIS stands for Geographic Information Systems

is included to assess the knowledge, attitudes and practices of the general population. In each country, the sample is selected in two stages. In the first stage, the clusters are selected from a list of enumeration areas from the latest national census (e.g. the 1994 Population and Housing Census in Ethiopia). For every selected cluster, a complete household listing is carried out and from this list, a given number of households are selected. In each selected household, all women age 15-49 years who were either usual residents or visitors present in the household on the night before the survey were eligible to be interviewed in the survey. For the male survey, only a fraction[2] of the sampled households were selected. In this subsample, all men aged 15-54 years [3] were eligible to be interviewed if they were either permanent residents or visitors present in the household on the night before the survey. All women and men living in the households selected for the male questionnaire and eligible for the individual interview were asked to voluntarily give a few drops of blood for HIV testing.

The new generation of the DHS records geographic coordinates of each of the sampled clusters. GIS data is essential for the empirical strategy of this paper because it enables us to locate each cluster on a map and relate it with the existing transportation roads and the national boundaries. For confidentiality issues, up to 2 kilometers of random error in any direction is added to cluster locations in urban areas, and up to 5 kilometers to cluster locations in rural areas. The three ingredients (survey, HIV testing and GIS) exist for a set of countries from which I select Cameroon, Ethiopia, Ghana, Kenya, Malawi and Zimbabwe[4]. In this empirical analysis, we restrict the sample to the usual residents that constitute 97.10% of the total sample because the relation of interest is that between road and behaviors and DHS provide no information about the place where the visitors live. Our final sample contains 86,644 individuals (see Table 1 in the appendix).

3.2 Prevalence and HIV/AIDS-knowledge

The average rate of HIV prevalence is 7.9% over the six countries, the lowest rate is found in Ethiopia at 1.8% and the highest levels are reached in Malawi (12.44%)

[2]all sampled households in Ghana and Zimbabwe; one half in Cameroon, Ethiopia and Kenya; one third in Malawi

[3]15-59 in Cameroon, Ethiopia, Ghana

[4]Other countries are available but present drawbacks. For instance, in Tanzania, the males were not surveyed and in Lesoto, there is no primary road built given the narrowness of the country.

and Zimbabwe (17.9%). Table 4 reports descriptive statistics for the prevalence rate by residence, sex, age cohort, wealth, educational attainment and marital status. Some common trends are emerging across countries. The rate of prevalence is higher among women than among men (9.5% v.s. 5.9%) and this pattern holds for each country. Never married individuals have a lower chance of being HIV-infected than their counterparts, and among the pool of married or previously married respondents, the formerly married group has a much higher rate of HIV-infection than the currently married one. This marital pattern is persistent across countries. One might imagine that the prevalence is low among the never married because they are likely to be the youngest respondents of the sample. The respondents are between 15 and 49 years old for women and 59 years old for men, and we can expect that the younger they are, the more likely they are still virgin and the shorter is their sexual experience[5]. From the simple descriptive statistics, the relationship between wealth and HIV status and between educational attainment and HIV status seem to be non linear. Here respondents are divided into five wealth categories. The category is drawn from a principal component analysis generated at the country level by the data provider and based on durable goods' ownership.

The level of risk taking crucially depends on the level of knowledge people have, because it is hard to argue that not using a condom is deliberated risk taking if the person does not know anything about the risk of contracting HIV and about the available means of prevention. This is a reason why in the context of the HIV/AIDS epidemic it is of particular importance to control for the level of knowledge and acquisition. In the Demographic and Health Surveys, two types of questions are asked to evaluate the knowledge about HIV/AIDS. First, spontaneous answers are required in the open-ended question *"What can a person do (to avoid contracting AIDS)?"*. This question provides insights about the ability of the individual to recover what he has learned and this ability depends on how the individual has integrated the means of prevention into his current behavior. Second, people are asked for prompted responses to statements like *"Can a person get the AIDS virus from mosquito bites?"* or *"Can people reduce their chance of getting the AIDS virus by not having sex at all?"*. Except in Kenya and Malawi, the answers to the open-ended questions are not as spontaneous as they could have been because they are

[5]Note that 18.69% of the total sample have declared to have never had sex while 13.82% of the respondents have reported having their first sexual intercourse before reaching 15.

asked after the prompted questions. The limit of this sequence is to give individuals potential good answers before asking them to list a set of preventive measures to protect against HIV. For this reason, and also because the fact that the respondents do not list one preventive measure does not mean that they do not know it, as it could be an omission, we prefer to use the prompted questions.

We measure the declared HIV/AIDS-knowledge at the individual level using six questions[6]. For each question, we observe whether the individual answers correctly, wrongly or if she does not know. Each answer reveals a given type of information about the individual HIV/AIDS-knowledge. The main concern is to deal with the "don't know" answer. In the empirical literature using survey data, the "don't know" are often recoded as missing values, but by itself, the "don't know" reveals more information than a missing answer because a missing value might result from many different reasons including the fact that the respondent does not want to answer or that the question was not applicable to the particular respondent, or that there was a mistake in the coding. An alternative to recoding the "don't know" as a missing value is to treat it as wrong answer, but we argue that here in the case of the HIV/AIDS-knowledge, it reveals ignorance while the wrong answer reveals misunderstanding or bad knowledge. Moreover treating them equally might be misleading when examining the role of knowledge in the adoption of safe or risky practices. To distinguish between the two and to keep as many information as possible, for each question k, we generate a variable $score_{ik}$ equal to 1 if individual i answers correctly to question k, -1 if wrongly and 0 if she does not know. Each score is summed up to generate an ordinal variable $closedscore6$ which takes values from -6 to +6, +6 being the score of an individual who answers correctly to every question. Even if 96.74% of the sampled respondents report that they have already heard about AIDS, the means of prevention are not widely understood or acquired and misconceptions are persistent. In particular, 24% of the respondents still think that HIV can be transmitted through mosquitos and 19% think that one can not protect against HIV with a condom. On the whole sample, the average score is 3.6

[6] The questions are as follows: 1) *"Can people reduce their chances of getting the AIDS virus by using condom every time they have sex?"*. 2) *"Can people reduce their chances of getting the AIDS virus by having just one partner who is not infected and who has no other partners?"*. 3) *"Can people reduce their chance of getting the AIDS virus by not having sex at all?"*. 4) *"Is it possible for a healthy-looking person to have the AIDS virus?"*. 5) *"can a person get the AIDS virus from mosquito bites?"*. 6) *"Can people get the AIDS virus by sharing food with a person who has AIDS?"*.

but the distribution of scores varies from one country to another. In Ethiopia, the average score is lower than the global average while it is higher among the Kenyan sample. This type of measurement sanctions the false statement more than the ignorance, because for further analysis, we want to distinguish the two situations.

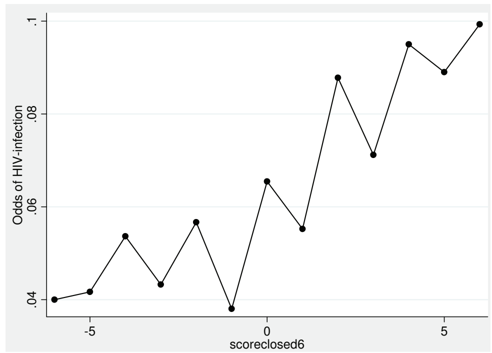

Figure 1: HIV prevalence by knowledge category, total sample

To gain some insights into the relationship between HIV/AIDS-knowledge and the risk of infection, Fig 2 shows the distribution of HIV prevalence by knowledge category. There exists a strong and deep belief that people get infected simply because they are ignorant and have no realization of the disease. Would this belief be supported by the data, one should find that HIV-infection drops as the level of knowledge improves. The distribution of HIV prevalence by knowledge category in Figure 2 suggests that it is much more complex than that.

This chart shows that over the individuals who have the poorest HIV/AIDS-knowledge (i.e. a score equal to -6 or -5), 4% are HIV-infected. At the opposite tail, the highest prevalence level is reached among the agents who hold the highest level of knowledge about HIV transmission and preventive means. About 10% of the respondents who have the highest level of knowledge are HIV-positive. These simple statistics reject the story that HIV-risk is due to ignorance and no realization of the risk of transmission and suggest that the demand for safer sexual practices is of crucial importance. We revisit this issue later in the paper.

3.3 Distance to the road and attributed traffic flows

We use ArcGis to project on a map the sampled clusters and the network of primary roads[7] for each of the six countries of the sample. The data used provides information

[7]The road network comes from the Digital Chart of the World, developed by ESRI.

on the whole network but we restrict the analysis to the paved and usable roads. Thus, when we refer to a road, we mean a primary road. Fig. 1 charts what is obtained for Zimbabwe, mapping both clusters and the primary road network.

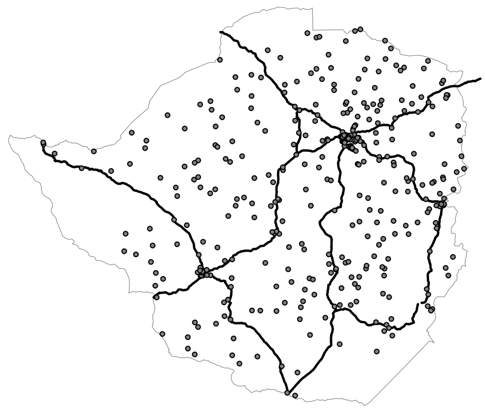

Figure 2: Cluster location and HIV prevalence, realized with 2005/06 Zimbabwe DHS.

Once the clusters and the primary roads are projected, we compute the distance from each cluster to its nearest primary road. The data we obtain is thus at the cluster level, given that the respondents are split into 466 clusters in Cameroon, 529 in Ethiopia, 412 in Ghana, 399 in Kenya, 521 in Malawi and 398 in Zimbabwe. ArcGis provides a measure of the distance expressed in meters. The variable *distroad* is generated as a measure of the distance in kilometers. Over the whole sample, 19% of the respondents (i.e. 16,280 individuals) are living on a primary road, ranging from only 7% in Malawi to at most, 28% in Ethiopia. Table 3 reports the distribution of the sampled individuals according to their proximity to the nearest primary road. Sampled individuals live on average 24.39 kilometers from the nearest primary road for the whole sample, 20 kilometers away in Cameroon and Malawi and up to 31 kilometers away in Zimbabwe. Cameroon and Kenya seem to have the most developed road network compared to the other countries of the sample since 75% of the respondents live less than 27 and 25 kilometers from a major road respectively. But in Kenya, the last 25% of the respondents who live the furthest away from a road are up to 288 kilometers away from their nearest primary road. In Malawi, the furthest distance to a road is 95 kilometers.

For further investigations, a proxy for road traffic is needed and one possibility is to use trade flows. We use the same data source as in Martin et al (2008) that comes from the Correlates of War project (Barbieri et al, 2008). This data source provides annual import and export data in current US dollars for each pair of trade partners. We restrict the trade flows to the flows between neighboring countries over the five years preceding the year of the survey. For each portion of road, we recover which countries it relates to and attribute to each cluster for which this road is the nearest the total amount of trade flows that were transported through it. In the Appendix, details are provided on the way these variables are generated. We do not include the internal trade flows assuming that they are uniformly distributed across the country. In other words, we assume that internal trade flows affect all regions and towns within a country in the same way and hence do not explain why one town is more hit by the HIV/AIDS epidemic than another one.

3.4 Other explanatory variables

Table 2 describes the data. A number of observations can be made. There is a vast majority of women in the sample except in Ghana and Zimbabwe as a consequence of the sample design detailed above. Respondents are 29 years old on average. 33% of the sample are living in urban areas. While over the whole sample, 25% have no formal education, this proportion ranges from 3% in Zimbabwe to 54% in Ethiopia. 38% and 34% have at least some primary and secondary education respectively. As far as religious affiliations, 47% of the respondents are protestant, 18% catholic and 17% muslims. In Zimbabwe, more than two thirds of the respondents are protestants. The highest proportion of muslims is found in Ethiopia, where 33% of the respondents declare themselves to be muslims. Ethiopia has also the particularity of having many people who are orthodox (49% of the sample). 15% of the respondents have previously been tested for HIV. In Ethiopia and Ghana, less than 10% have done so. Understanding the reasons why the proportion of HIV testing is so low is beyond the scope of the paper, but it is worth keeping in mind that the vast majority of the population does not know her HIV status, whether it is positive or negative. On average, about 45% of the respondents know someone who has HIV or who died from AIDS. Among the two most affected countries, Malawi and Zimbabwe, the rate goes to 66% and 30% respectively. In Ghana, a country with low HIV-prevalence, 39% of the respondents report knowing an HIV-infected person, and Kenya has the highest proportion with 74%. These summary statistics are suggestive that this proportion is not linearly positively associated with the national HIV-prevalence.

4 Main results: road proximity and HIV-infection

The baseline equation consists in estimating the probability of being HIV-infected through a random effects probit specification. Denoting by i the index for individual, j for the cluster and c for the country, we estimate the following panel model :

$$HIV_{ij} = c + \beta_1 D_j ldistroad_j + X_i' \delta + \gamma_c + u_{ij}, \forall i, \forall j \tag{1}$$

where c is the constant, D_j is a dummy variable equal to zero or one if the distance of cluster j to its nearest road is equal to zero or positive respectively, $ldistroad_j$ is the logarithm of the distance of cluster j to its nearest road, X_i is the set of individual characteristics, γ_c is the country-specific effect and u_{ij} the disturbances of the model. We use a random effects model in which the composite error term is written as follows:

$$u_{ij} = \alpha_j + \varepsilon_{ij} \tag{2}$$

where α_j is the cluster random effect which captures the unobserved heterogeneity of the clusters and ε_{ij} is the error term. The random effect specification is preferred over a fixed effect one in order to identify the effect of the distance to the road, which is cluster-invariant. Note that the idea of interacting a dummy variable with the logarithm to manage the observations for which the distance to the nearest road is equal to zero comes from Battese and Tessema (1993).

4.1 Primary results

Table 5a presents the results of the random effects probit model, which includes the distance to a road and the additional explanatory variables in the Column 2. These regressions have HIV-status as the dependent variable and always include country dummies on the right-hand side. The coefficients on distance to a road indicate a negative and statistically significant relationship with the likelihood of being HIV-infected.

Magnitude of the effect Empirical findings suggest that increasing the distance to a road by 10% decreases the risk of being HIV-infected by 0.049 percentage point for the whole sample. The magnitude of the impact varies from country to country. The road proximity has a moderated impact on HIV-infection in Ethiopia and Ghana where a 10% increase in the distance leads to a decrease in the risk of 0.011% and

0.017% respectively. The effect is in the average range for Cameroon and Kenya where increasing the distance by 10% induces a fall in the likelihood of infection by 0.035 and 0.040 percentage point respectively. The impact of road infrastructure is found the highest in Malawi and Zimbabwe, the two most infected countries of the sample. Increasing the proximity to a major road by 10% leads to a rise in the risk of infection of 0.084% in Malawi and 0.10% in Zimbabwe. The predicted probability of being infected for someone who lives at less than 10 kilometers away from a road is 8.39% while it is 6.19% for someone who lives at more than 10 kilometers.

Controls We find a significant and positive association between urban residence and the risk of HIV-infection. Living in an urban area might play a role similar to that of road proximity which is to facilitate the contact between people. In addition, the dummy for living in a urban area may be viewed as a proxy for the influence of social norms. Urban individuals may be more able to build extramarital ties with the people they meet because they handle less social pressure with respect to social and familial norms than someone living in a village, and also because urbanization makes people more anonymous, which confers them the possibility to do things in secret, such as having multiple partners. We found no significant association between HIV/AIDS-knowledge and the risk of infection when other variables are controlled for. The marital status is a significant determinant of HIV-infection, as previously suggested by the summary statistics in Table 4. Someone who is currently married or has been previously married has a higher probability of infection than a single individual. The agents who were previously in union are those with the highest probability of being HIV-infected. A deeper analysis could be done to see whether the effect is different for the separated and widows. Religion appears statistically significant in predicting HIV-infection and the muslims are found less likely to get HIV-infected than their counterparts having other religious affiliations except for the protestants.

Sensitivity To test for the sensitivity of the results to some individual characteristics, we reestimate the model separately for men and women, for urban and rural groups, for different educational attainments and wealth levels (see Table 5b). First, the effect of road distance on the likelihood of HIV-infection is negative and statistically significant for both sexes. The effect of road proximity is greater for males than for females. This difference in the size of the effect might be due to gender differences in sexual patterns. In particular, it has already been shown in the literature that men are more likely to have multiple partners than women and that the number of lifetime sexual partners is higher for men than for women. Both elements might explain why proximity to a road has a greater effect on males than on females because if men have the opportunity to have more sex they will probably take it while women do not necessarily do so. Second, the effect of a road on HIV-infection should depend upon the density of the population living there that drives the pressure of social norms and the possibility from anonymity and secrecy. If the road is passing through a dense area, one might guess the effect on individual HIV-status be different than that for someone living closed to a road but in a less populated area. We test for the sensitivity of the effect to social norms by performing separate estimations for rural and urban agents (see Panel B). We found that the effect is greater for rural agents than for urban agents indicating that prior intuition is not validated by the data. It appears that rural agents react more to the presence of road than the urban ones. This finding might be explained by the fact that urban agents do not need the presence of a road to meet people and to have multiple sexual partnership since living in a town or city confers them a higher potential sexual network than rural agents. By contrast, the possibility to have multiple sexual partners for a rural individual appears to be much more dependent upon the presence of a road. Third, the association between road distance and HIV-infection remains negative and significant for every education group, but the magnitude of the effect varies from one group to another (see Panel C). The impact of a road is particularly great for the two extremes: the agents with no formal education and those who went beyond the secondary school level. Note that the effect is found to be lowest for those with a secondary education. Fourth, Panel D shows a negative and significant relationship between distance to a road and HIV-infection for every wealth quintile.

4.2 The traffic scenario

Next, we examine the traffic scenario that suggests that the observed relationship between proximity to a road and HIV-infection is driven by the increased opportunities to have sex with multiple partners induced by the presence of a road. A road increases physical communication since sedentary people who live close to a road potentially get in touch with the mobile population who travels along this road. To test the validity of this traffic scenario, we test whether the effect of road distance on the risk of HIV-infection is sensitive to the use of the road using two complementary estimation strategies.

First, to estimate the role of road traffic in amplifying the effect of road infrastructure on HIV-risk, we interact *logdistance* with one of the measures of traffic flows detailed above in the following extended model:

$$HIV_{ij} = c + \beta_1 D_j ldistroad_j + \beta_2 log(traffic_j) + \\ \beta_3 D_j ldistroad_j) \times log(traffic_j) + X_i'\delta + \gamma_c + \alpha_j + \varepsilon_{ij}, \forall i, \forall j \qquad (3)$$

Table 5c presents the results of the extended model. Column 1 reports the benchmark empirical results. In Column 2, the proxy for the road traffic is specific to each road portion and is equal to the average traffic flows in that road portion in logarithm. The estimated coefficient of distance to road remains negative and significant but increases in magnitude compared to the baseline coefficient. The interaction term turns out to be significantly and positively related to the likelihood of HIV-infection meaning that the reductive effect of road distance declines as the traffic flows increase. To illustrate the mechanism, consider two clusters A and B located two kilometers from a primary road. If you displace the clusters further away from the road, the decrease in the risk of infection will be greater for the individuals living in the cluster whose primary road is used less to transport goods.

In column 3, the national trade openness[8] in logarithms is used as a proxy for traffic flows. This variable is country-invariant and to be able to estimate its effect, country-specific effects are removed. The effect of trade openness is found to be a significant and positive predictor of HIV-infection. Our findings confirm Oster (2009) who investigates the relationship between trade openness and HIV prevalence at the regional level and found that HIV prevalence increases with the level of trade flows both in volume and in value suggesting that the flow of people resulting from trade might increase the risk of infection. The coefficient of the interaction term shows that the effect on trade openness increases with the proximity to the road meaning that the effect of trade openness on HIV-infection is sensitive to the individual's place of residence. Column 4 includes both measures of traffic flows and their respective interaction terms in the right-hand side and confirms previous findings that living in a town where the road traffic is large leads to a higher risk of

[8] The measure of trade openness comes from the Penn World Table (Heston et al, 2006) and stands for the ratio of the total imports and exports over the gross domestic product. The trade openness is averaged over the last decade.

HIV-infection than living near a less busy road.

In summary, we found that the traffic scenario is supported by the data since the relationship between proximity to road and the risk of infection is magnified by the use of the road. Roads facilitate physical communication in such a way that people living in accessible areas get in touch with the mobile population passing through. However the increase in the risk of contamination might be dependent on the prevalence rate of the mobile people following the road. If the mobile agents have a tiny probability of being HIV-infected, the risk to contaminate sedentary people with whom they engage in sexual intercourse is very limited. One concern with our previous estimation is that we do not observe the HIV-prevalence for the people passing through the road portions.

To incorporate this latter element, we propose an original estimation strategy that comes from the fact that bordering areas exhibit great human flows and population mixing for informal trade purposes. Indeed, a large proportion of the bilateral trade is informal and takes place at borders, which implies that bilateral trade flows are tremendously underestimated in the national accounts (Azam, 2007). Grounded on this reality, we identify the effect of the distance of the sampled clusters from the nearest neighboring country, of the prevalence rate on the other side of the frontier and of their interaction terms. Prevalence rates come from UNAIDS (2004) and we use the average of the prevalence rates among the adult population in 2001 and 2003. Applying the same techniques as for the measure of the distance to the nearest road, we generate a continuous variable equal to the distance in kilometers between the cluster and the nearest neighboring country[9]. We expect the effect of living in a bordering area on HIV status to be positive and to increase with HIV-prevalence in the neighboring country. Results displayed in Column 4 show that the prevalence exhibited in the neighboring country has a positive but non significant effect on HIV-infection. The higher the level of HIV prevalence in the neighboring country, the higher is the probability of getting HIV-infected, and this effect decreases with the distance to the border, or alternatively, increases with the proximity to the border.

[9]The layer for borders was developed by ESRI using boundary data from ESRI, AND, and Tele Atlas.

4.3 Identification

The results above indicate that there is a reduced-form relationship between distance to the nearest road and the risk of HIV-infection. One concern is whether this relationship is causal, meaning whether a change in the distance to a road causes a variation in the likelihood of HIV-infection.

Our concern is about the non random placement of the individuals. First-of-all, it is reasonable to assume that road investment does not depend upon the HIV prevalence and that people do not choose to live close to road just to have more sex. The first assertion is reasonable in our context even though it is not generalizable to each and every setting. In fact, disease prevalence, and especially malaria prevalence, was a good predictor of investment in colonial times. However, people might have observable or unobservable characteristics that drive both the choice of residence location and the risk of infection. Let us consider an agent who is born in a remote village and moves to live close to a road. The ability to migrate is not uniform and it might happen that people who have a lower risk aversion are more likely to move than someone who is highly risk averse and this risk aversion influences also HIV-related risk taking behavior.

If we think about the potential determinants of individual placement in the African context, household size and job occupation are good candidates. Firstly, household size might influence the decision to live in a given place because in Africa, the agents often live close to other family members in the extended family sense in order to benefit from their social network. Secondly, job occupation might predict where people live because some jobs are only found in proximity to a road. We use the job categories coded by DHS such that binary variables are generated depending on whether the individual is working in sales, in agriculture (self-employed/employee), in services, in household and domestic jobs, in skilled or unskilled jobs or not currently working. One concern with this instrument could be through income and earnings because wealth is known to be a good predictor of HIV-infection and HIV-related behaviors. Here we use categories for job occupation that are broad enough to have both rich and poor individuals in each and every category. To have consistent estimators, the exclusion restriction that household size and job occupation do not drive HIV-infection for reasons other than through its effect on individual location needs to be valid, which is quite reasonable.

In Table 5d, each column reports the results of a different estimation model.

To correct for the potential endogeneity bias, we use a linear probability model in panel assuming a random effects specification. Columns 1 and 2, we report the results of the baseline equation that is estimated through a random effects probit specification and a random effects linear probability model respectively. The last three columns estimate the likelihood of HIV-infection in a two-stage least squares model in which the disturbances are assumed to include a random effect component. Column 3 instruments the distance to a road with a set of dummy variables for job occupation, Column 4 instruments with household size and Column 5 uses both sets of instrumental variables. We found that the estimated coefficient remains negative and statistically significant when correcting for potential endogeneity bias but it increases in size. This finding is supportive of the fact that the impact of road proximity is not only driven by selection. Indeed, the individuals who decided to live close to the road would have undertaken HIV-related risky behaviors even if they had lived far away from a road. In addition, the results imply that even people who are not risk lovers will be at higher risk of infection while living at proximity to road.

5 Ignorance or deliberated risk taking?

Our main finding suggests that road proximity rises the risk of infection. Now we examine whether this observed relationship between road proximity and HIV-infection is supported by the story of misfortune. What we call the story of misfortune is telling us that the AIDS contamination is due to ignorance and deficiencies in the supply of preventive measures. The alternative story would be that HIV-risk results from deliberated risk taking. We test which story is validated by our data. If the story of misfortune holds, one would find that the level of HIV/AIDS-knowledge and access to condoms increases with the distance to the road and this would explain why the individuals living in remote areas are prevented from getting infected. Alternatively, if this story is rejected by the data one would find the knowledge and access to condoms is facilitated by the presence of road infrastructures and then that the positive relationship between proximity to road and HIV-risk is related to a deficiency in the demand for protection rather than to a deficiency in the supply.

5.1 Effect on the supply of self-protective measures

We estimate the following random effects model:

$$y_{ij} = c + \beta D_j ldistroad_j + X_i'\delta + \gamma_c + \alpha_j + \varepsilon_{ij}, \forall i, \forall j \tag{4}$$

where y_{ij} is the dependent variable, c the constant and X_i a set of individual control variables. D_j $ldistroad_j$, α_j and ε_{ij} are defined as in the estimation of HIV-infection. The dependent variable will be alternatively the measure of HIV/AIDS-knowledge, $scoreclosed6$, or the binary measures of access to condoms. To examine whether the story of misfortune is supported by the data, we test $H_0 : \beta > 0$ against the alternative that proximity to road improves knowledge and facilitates access to condoms, i.e. $H_1 : \beta < 0$.

5.1.1 Effect on HIV/AIDS-knowledge

First, we estimate HIV/AIDS-knowledge through Ordinary Least Squares. The error term, ε_{ij}, is likely to exhibit correlation patterns within the clusters, thus we cluster the robust standard errors at the cluster level to take this into account. Table 6 displays the random effects results. The distance to a primary road facility is negatively and significantly associated with the quality of the knowledge. The further to a road the individual's place, the worse is her level of knowledge. This finding is supportive of the fact that knowledge about HIV-transmission is spread similarly to any other good and especially knowledge goods; the further to the road the individual lives, the weaker is the acquisition.

In the first column, we only control for standard demographic variables (sex, urban residence, education, age). The coefficient for age is found to be positive and significant in the first two equations but it remains low. If, instead of entering age linearly, dummy variables by age cohort are used as in column 3, the effect of age turns out to be non linear. The 15-19 years old are found to be less well-informed than the oldest respondents (40 and older) as suggested in the first two equations, but the agents who are between 20 and 39 are better informed than are the 40 and older. The 25-29 is the best informed age group. Males and educated people exhibit a higher score of knowledge than their counterparts. Columns 2 - 3 add two variables related to the individual's exposure to the AIDS epidemic. First, a variable equal to one if the agent knows someone who has HIV or died from AIDS and zero otherwise captures to what extent this element induces people to become better informed in order to avoid being infected in turn themselves. The results confirm the "model of confrontation" stressed in de Loenzien (2005) and suggest that knowing someone infected increases the adequacy of the knowledge about HIV/AIDS certainly due to the fear of becoming infected that follows after seeing or caring for someone who has developed AIDS symptoms. Second, the likelihood of having ever been tested for HIV is used in order to control for the fact that if the individual has ever been tested, this implies that she has received at least pre-test counseling, and even post- test counseling if she has had her result back. This pre- and post-test counseling is the most customized way of transmitting information about AIDS and about preventive methods and this seems efficient since it turns out that someone who has previously been tested has a significantly better level of knowledge than someone who has never been tested.

The negative effect of the distance to a road on HIV/AIDS-knowledge is a reduced-form effect that might capture both the increased capability for the associations leading sensitization campaigns to reach people and the increased access to media, as sensitization messages are broadcasted through TV, radio, magazines and newspapers. In Column 3, we replicate specification (2) and add dummy variables to control for the fact that the respondents report watching TV, listening to the radio and reading a magazine less than once a week, at least once a week or almost every day. If a remote area is defined as being located further than 10 kilometers from the nearest road, 75% (36%) of the sampled individuals living in remote areas and 43% (19%) in accessible areas report not watching TV (listening radio) at all.

The magnitude of the coefficient for *ldistroad* is reduced by the introduction of these variables suggesting that indeed the coefficient for *ldistroad* incorporates the effect of the access to media on the quality of knowledge. However the negative impact of road distance on knowledge is robust in terms of statistical significance. The use of each of the three media increases significantly the level of HIV/AIDS-knowledge. The more frequently people read newspapers or magazines, listen to the radio and/or watch TV, the better is their knowledge. Since we failed to reject the null hypothesis that the effects of reading magazines or newspapers less than once a week, at least once a week and almost every day are equal, we aggregate these three possibilities in one dummy variable and include the latter in the estimation. Among the three media, the strongest effect is found for newspapers and magazines.

Note that even if educational attainment is controlled for, access to magazines and newspapers is restricted to the literate individuals. The results on radio are of particular interest as radio remains the most democratic media. In terms of country-level heterogeneity, we found that all countries of the sample have a better level of knowledge than Cameroon, the highest scores being for Kenya and Zimbabwe.

Results indicate that spatial inequalities in the knowledge about AIDS persist due to the unequal access to information and technology. People who have access to radio, TV and magazines may receive prevention messages not only from a sensitization campaign group but also from additional sources. This diversified source of information allows the individual to hear different messages and to make her own perceptions of her risk of HIV-contamination. On the contrary, for someone who only hears prevention message from one single source, the limitation comes from the competency, honesty and trustworthiness of that source.

5.1.2 Effect on access to condoms

Second, we complete the analysis of supply of preventive measures by investigating whether road infrastructure facilitates access to (male) condoms and the ability to buy a (male) condom. Two types of questions are used for this analysis. First, we want to estimate the probability of knowing a place where one could purchase a condom. Respondents are asked *"Do you know a place where a person can get condoms?"*. If yes, they are asked to cite all the places they know. We aggregate all the different possibilities in three categories: the public medical sector (e.g. government hospital, government health center), the private medical sector (e.g. pharmacy, private clinic) and the other private sector (e.g. shops). The subsample of respondents who report knowing where to find condoms are asked about their ability to buy it through the question: *"If you wanted to, could you yourself get a condom?"*. We estimate equation (4) through a random effects probit model where the dependent variable, y_{ij}, will be alternatively the likelihood of knowing where to find a condom, the likelihood of citing at least one place from the public medical sector, from the medical or from the non medical private sector where condoms can be found, and eventually, the ability to buy one.

Table 7 reports the empirical results from a random effects probit estimation. The first column suggests that the ability to know of at least one place where one can find a condom decreases with the distance to a road. Increasing the distance to a road by 10% reduces the likelihood of citing one place by 0.11 percentage point on the whole sample and the magnitude of the effect is homogeneous across the countries of the sample. The predicted probability of citing at least one place is 57.54% over the whole sample. This predicted probability varies significantly according to the distance to a road and is equal to (i) 68.5% for someone who lives at less than 5 kilometers away from the nearest road, (ii) 58% for someone who live between 5 and 10 kilometers away from a road, (ii) 52% for someone who live between 10 and 15 kilometers away and (iv) 38% for someone who live at more than 100 kilometers away from a road.

Additional results differ according to the type of places cited by the respondents (see Columns 2 - 4). It turns out that the distance to a road increases the likelihood of citing a public place and reduces the likelihood of citing a private place from either the medical or the non medical sector. One possible interpretation is that public places behave as a substitute for private places. If the private medical sector is limited in remote areas, it might be that public places are more likely and are eventually the only places where someone can find and buy a condom. Increasing the distance to a road by 10% leads to a decrease in the likelihood of citing a place from the medical private sector of 0.11 percentage point and from the non medical private sector by 0.15 percentage point. Column 5 reports the estimates of the ability to get a condom from the subsample of respondents who know at least one place. It is shown that the distance to a road is significantly and negatively associated with the ability to get a condom. This might be driven by the anonymity implied by the presence of a primary road. Cities crossed by a road are places of population mixing that might facilitate the purchase of condoms. Road infrastructure facilitates the ability to buy a condom such that decreasing the distance to a road increases this ability by 0.11 percentage point. From most of these estimations, access to condoms is found to be largely improved by the proximity to road infrastructures.

As far as the control variables are concerned, the empirical results suggest that gender matters since women are less able to find and buy condoms than males. This result is not surprising in the sense that males are often in charge of the purchase of condoms as they have a higher bargaining power as far as sexual affairs and condom use in particular are concerned. Higher educational attainment is associated with increased ability to find a condom. The access to condoms is found to be significantly and positively related to the HIV/AIDS-knowledge, suggesting that sensitization campaigns might tell people where to find protective measures. Moreover the level of knowledge increases the ability to buy a condom suggesting that stigma are also cleared up during sensitization campaigns. The relation between wealth and access to condoms is significant. All groups are less able than the richest one to know where to find a condom and to dare to buy one. Muslims are in general less likely to find a condom than people from other religious groups.

Our findings are suggestive that spatial inequalities are persistent in terms of access to preventive measures. However spatial inequalities do not drive the association between HIV-infection and road proximity since this section shows that a road reduces the cost of prevention as it makes condoms more-readily available and people aware of the risk of infection and aware of the preventive methods. HIV-infection is not a result of deficiencies in the supply of protection, the mechanism must be found in the demand for protection.

5.2 Mechanism: Lower condom use or more sex

The last point is to examine which behavioral mechanism is driving the observed relationship between proximity to road and HIV-infection. We investigate whether agents living in proximity to a road adopt riskier sexual behaviors than their counterparts living in remote areas. In particular, this section is an attempt to state whether the positive relation between proximity to a road and HIV-infection is due to a decrease in condom use, an increase in sex, or both.

5.2.1 Revealed preferences

The individual choice is formalized as follows. Denote by y^*_{i1} and y^*_{i2} the indirect utility of individual i when he decides whether to have sex with a usual partner and whether to use a condom respectively. We do not observe their indirect utility but we are able to observe their decision to do so. We assume that a rational agent chooses to have sex with a usual partner or to use condom if his indirect utility to do so is positive. Formally, we have:

$$y_{i1} = \begin{cases} 1 & \text{if } y^*_{i1} > 0 \\ 0 & \text{if } y^*_{i1} \leq 0 \end{cases}$$

$$y_{i2} = \begin{cases} 1 & \text{if } y^*_{i2} > 0 \\ 0 & \text{if } y^*_{i2} \leq 0 \end{cases}$$

where $y_{i1}^* = x_{i1}^{'}\beta + \epsilon_{i1}$, $y_{i2}^* = x_{i2}^{'}\beta + \epsilon_{i2}$. It is assumed that ϵ_{i1} and ϵ_{i2} are joint normal with means equal to zero, variances of one, and a correlation ρ and that both sets of explanatory variables are identical (i.e. $x_{i1} = x_{i2}$) except that we also include a dummy variable equal to one if the individual knows at least one place where one can find a condom and zero otherwise in the estimation of condom use. A bivariate probit specification is performed to describe the joint probability of using a condom and of having sex with a usual sexual partner. Note that this specification does not allow for taking into account the unobserved heterogeneity among clusters anymore. The error term is likely to exhibit correlation patterns within the clusters, thus we cluster the robust standard errors at the cluster level to take this into account.

We perform a bivariate probit specification as in Kazianga (2005) for two reasons. First, the choice of partner and the choice of condom use are not independent. The condom use depends upon the partner with whom the individual is having sex. When having sex with one's spouse, other elements such as the desire for a child enter the choice of not using a condom. Second, estimating the individual choice of condom use in a standard probit specification does not allow for capturing the level of HIV-related risk taking since the riskiness directly depends on the type of partner. The risk of not using a condom is reduced when the agent has sexual intercourse with her spouse while it is rather high when the sexual partner is a commercial sex worker. Accordingly it is worth estimating the choice of condom use jointly with the estimation of the choice of partner. The choice of this specification is validated by the fact that the coefficient of correlation between the error terms in the two equations is statistically different from 0. In the Demographic and Health Surveys, respondents are asked about the nature of the relation they have with their last sexual partner (e.g. spouse, casual acquaintance, relative, commercial sex worker) and whether they have used a condom during their last sexual intercourse. 12% of the sample report condom use and 80% report that the last intercourse partner was the spouse or cohabiting partner. Similar information are provided for the second to last sexual intercourse but the data are poorly reported.

5.2.2 Empirical results

First, considering the choice of last sexual partner, empirical findings suggest that the relation between the distance to a road and the probability that the last sexual partner was one's spouse is significant and positive (see Table 8a). Data supports the argument that living in accessible areas increases the opportunities to have sex with casual partners. This is induced by the fact that living in areas crossed by roads brings people in contact with mobile people and makes them more easily move to other places, and in turn increases the perimeter of the sexual network. Once the opportunity to have sex increases, people decide to use it or not. Here results are supportive of the fact that people prefer to have more sex if proximity to road makes it possible. When adding a variable to control for road traffic, we found that road traffic significantly increases the likelihood to have had the last intercourse with a casual partner[10]. Catholics and those affiliated to the category "other religions" are the religious groups the most likely to have had the last sexual intercourse with a casual partner. Wealth does not appear to play a significant role in the choice of the last sexual partner for most income categories except for the poorer group who are more likely than the richest one to have had their last intercourse with their spouse. We found out that the better informed people are in terms of HIV transmission, the less likely they are to have had their last sexual intercourse with their spouse. This suggests that multiple and casual partnerships are more prevalent among people who know the risk they undertake by choosing to do so. Females are found to be more faithful than males since being a woman increases the probability of having one's last sex with one's spouse. Those in Cameroon are the least likely to have had their last sexual intercourse with their spouse. The faithfulness is more commonplace in Ethiopia than everywhere else.

Second, considering the choice of condom use during the last sexual intercourse, it is found that road infrastructure facilitates condom use (see Column 2 in Table 8). This means that the additional risk of getting infected by having multiple partners is somehow mitigated by the rise in condom use. These findings suggest that people are choosing the preventive methods that fit best their ways of life and that hurt their utility the least. It appears here that condom use is preferred to faithfulness, which is the prevention measure that consists in knowing one's partner and one's partner's sexual life, contradicting strong beliefs that Africans are reluctant to use condoms. The coefficients for the other explanatory variables have the expected signs. The relation between age and condom use is significant and negative, meaning that condoms are more readily used by the young cohorts than their elders. The higher the educational attainment and the higher the wealth level, the more likely

[10] Results are not reported here.

people are to use condoms. The link between wealth and demand for condoms might be explained either by the increased ability to purchase condoms or by the increased incentives wealthy agents have to invest in health. The condom is more likely to be used by Catholics compared to the other religious groups. Improving knowledge about HIV transmission increases the probability that a condom was used during the last intercourse. HIV/AIDS-knowledge appears as a pre-requisite for the condom use since the level of knowledge is positively related to the likelihood of using a condom during the last sexual intercourse. Heterogeneities are found across the countries. In all countries, people are less likely to use condoms than Cameroonian people and it is in Ethiopia where the recourse to condoms is the least commonplace.

Third, Table 8b reports the four joint probabilities. We found that the probability of having sex with one's spouse and using a condom is significantly and negatively related to the distance to the road infrastructure. People living close to a primary road have a higher probability of doing so than people living in remote areas. Column 2 provides the marginal effect of the distance to the road on the probability of having sex with one's spouse and not using a condom, which turns out to be significant and positive. The last two columns deal with extramarital sexual relations and show that the probability of having sex with a casual partner and using a condom or not decreases with the distance to a road. Accordingly, it appears from this estimation that proximity to a road increases the likelihood of having casual partners. One might explain this relation by the extended sexual network resulting from road facilities or by the fact that it might be easier to hide such a casual relationship in places where human movement is high. In line with the results from Table 8, women are found to be less likely to engage in extramarital sex than men and even less likely without a condom. The likelihood of engaging in extramarital sex increases with urban residence, suggesting that the relaxed constraints of social norms and anonymity induced by large cities facilitate the relations outside marriage. Of particular interest are the regression results from the last column that give the determinants of the riskiest HIV-related behavior. We found that having sex with a casual partner and without a condom is less likely for people living far from a major road, and also for older individuals and for non educated or highly educated people. HIV/AIDS-knowledge plays a counter-intuitive role since it turns out that the likelihood of engaging in risky sex increases with the level of knowledge.

The behavioral analysis suggests that road proximity has two competing effects: on one hand, it increases the likelihood of using a condom and on the other hand, it increases the likelihood of having sex with a casual partner. Findings are robust to the 2SLS estimation using the same set of instrumental variables as in the estimation of HIV-infection. We found that access to condoms and to information increase the demand for condom and the agents seem to choose the preventive measure that hurts their utility the least. Individuals reveal preferences for having more sexual partners even with a condom. This suggests that as condoms become available, people tend to use them but increase or maintain their demand for casual sexual partners. This finding is related to the literature on risk compensation about road safety. Previous works in this literature have shown that when road safety devices became compulsory the occurrence of road traffic accidents did not decrease as much as it was expected because people adjusted their behavior to the fall in the probability of accident and in the probability of having a mortal accident induced by the seat belts by driving faster (Peltzman, 1975; Evans and Graham, 1991; Peterson et al, 1994; Sen and Mizzen, 2007).

6 Robustness checks

6.1 The measure of knowledge

One concern about the validity of our results about individual knowledge might be the way we measure the level of knowledge and especially the fact that we distinguish among the lack of knowledge the fact of ignorance from the fact of having false knowledge. To check for the robustness of the association between distance to a road and the level of knowledge, we measure knowledge in three different ways. First, we use the same idea as above with the exception that we attribute a score equal to 0 in case of false statement instead of -1. The second measurement relies on the principal component analysis method to generate a score of knowledge based on the six initial variables. The third measurement is a binary variable equal to one if the individual gives the right answer for every question and zero otherwise, what DHS calls the "comprehensive knowledge" in its reports. For each measurement used to estimate the effect of road distance, the results are qualitatively the same. The level of knowledge is decreasing with the distance to the road. These results[11] are suggestive that the definition of our variable for knowledge about HIV-transmission we adopt throughout the paper is not a major source of bias in our estimates.

[11] Data description and output tables are available from the author upon request.

6.2 Sensitivity analysis

In Table 9, we perform a sensitivity analysis on each estimation of the paper to see whether the previous findings are sensitive to changes in the sample. we re-estimate each estimation by removing one country at a time from the sample. For the sake of exposition, we report only the coefficient of the variable of interest, the distance to a road, in log. The sensitivity analysis performed on HIV-infection, access to condoms and sexual behaviors stresses that the previous results are robust to the sample used (Panels B, C and D). The negative association between the distance to a road and HIV/AIDS-knowledge is robust to changes in the sample in terms of statistical significance and in terms of magnitude. One exception is the exclusion of Malawi in re-estimating the equation of Column 3, suggesting that once controlling for the access to mass media, the effect of road distance disappears (see Panel A).

6.3 Country-by-country

Of particular interest is to see whether our findings hold for each particular country of the sample. We replicate the equations country-by-country to be able to draw specific policy recommendations. The estimated effects of road distance are reported in Table 10.

HIV-infection Three groups of countries emerge from the country-by-country analysis of the effect of road distance on the risk of being HIV-infected. When estimating the baseline equation, the negative association between distance and HIV-infection is statistically significant in Ethiopia, Ghana, Malawi and Zimbabwe, the highest effect being found in Malawi. In Cameroon and Kenya, geographical disparities in terms of access to road and markets as a large do not appear to be a good predictor of HIV-status. Later when controlling for the potential traffic flows being carried through each particular road portion, the effect of road distance is found to lose its predictive power in Ethiopia and Malawi, while it remains negative and statistically significant in Ghana and Zimbabwe.

Knowledge A country-by-country analysis is suggestive that the role of proximity to a road on HIV/AIDS-knowledge is not homogeneous across the countries of the sample (see Panel A). Three groups of countries emerge. First, in Kenya and Malawi, among the most infected countries of our sample, the effect of road distance on knowledge is found to be high and significant, suggesting that spatial inequalities in the access to information about HIV/AIDS remain present. In this category, the two countries have opposite patterns in terms of road density. Second, in Cameroon and Ghana, the effect of road proximity is significant in the first two specifications but loses its predictive power as soon as media are controlled for. This suggests that geographical disparities in the access to media drive the positive relation between road proximity and the quality of knowledge in these two countries. Third, in Ethiopia and Zimbabwe, the quality of knowledge is not statistically different for people living in areas crossed by roads and those living in remote areas. It seems that both countries have succeeded in spreading the information about HIV/AIDS even in remote areas.

Access to condoms The empirical country-by-country analysis of Panel B suggests that spatial inequalities in the access to condoms vary across countries and might result from heterogeneity in the supply of public and private medical services that we are unable to control for in this paper. What is supported by the data is that people living far from a primary road have a higher probability of knowing a place from the public medical sector where a condom can be found in Ethiopia, Ghana, Kenya and Zimbabwe. A private medical place is less likely to be known by people living in remote areas compared to their counterparts living in accessible areas in Cameroon, Ethiopia, Kenya and Malawi. The negative and significant relation between knowing a place from the non medical private sector and distance to the road holds for all countries of the sample except Malawi. The distance to a road decreases the ability to get condoms in Cameroon and Malawi. One possible reason why the relation is significant there and non significant in the other countries might be that taboos are deeper in the countryside in these two countries than in the others.

Sexual behaviors Panel D shows that the role of roads on partner choice and condom use is found only statistically significant in Cameroon, in Malawi (for partner choice) and in Cameroon (for condom use).

7 Conclusion

In this paper, using individual survey data from Demographic and Health Surveys and geographical data on road infrastructure, we have analyzed the effect of proximity to a road on the risk of HIV-infection. Although the contribution of the paper is purely empirical, we have suggested a mechanism by which this observed relationship might hold. The empirical results indicate that the risk of being HIV-infected decreases with the distance to the nearest major road, suggesting that living far from physical communication means takes people away from the risk of infection. The traffic scenario has been validated by our empirical analysis meaning that the observed effect of a road on HIV-risk is sensitive to the human flows and population mixing induced by the presence of roads. A road that is not used at all does not bring any additional risk of infection compared to a setting in which the road is very far from where the agents live. However even if this analysis provides insights into the relationship between road proximity and infection, it does not explain why people get infected since self-preventive measures exist and people living in accessible areas may have decided to use them to reduce the probability of being infected. The increased opportunity to have sex can not explain as such the observed relationship.

Considering the supply of preventive measures, we show that proximity to a road plays a strong role in improving HIV/AIDS-knowledge and in facilitating access to condoms and the ability to buy condoms. The fact that proximity to a road increases the risk of infection even if it also increases the access to protection (and hence reduces the cost of protection) is inconsistent with the story that ignorance and lack of access to preventive measures are driving the spread of HIV in Africa. Empirical results reject the story of ignorance and misfortune as unique determinant of HIV-infection. People living at proximity to a major road are found to be more likely to be contaminated and at the same time more likely to have a comprehensive knowledge about HIV-transmission and have a better access to preventive measures. Thus the results support the story that the incentives to invest in health remain too low in these countries of analysis and that the demand for risky sex depends upon the agent's place of residence. Considering the spatial variation in individual behavior, we show that condom use and multiple sexual partnerships are more likely in accessible areas. This finding reveals two important things. The first implication is that access to condoms and to information about the importance of using them have facilitated their use. The second implication is that people express their preferences towards the set of available preventive measures and choose the one that hurts their utility the least. The agents living close to a road are found to be more likely to prefer to use a condom and have multiple partnerships than their counterparts living in remote areas.

Should policy implications have to be drawn from our empirical analysis, we would say the following. First, this paper does not promote autarky nor the freeze of every investment in road infrastructure. It is rather an attempt to show that prevention efforts need be reinforced in proximity to roads. Second, from our results, we are able to state that exerting more efforts on sensitization or subsidizing condoms would not be efficient in inducing people to change their behaviors and to adopt safer sexual practices. By contrast, prevention policies that provide people with more incentives to take care of their health status may help in reducing the risk of infection. Third, the choice of sexual practices is a joint decision and if sedentary people take advantage of living close to a road to have sex with mobile population, this population should also benefit from particular programs of prevention.

References

[1] Adaji Nwokoji, U., Ajuwon, A.J., 2004. Knowledge of AIDS and HIV Risk-related sexual Behavior among Nigerian Naval Personnel. BMC Public Health 4 (24).

[2] Azam, J.P., 2007. Trade, Exchange Rate and Growth in Sub-Saharan Africa. Cambridge University Press.

[3] Banerjee, A., Duflo, E., Qian, N., 2009. On the Road: Access to Transportation Infrastructure and Economic Growth in China. Working Paper.

[4] Barbieri, K., Keshk, O., Pollins, B., 2008. Correlates of War Project Trade Data Set Codebook, Version 2.0. Online: http://correlatesofwar.org.

[5] Battese, G.E., Tessema, G.A., 1993. Estimation of stochastic frontier production functions with time-varying parameters and technical efficiencies using panel data from Indian villages. Agricultural Economics 9, 313-33.

[6] de Loenzien, M., 2005. Connaissances et attitudes face au VIH/SIDA. L'Harmattan, Collection Populations.

[7] Digital Chart of the World, http://www.maproom.psu.edu/dcw/.

[8] Donalson, D., 2008. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. Mimeo.

[9] Evans, W., Graham, J., 1991. Risk reduction or risk compensation? The case of mandatory safety belt use laws. Journal of Risk and Uncertainty, 4, 61-73.

[10] Fay, M., Leipziger, D., Wodon, Q., Yepes, T., 2005. Achieving Child-Health Related Millenium Development Goals: The Role of Infrastructure. World Development 33 (8), 1267-1284.

[11] Ferguson, A.G., Morris, C.N., 2007. Sexual and Treatment-Seeking Behaviour for Sexually Transmitted Infection in Long-Distance Transport Workers in East Africa. Sexuall Transmitted Infections 83, 242-245.

[12] Gouws, E., Ramjee, G., 2002. Prevalence of HIV Among Truck Drivers Visiting Sex Workers in KwaZulu-Natal, South Africa. Sexually Transmitted Diseases 29 (1), 44-49.

[13] Heston, A., Summers, R., Aten, B., 2006. *Penn World Table Version 6.2*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. September 2006.

[14] Huillery, E., 2009. History Matters: The Long Term Impact of Colonial Public Investments in French West Africa. American Economic Journal Applied Economics, 1(2), 176-215.

[15] International mapping. On line: $www.internationalmapping.com/index/Goodies_Africa.html$.

[16] Jacoby, H.G., 2000. Access to Markets and the Benefits of Rural Roads. The Economic Journal 110 (465), 713-737.

[17] Kazianga, H., 2005. HIV/AIDS Prevalence and the Demand for Safe Sexual Behavior: Evidence from West Africa. Unpublished Manuscript.

of Burkina Faso. AIDS Behaviour 10, 607-17.

[18] Martin, P., Mayer, T., Thoenig, M., 2008. Make Trade Not War. Review of Economic Studies 75, 865-900.

[19] Meekers, D., 2000. Going Underground and Going After Women: Trends in Sexual Risk Behaviour among Gold Miners in South Africa. International Journal of STD and AIDS 11, 21-26.

[20] Oruboloye IO, Caldwell P and Caldwell JC. 1993. The role of high-risk occupations in the spread of AIDS: truck drivers and itinerant market women in Nigeria. Inter Fam Plann Persp 19, 43-48.

[21] Oster, E., 2009. Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa, Working Paper

[22] Peltzman, S., 1975. The Effects of Automobile Safety Regulation. Journal of Political Economy, 83 (4), 677-726.

[23] Peterson, S., Hoffer, G., Millner, E., 1994. Are Drivers of Airbag-Equipped Cars More Aggressive? A test of the offsetting behaviour hypothesis. Journal of Law and Economics, 38, 251-265.

[24] Rakwar, J., Lavreys, L., Thompson, M.L., Jackson, D., Bwayo, J., Hassanali, S., Mandaliya, K., Ndlinya-Achola, J., Kreiss, J., 1999. Cofactors for the Acquisition of HIV-1 among Heterosexual Men: Prospective Cohort Study of Trucking Company Workers in Kenya. AIDS 13, 607-614.

[25] Sen, A., Mizzen, B., 2007. Estimating the Impact of Seat-Belt Use on Road Traffic Fatalities: Empirical Evidence from Canada. Canadian Public Policy, 33(3), 315-335.

[26] Straub, S., Vellutini, C., Warlters, M., 2008. Infrastructures and Economic Growth in East Asia. Policy Research Working Paper n4589. World Bank.

[27] Tanser, F., LeSueur, D., Solarsh, G., Wilkinson, D., 2000. HIV Heterogeneity and Proximity of Homesteads to Roads in Rural South Africa: An Exploration using a Geographical Information System. Tropical Medecine and International Health 5 (1), 40-46.

[28] UNAIDS, 2004. 2004 Report on the Global AIDS Epidemic. 4th global report.

APPENDIX A: Distance of the clusters to the primary roads

Data source We use ArcEditor to compute the distance between each cluster and the nearest primary road. Two sources, labelled "Digital Chart of the World", are used for the primary roads. One is from Harvard Geospatial Library and provides the road network in a shapefile. The limitation of this first source is that includes all types of roads without distinguishing the primary roads from the secondary roads or even the trails. The other source used comes from the software provider, ESRI, and gives the primary roads in a layer format. With a layer file, it is not possible to compute distance, hence we select all the small portions of the Harvard Geospatial Library shapefile that correspond to the primary roads of the second source by hand to generate a new shapefile that will contain only the primary roads. This shapefile is linked to the DHS GIS data to map the clusters in the map and to compute distances.

The shapefile for the borders comes from International mapping. We generate one file corresponding to the border between Zimbabwe and Mozambique, one for Zimbabwe and Zambia and so on for every country that shares a common border with Zimbabwe. Later, for each sampled cluster, we compute the distance between the cluster and each border, and afterwards, we take the distance to the nearest country.

Proxy for mobility Consider the case of Zimbabwe (see Figure above). We follow a stepwise procedure in which first, we divide the primary road network in several road portions such that each portion corresponds to a trade route (e.g. Zimbabwe-South Africa, Zimbabwe-Botswana). In the case of Zimbabwe, two portions are central and clearly link to every neighboring country and thus are considered as a single network. Second, we compute the distance from the cluster to each road portion generated. Third, we generate a variable equal to the distance to the nearest road portion and then we attribute a value of traffic flows that is a function of the road portions. If it turns out that the cluster a is close to the road portion that related Zimbabwe to South Africa, the value of traffic flows for all the respondents of this cluster a is equal to the average bilateral trade flows between the two countries during the five years preceding the survey. If the cluster is located close to the road portion that might be used to circulate the goods for every bilateral relationship, the traffic variable is the sum of each bilateral average trade flow.

For the prevalence rate in the neighboring country, we proceed in the same manner. we attribute the value of the prevalence rate in the nearest neighboring country once we have computed all the distances from the cluster to the borders.

Table 1: Sample size

	Obs.	Women	nb clusters	nb households	nb hh/hiv†
Cameroon	14,927	9,940	466	10,462	5,319
Ethiopia	19,456	13,628	529	13,721	6,689
Ghana	10,570	5,607	412	6,251	6,251
Kenya	11,360	7,891	399	8,561	8,561
Malawi	14,679	11,503	521	13,644	13,644
Zimbabwe	15,652	8,664	398	9,285	9,285
Total	86,644	57,233	2,725	61,924	36,358

† the number of households in the subsample eligible for the blood sample collection is reported in this column.

Table 2: Summary statistics

Variable	**All**	**CMR**	**ETH**	**GHA**	**KEN**	**MWI**	**ZWE**
HIV+	.0787	.0539	.0184	.0213	.0650	.1244	.1788
	[.2693]	[.2258]	[.1344]	[.1444]	[.2466]	[.3301]	[.3832]
HIV testing	.1536	.1907	.0743	.0921	.1563	.1488	.2204
	[.3606]	[.3929]	[.2623]	[.2891]	[.3631]	[.3559]	[.4145]
know s^{one} HIV+	.4549	.4468	.1390	.3862	.7402	.6567	.2975
	[.5018]	[.4972]	[.3799]	[.4869]	[.4385]	[.4748]	[.4572]
scoreclosed6	3.5664	3.2904	2.9479	3.4820	4.5207	3.4346	4.1088
	[2.3278]	[2.2833]	[2.5567]	[2.3932]	[1.7328]	[2.2837]	[2.1073]
women	.6606	.6659	.7005	.5305	.6946	.7836	.5535
	[.4735]	[.4717]	[.4581]	[.4991]	[.4606]	[.4118]	[.4971]
age	28.6032	28.5550	28.7690	30.2568	28.5342	28.1460	27.8055
	[10.1624]	[10.4616]	[10.2714]	[10.9101]	[9.8735]	[9.4432]	[9.9358]
urban	.3332	.4993	.2994	.3988	.3312	.1422	.3530
	[.4714]	[.5000]	[.4580]	[.4897]	[.4707]	[.3493]	[.4779]
noeducation	.2462	.1755	.5440	.2851	.1370	.2087	.0316
	[.4308]	[.3804]	[.4981]	[.4515]	[.3438]	[.4064]	[.1750]
primary educ	.3795	.3997	.2439	.1848	.5313	.6271	.3181
	[.4853]	[.4898]	[.4295]	[.3881]	[.4990]	[.4836]	[.4658]
secondary educ	.3378	.3935	.1815	.4898	.2501	.1561	.6101
	[.4730]	[.4885]	[.3854]	[.4999]	[.4331]	[.3630]	[.4877]
higher educ	.0365	.0313	.0305	.0404	.0816	.0081	.0401
	[.1875]	[.1741]	[.1720]	[.1969]	[.2738]	[.0897]	[.1963]
catholic	.1753	.3892	.0099	.1587	.2383	.2182	.1029
	[.3803]	[.4876]	[.0994]	[.3654]	[.4260]	[.4130]	[.3038]
protestant	.4670	.3539	.1533	.5353	.6044	.6158	.6790
	[.4989]	[.4782]	[.3604]	[.4988]	[.4890]	[.4864]	[.4669]
muslim	.1696	.1720	.3269	.1941	.1209	.1518	.0075
	[.3753]	[.3774]	[.4691]	[.3955]	[.3260]	[.3589]	[.0865]

Note: Standard deviations are in brackets.

Table 3: Distance to the nearest primary road in kilometers

Country	**Mean**	**Std dev.**	**p25**	**p50**	**p75**	**Max**
All	24.3931	32.5844	2.6582	10.9791	35.6629	287.8292
Cameroon	19.6884	28.2367	2.6582	5.3163	26.7141	160.0642
Ethiopia	26.9754	35.5261	0	11.8218	39.0099	192.2833
Ghana	25.0190	32.7205	1.7829	10.0855	36.1007	176.7575
Kenya	22.2619	37.6028	0	10.2895	24.7803	287.8292
Malawi	19.9984	22.0936	2.7278	10.3512	32.2064	94.88632
Zimbabwe	30.8791	35.3058	2.7647	13.8236	50.8287	172.0802

Table 4: HIV prevalence by characteristics and by country

Variable	**All**	**CMR**	**ETH**	**GHA**	**KEN**	**MWI**	**ZWE**
Sex							
Female	.0954	.0676	.0229	.0257	.0830	.1459	.2071
Male	.0589	.0401	.0132	.0159	.0450	.0988	.1407
Education							
No education	.0340	.0354	.0111	.0174	.0307	.1430	.2184
Primary	.0907	.0567	.0173	.0254	.0722	.1182	.1947
Secondary	.0947	.0586	.0412	.0227	.0688	.1318	.1695
Higher	.0645	.0485	.0237	.0120	.0586	.0455	.1542
Wealth							
Poorest	.0599	.0288	.0078	.0135	.0326	.1000	.1690
Poorer	.0701	.0326	.0063	.0213	.0610	.0779	.1763
Middle	.0794	.0593	.0082	.0308	.0486	.1277	.1696
Richer	.0997	.0711	.0067	.0225	.0672	.1436	.2176
Richest	.0825	.0673	.0421	.0211	.0968	.1692	.1559
Urban							
Rural	.0730	.0429	.0079	.0197	.0519	.1147	.1745
Urban	.0906	.0654	.0498	.0239	.0963	.1863	.1880
Age cohort							
15-19	.0194	.0132	.0055	.0035	.0197	.0206	.0425
20-24	.0633	.0533	.0175	.0114	.0586	.1057	.1175
25-29	.1015	.0803	.0258	.0228	.0968	.1435	.2271
30-39	.1275	.0846	.0277	.0360	.0946	.1872	.3177
40-49	.0868	.0474	.0167	.0275	.0628	.1531	.2435
Marital Status							
Single	.0291	.0259	.0102	.0065	.0285	.0276	.0607
Married	.0849	.0583	.0170	.0252	.0715	.1365	.2031
Formerly Married	.2229	.1198	.0630	.0556	.1914	.2796	.4559
Total	.0787	.0539	.0184	.0213	.0650	.1244	.1788

Table 5a: Determinants of HIV-infection
Random Effects panel probit estimates

	(1)		(2)	
	hivpos		hivpos	
Dldistroad	-0.0818***	(0.008)	-0.0548***	(0.009)
equal 1 for woman			0.2176***	(0.021)
married			0.5079***	(0.030)
prev. married			1.0622***	(0.037)
age			0.0116***	(0.001)
urban			0.0718**	(0.033)
primary educ			0.2017***	(0.035)
secondary educ			0.2202***	(0.038)
higher educ			0.0120	(0.064)
wpoorest			-0.1659***	(0.043)
wpoorer			-0.1316***	(0.040)
wmiddle			-0.0558	(0.037)
wricher			0.0018	(0.031)
catholic			0.0769*	(0.044)
protestant			0.0521	(0.040)
other religion			0.2104***	(0.045)
scoreclosed6			0.0019	(0.005)
Ethiopia	-0.5080***	(0.045)	-0.4999***	(0.053)
Ghana	-0.4128***	(0.045)	-0.4122***	(0.047)
Kenya	0.1040**	(0.043)	0.1651***	(0.047)
Malawi	0.4890***	(0.040)	0.5235***	(0.044)
Zimbabwe	0.7962***	(0.036)	0.7768***	(0.039)
Constant	-1.5348***	(0.032)	-2.7613***	(0.078)
Observations	53,405		50,830	
Number of clusters	2,704		2,702	

The reported coefficients are the average marginal effects of the random effects probit model. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5b: Separated effect of distance to a road on infection

Panel A: Gender			
	coefficient	std error	obs.
female	-0.0453***	(0.0105)	27,125
male	-0.0577***	(0.0132)	23,705

Panel B: Urban			
	coefficient	std error	obs.
yes	-0.0427***	(0.0144)	17,151
no	-0.0575***	(0.0125)	33,679

Panel C: Educational attainment			
	coefficient	std error	obs.
no education	-0.0795***	(0.0228)	10,321
primary education	-0.0488***	(0.0136)	18,340
secondary education	-0.0348***	(0.0130)	20,270
higher education	-0.0851*	(0.0439)	1,899

Panel D: Wealth quintile			
	coefficient	std error	obs.
poorest	-0.0582**	(0.0263)	9,132
poorer	-0.0701***	(0.0242)	9,281
middle	-0.0602***	(0.0197)	9,753
richer	-0.0607***	(0.0180)	10,296
richest	-0.0273*	(0.0165)	12,368

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All probit estimations control for country-specific effects.

Random effects specification in the error terms.

Table 5c: Effect of distance to a road on infection- Mechanism test: traffic scenario

	(1)	(2)	(3)	(4)	(5)
Dldistroad	-0.0548***	-0.1721***	-0.5012***	-0.5659***	
	(0.009)	(0.028)	(0.129)	(0.129)	
ltraffic		-0.0287*		0.0271	
		(0.015)		(0.018)	
$Dldistroad \times ltraffic$		0.0237***		0.0277***	
		(0.005)		(0.007)	
lopenness			0.1906**	0.1793**	
			(0.084)	(0.085)	
$Dldistroad \times lopen$			0.1105***	0.0926***	
			(0.032)	(0.032)	
Dldistborder					0.0443
					(0.041)
lHIVprevneig					0.1347
					(0.087)
$Dldistborder * lHIVprevneig$					-0.0212
					(0.018)
Observations	50,830	50,830	50,830	50,830	50,869
Number of clusters	2,702	2,702	2,702	2,702	2,704
Controls	YES	YES	YES	YES	YES
Country FE	YES	YES	NO	NO	YES
Cluster RE	YES	YES	YES	YES	YES

Coefficient from a random effects probit model are reported. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The omitted dummies are male, single, rural, no education, richest, muslim, Cameroon.

Table 5d: Effect of distance to a road on infection, instrumented by job occupation and/or household size

	(1)	(2)	(3)	(4)	(5)
Dldistroad	-0.0548***	-0.0066***	-0.0972***	-0.5024***	-0.1386***
	(0.009)	(0.001)	(0.019)	(0.094)	(0.018)
Country effects	YES	YES	YES	YES	YES
Cluster RE	YES	YES	YES	YES	YES
Estimation	Probit	OLS	2SLS	2SLS	2SLS
Instruments					
job occupation			YES	NO	YES
hh size			NO	YES	YES
Observations	50,830	50,830	50,605	50,830	50,605

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Country-specific are included in each estimation.

Table 6: Knowledge about HIV-transmission
Random Effects model estimates

	(1)		(2)		(3)	
	scoreclosed6		scoreclosed6		scoreclosed6	
Dldistroad	-0.0972***	(0.011)	-0.0905***	(0.011)	-0.0471***	(0.011)
1 for woman	-0.5553***	(0.024)	-0.5525***	(0.025)	-0.3821***	(0.025)
1 for urban	0.6048***	(0.037)	0.5211***	(0.036)	0.2805***	(0.035)
no education	-0.9971***	(0.028)	-0.9335***	(0.029)	-0.6804***	(0.030)
age	0.0035***	(0.001)	0.0014	(0.001)		
knows s^{one} HIV+			0.2676***	(0.020)	0.1974***	(0.020)
ever tested for aids			0.2846***	(0.021)	0.1755***	(0.021)
age1519					-0.1057***	(0.027)
age2024					0.1011***	(0.025)
age2529					0.1394***	(0.027)
age3039					0.0843***	(0.024)
magazines and newspapers					0.4639***	(0.021)
radio less than once a week					0.2552***	(0.031)
radio at least once a week					0.3309***	(0.033)
radio almost every day					0.4512***	(0.029)
tv less than once a week					0.2070***	(0.030)
tv at least once a week					0.1222***	(0.034)
tv almost every day					0.3488***	(0.029)
Ethiopia	0.0221	(0.055)	0.1567***	(0.056)	0.1503***	(0.054)
Ghana	0.2913***	(0.057)	0.3232***	(0.056)	0.2123***	(0.053)
Kenya	1.2290***	(0.049)	1.1372***	(0.047)	0.9514***	(0.043)
Malawi	0.4797***	(0.050)	0.3975***	(0.049)	0.3236***	(0.048)
Zimbabwe	0.7670***	(0.047)	0.8269***	(0.048)	0.7911***	(0.048)
Constant	3.6232***	(0.058)	3.5291***	(0.059)	2.8849***	(0.058)
Observations	81,157		70,025		69,775	
Number of clusters	2,713		2,713		2,713	
R squared within	0.0392		0.0429		0.0641	
R squared between	0.4980		0.4800		0.5215	
R squared overall	0.1418		0.1374		0.1624	

Note: Clustered robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The omitted dummies are male, rural, any formal education, Cameroon, 40 and older; and for equation (3): never listen to the radio, never watch tv, never read newspapers.

Table 7: Access to condoms- Random Effects probit model estimates

	(1) 1 if know any place		(2) 1 if know a public place		(3) 1 if know a privhealth place		(4) 1 if know a other priv place		(5) 1 if could get a condom	
Dldistroad	-0.0427***	(0.008)	0.0433***	(0.009)	-0.0745***	(0.009)	-0.0742***	(0.008)	-0.0509***	(0.009)
equal 1 for woman	-0.5624***	(0.014)	0.0302**	(0.013)	-0.0999***	(0.014)	-0.5096***	(0.013)	-0.8467***	(0.024)
married	0.3803***	(0.016)	0.3566***	(0.016)	0.2022***	(0.017)	0.0751***	(0.016)	0.4564***	(0.023)
prev. married	0.3297***	(0.024)	0.2603***	(0.023)	0.1862***	(0.025)	0.1210***	(0.023)	0.4333***	(0.035)
age	-0.0039***	(0.001)	0.0029***	(0.001)	-0.0013	(0.001)	-0.0090***	(0.001)	-0.0068***	(0.001)
urban	0.2536***	(0.028)	0.0216	(0.031)	0.3718***	(0.031)	0.2542***	(0.028)	0.1425***	(0.033)
primary educ	0.5794***	(0.017)	0.4686***	(0.018)	0.4375***	(0.021)	0.4065***	(0.018)	0.1047***	(0.027)
secondary educ	0.9924***	(0.021)	0.7170***	(0.021)	0.8187***	(0.023)	0.6190***	(0.022)	0.2968***	(0.031)
higher educ	1.2892***	(0.044)	0.7612***	(0.034)	1.1338***	(0.035)	0.7807***	(0.035)	0.4863***	(0.052)
wpoorest	-0.4041***	(0.026)	-0.1078***	(0.026)	-0.4818***	(0.029)	-0.3620***	(0.026)	-0.0438	(0.036)
wpoorer	-0.3464***	(0.025)	-0.0773***	(0.025)	-0.3865***	(0.027)	-0.3003***	(0.024)	-0.1057***	(0.033)
wmiddle	-0.3025***	(0.024)	-0.0717***	(0.023)	-0.3149***	(0.025)	-0.2519***	(0.022)	-0.0951***	(0.031)
wricher	-0.2359***	(0.021)	-0.0445**	(0.020)	-0.2557***	(0.021)	-0.1753***	(0.019)	-0.0603**	(0.027)
catholic	0.1718***	(0.026)	0.0925***	(0.026)	0.0749***	(0.027)	0.1060***	(0.026)	0.0691**	(0.035)
protestant	0.1708***	(0.023)	0.0842***	(0.024)	0.0439*	(0.025)	0.0840***	(0.023)	-0.0020	(0.032)
other religion	0.0670***	(0.025)	0.0314	(0.026)	0.0236	(0.027)	0.1080***	(0.026)	0.1665***	(0.040)
scoreclosed6	0.0933***	(0.003)	0.0597***	(0.003)	0.0652***	(0.003)	0.0665***	(0.003)	0.0519***	(0.004)
Constant	-0.1519***	(0.053)	-2.3035***	(0.059)	-1.1652***	(0.059)	0.0432	(0.053)	0.6620***	(0.072)
Observations	76,030		76,030		76,030		76,030		42,688	
Number of clusters	2,713		2,713		2,713		2,713		2,652	

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Country dummies are included.

The omitted dummies are male, single, rural, no education, richest, muslim.

Table 8a: Last sexual intercourse with spouse and condom
Bivariate Probit Model

	(1)		(1)	
	1 if sex with spouse		1 if condom use	
Dldistroad	0.0391***	(0.011)	-0.0245***	(0.009)
knowplacecondom			-0.4143***	(0.037)
equal 1 for woman	0.3092***	(0.036)	-0.4944***	(0.023)
married	4.3634***	(0.077)	-1.2986***	(0.028)
prev. married	1.8267***	(0.079)	-0.2791***	(0.037)
current age	0.0170***	(0.002)	-0.0141***	(0.001)
urban	-0.1660***	(0.045)	0.1417***	(0.031)
primary educ	-0.2396***	(0.044)	0.3719***	(0.041)
secondary educ	-0.3457***	(0.049)	0.6447***	(0.045)
higher educ	-0.2496***	(0.074)	0.8022***	(0.061)
wpoorest	0.0687	(0.059)	-0.3212***	(0.046)
wpoorer	0.0919*	(0.055)	-0.2525***	(0.042)
wmiddle	0.0042	(0.049)	-0.2130***	(0.037)
wricher	-0.0526	(0.041)	-0.0916***	(0.029)
catholic	-0.2465***	(0.053)	0.1269***	(0.041)
protestant	-0.0929*	(0.051)	0.0418	(0.039)
other religion	-0.2823***	(0.058)	0.0827*	(0.047)
scoreclosed6	-0.0198***	(0.006)	0.0447***	(0.005)
Ethiopia	1.4434***	(0.079)	-0.7446***	(0.053)
Ghana	0.4906***	(0.054)	-0.3449***	(0.040)
Kenya	0.8997***	(0.052)	-0.5856***	(0.042)
Malawi	1.1966***	(0.051)	-0.3142***	(0.041)
Zimbabwe	0.9192***	(0.055)	-0.3568***	(0.037)
Constant	-3.4409***	(0.125)	0.4911***	(0.082)
Observations	48,798			
Number of clusters	2,703			
Wald test of rho= 0:	chi2(1)= 514.522 - $Prob > chi2 = 0.0000$			

Note: Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The omitted dummies are male, single, rural, no education, richest, muslim, Cameroon.

Table 8b: Last sexual intercourse with spouse and condom
Bivariate Probit Model (marginal effects)

	(1)		(2)		(3)		(4)	
	P[y1=1, y2=1]		P[y1=1, y2=0]		P[y1=0, y2=1]		P[y1=0, y2=0]	
Dldistroad	-.0014*	(.001)	.0066***	(.002)	-.0013***	(.000)	-.0039***	(.001)
knowplacecondom	-.0433***	(.004)	.0433***	(.004)	-.0097***	(.001)	.0097***	(.001)
female	-.0431***	(.003)	.0876***	(.006)	-.0206***	(.002)	-.0239***	(.004)
married	.0305***	(.001)	.9400***	(.003)	-.2960***	(.009)	-.6745***	(.009)
prev. married	-.0084***	(.003)	.0921***	(.004)	-.0170***	(.001)	-.0666***	(.003)
age	-.0009***	(.000)	.0032***	(.000)	-.0006***	(.000)	-.0016***	(.000)
urban	.0096***	(.003)	-.0326***	(.007)	.0069***	(.001)	.0161***	(.005)
primary educ	.0300***	(.004)	-.0628***	(.007)	.0140***	(.002)	.0188***	(.005)
secondary educ	.0602***	(.006)	-.1104***	(.009)	.0266***	(.002)	.0236***	(.006)
higher educ	.1149***	(.014)	-.1539***	(.017)	.0373***	(.006)	.0017	(.009)
wpoorest	-.0231***	(.003)	.0319***	(.008)	-.0073***	(.001)	-.0015	(.006)
wpoorer	-.0183***	(.003)	.0299***	(.007)	-.0065***	(.001)	-.0052	(.006)
wmiddle	-.0169***	(.003)	.0175***	(.007)	-.0045***	(.001)	.0039	(.006)
wricher	-.0088***	(.002)	.0016	(.006)	-.0010	.	.0081*	(.005)
catholic	.0062	(.004)	-.0428***	(.009)	.0088***	(.002)	.0278***	(.007)
protestant	.0019	(.003)	-.0142**	(.007)	.0028*	(.001)	.0095*	(.006)
other religion	.0012	(.004)	-.0442***	(.010)	.0084***	(.002)	.0346***	(.009)
scoreclosed6	.0036***	(.000)	-.0062***	(.001)	.0014***	(.000)	.0013*	(.001)
Ethiopia	-.0334***	(.002)	.1302***	(.004)	-.0217***	(.001)	-.0751***	(.003)
Ghana	-.0192***	(.002)	.0693***	(.005)	-.0123***	(.001)	-.0377***	(.004)
Kenya	-.0293***	(.002)	.1047***	(.004)	-.0177***	(.001)	-.0576***	(.003)
Malawi	-.0098***	(.003)	.1136***	(.005)	-.0207***	(.001)	-.0830***	(.003)
Zimbabwe	-.0158***	(.003)	.0995***	(.005)	-.0179***	(.001)	-.0658***	(.003)

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The omitted dummies are male, single, rural, no education, richest, muslim, Cameroon.

Table 9: **Sensitivity analysis**

Panel A: Knowledge about HIV-transmission

	(1)	(2)	(3)
	scoreclosed6	scoreclosed6	scoreclosed6
Cameroon	-0.0629***	-0.0603***	-0.0371***
Ethiopia	-0.0669***	-0.0660***	-0.0357***
Ghana	-0.0656***	-0.0612***	-0.0319***
Kenya	-0.0584***	-0.0553***	-0.0262**
Malawi	-0.0596***	-0.0529***	-0.0168
Zimbabwe	-0.0780***	-0.0693***	-0.0407***

Panel B: Access to condoms

	(1)	(2)	(3)	(4)	(5)
	1 if know any place	1 if know a public place	1 if know a privhealth place	1 if know a other priv place	1 if could get a condom
Cameroon	-0.0262***	0.0384***	-0.0491***	-0.0581***	-0.0497***
Ethiopia	-0.0214**	0.0442***	-0.0492***	-0.0545***	-0.0478***
Ghana	-0.0322***	0.0323***	-0.0650***	-0.0539***	-0.0526***
Kenya	-0.0272***	0.0380***	-0.0560***	-0.0559***	-0.0585***
Malawi	-0.0303***	0.0472***	-0.0495***	-0.0649***	-0.0412***
Zimbabwe	-0.0466***	0.0264***	-0.0721***	-0.0527***	-0.0491***

Panel C: Likelihood of HIV-infection

	(1)	(2)	(3)	(4)
	hivpositive	hivpositive	hivpositive	hivpositive
Cameroon	-0.0525***	-0.1902***	-0.5210***	-0.5488***
Ethiopia	-0.0445***	-0.1802***	0.0227	-0.3677
Ghana	-0.0461***	-0.1676***	-0.6661***	-0.6284***
Kenya	-0.0559***	-0.1719***	-0.5093***	-0.5749***
Malawi	-0.0333***	-0.1314***	-0.5342***	-0.5958***
Zimbabwe	-.0555***	-.1593***	-.4426***	-.4600***

Panel D: Last sexual intercourse

	(1)	(2)
	1 if spouse	1 if condom
Cameroon	0.0305**	-0.0219**
Ethiopia	0.0413***	-0.0208**
Ghana	0.0432***	-0.0266**
Kenya	0.0409***	-0.0277***
Malawi	0.0319**	-0.0266***
Zimbabwe	0.0437***	-0.0201**

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient reported is the coefficient for $D_j ldistroad_j$.

One country is removed at a time and the country removed is given in the first column.

Table 10: Country-by-country analysis

Panel A: Knowledge about HIV-transmission

	(1) scoreclosed6	(2) scoreclosed6	(3) scoreclosed6
Cameroon	-0.0726***	-0.0647**	-0.0108
Ethiopia	-0.0296	-0.0140	0.00938
Ghana	-0.0642*	-0.0626*	-0.0383
Kenya	-0.0984***	-0.0810***	-0.0554***
Malawi	-0.0777***	-0.0768***	-0.0728***
Zimbabwe	-0.00456	-0.0304	-0.00264

Panel B: Access to condoms

	(1) 1 if know any place	(2) 1 if know a public place	(3) 1 if know a privhealth place	(4) 1 if know a other priv place	(5) 1 if could get a condom
Cameroon	-0.0428**	0.0249	-0.0945***	-0.0339**	-0.0400**
Ethiopia	-0.0321**	0.0362**	-0.0610***	-0.0403***	-0.0066
Ghana	-0.0183	0.0793***	-0.0154	-0.0797**	-0.0422
Kenya	-0.0444**	0.0390*	-0.0608***	-0.0503***	-0.0012
Malawi	-0.0155	0.0068	-0.0676***	-0.0044	-0.0664***
Zimbabwe	0.0127	0.0676***	0.0101	-0.0741***	-0.0441

Panel C: Likelihood of HIV-infection

	(1) hivpositive	(2) hivpositive
Cameroon	-.0138	0.1974
Ethiopia	-.0516*	-0.0373
Ghana	-.0587**	-0.3516**
Kenya	.0301	-0.0985
Malawi	-.1241***	-0.0679
Zimbabwe	-.0465***	-0.1238**

Panel D: Last sexual intercourse

	(1) 1 if spouse	(2) 1 if condom
Cameroon	0.0496**	-0.0373**
Ethiopia	-0.0262	-0.0189
Ghana	0.0195	-0.0040
Kenya	0.0241	-0.0033
Malawi	0.0855***	-0.0120
Zimbabwe	0.0193	-0.0354

Note: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient reported is the coefficient for $Djldistroad_j$.

The first column reports the country under analysis.