HIV and Access to Road*

Elodie DJEMAI†
Population Studies Center, University of Michigan
Toulouse School of Economics

This version: February 2011
First draft: June 2009

Abstract

Using GIS and HIV data from six African countries, we estimate the effect of road proximity on HIV. Access to roads is shown to improve knowledge of STI transmission and availability of condoms, but also increase the propensity to engage in casual sex. The latter effect more than offsets the former, so that at distance mean, a one standard deviation increase in distance to the nearest road decreases the risk of HIV infection by 0.6 percentage point. Results are robust to controlling for endogenous road placement, using slope and ruggedness as instrumental variables. Migration and endogenous individual placement are also discussed.

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

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

1 Introduction

Transportation infrastructure is essential to development as it facilitates the movement of people, goods and services. A large set of economic literature has examined the effects of infrastructure on various outcomes of interest, such as poverty, trade, and economic performance. The role

*Raj Arunachalam, Martine Audibert, Jean-Paul Azam, Sonia Bhalotra, Samuel Bowles, Andrew Clark, Pascaline Dupas, Youenn Lohéac, Céline Nauges, Emily Oster, Paul Seabright, Stéphane Straub and Rebecca Thornton provided insightful comments. The previous versions of the paper have been presented in seminars at Toulouse School of Economics, Paris School of Economics, Edinburgh, Bristol (CMPO), Santa Fe Institute, Université Paris 1-Sorbonne, Université de Paris Dauphine (DIAL) and in conferences at the Journées de Microéconomie Appliquée (2009), International Conference on Infrastructure Economics and Development (2010), the ADRES doctoral conference (Lyon, 2010), the CSAE conference (2010), the annual congress of the European Society for Population Economics (2010), the European Economic Association Annual Congress (Glasgow, 2010) and AFSE Congress (Paris, 2010). Previous versions of the paper have circulated under the title "How do Roads Spread AIDS in Africa? A Critique of the Received Policy Wisdom". The usual disclaimer applies.

†elodiedjemai@gmail.com

of transport infrastructure in alleviating poverty has been examined in a number of countries, especially in Papua New Guinea (Gibson and Rozelle 2003), in Ethiopia (Dercon *et al* 2008), in Bangladesh (Khandker *et al* 2009), in Cameroon (Gachassin *et al* 2010) and in Nepal (Dillon *et al* 2011). In low-income countries, investing in infrastructure facilitates trade as it reduces transportation costs (Jacoby 2000), and thus enhances non farm earnings (Jacoby and Minten 2009). Buys *et al* (2010) estimate that a major investment in upgrading the road network and in maintaining the existing roads could lead to a rise in trade flows within Sub-Saharan Africa by about USD 250 billion over 15 years. A body of research along these lines has examined the role of railways infrastructure in enhancing economic performance in Asian countries (Straub 2008) and in particular in China (Banerjee 2009) and India (Donaldson 2010).

Yet despite a large series of studies on the road impact on economic outcomes, little is known about its impact on disease outcome. Diseases are often geographically transmitted by trade and travel. Trade along roads may contribute to rapid propagation of communicable diseases through the induced displacement of people. However in the case of an infectious disease like HIV/AIDS, the relationship between road access and infection is ambiguous, and needs to be explored empirically. Road proximity may have two competing effects on the HIV risk among the general population. On one hand, roads may reduce risk by increasing access to condoms and knowledge about HIV/AIDS, both diminishing the cost of protection. On the other hand, roads may increase the risk of infection by bringing people living in accessible areas in contact with mobile populations who are at high risk of infection.

This paper explores the net effect of road infrastructure on the risk of HIV-infection using a unique dataset constructed by matching data from large household surveys from the Demographic and Health Surveys to geographical data on road infrastructure for a sample of African countries. The most recent Demographic and Health Surveys collect geographic locations for respondent communities in a way that allowed us to ascertain the distance between the community and the nearest road by applying cartographic techniques. We find compelling evidence that people living in proximity to a road are actually more likely to be infected with HIV than people living in remote areas. For mean values of all covariates, a one standard deviation increase in distance to the nearest road, that is about 3.4 kilometers, reduces the risk of infection by between 0.5 and 1.0 percentage point. We also examine the role of means of transportation, showing that owning a car, bike, or motorcycle makes the presence of the road irrelevant to HIV risk, while owning no means of transportation makes the distance to a road be highly protective. The effect of road distance on HIV risk is shown to be sensitive to volume of traffic passing along the road.

The main issue when estimating the effect of road on HIV infection is that roads are non-randomly placed and individuals may decide to move to live close to a road. The approach to address the issue of endogenous road placement is two-fold. First, we control for a set of potential confounding factors at the community level, such as urbanization, population density, distance to the nearest city and wealth. Second we estimate the effect of road in a IV strategy instrumenting the road distance variable using the community land gradient and community terrain ruggedness. The rationale for using these geographical characteristics as instrumental variables for road distance is that both land gradient and terrain ruggedness affect the cost of building a paved road. A few recent papers have used slope measures to predict or instrument for the placement of physical infrastructure such as dams (Duflo and Pande 2007), cellular network (Batzilis *et al* 2010) or electricity (Dinkelman 2011). The issue of non-random individual placement is also discussed in the paper by observing the individuals' migratory patterns and HIV testing, and the community recent in-migration. We find that the risk-reducing effect of road distance is robust in any case, and statistically significant in most cases. Other threats to validity of the estimates are discussed, such as the refusal to be tested for HIV.

Last, the role of road distance in the supply and demand for protection is examined to disentangle which channel is driving the observed relationship. As expected, we found that the likelihood of having access to condoms and the quality of HIV/AIDS-knowledge decrease with the distance to a road, so that at mean value of all covariates, the predicted probability of knowing at least one place where one can get a condom is equal to 0.71 and a one standard deviation increase in road distance reduces this likelihood by 0.9 percentage point. The observed positive relationships between road proximity and HIV/AIDS knowledge, access to condoms, and HIV infection suggest that ignorance and misfortune are not driving our results. That is, although people living close to road infrastructure tend to have greater HIV/AIDS knowledge and more access to condoms than their more remote counterparts, these factors appear to be insufficient to prevent them from being infected in larger proportions. We found that a one standard deviation increase in road distance increases the likelihood of having sex with one's spouse and reduces the likelihood of using a condom by 4.6 and 5.1 percentage point respectively, at mean distance. This empirical finding suggests that the observed positive relationship between road proximity and HIV risk is due to deficiencies in the demand for protection rather than the supply. Indeed, the increase in the risk of infection due to road infrastructure is found to be driven by an increase in the likelihood of having casual sexual partners - an increase that offsets the rise in condom use that is found in proximity to a road.

This paper contributes to two literatures. First, the paper adds to what is known about the effects of physical infrastructure in the developing world, placing a new emphasis on its impact on the spread of AIDS, and especially on the risk of being infected with HIV. A variety of works examine the impact of physical infrastructure, railways and roads in particular, on welfare and economic performance. At the national level, these infrastructures are found to be strong determinants of development. Investing in infrastructures facilitates trade (Buys, Deichmann and Wheeler 2010) as it immediately reduces transportation costs (Jacoby 2000). History shows the persistent benefits of such investments; disparities in colonial investments in West Africa are found to be one of the main determinants 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. This literature tends to focus on Asian countries (Straub *et al* 2008) and particularly China (Banerjee *et al* 2009) and India (Donaldson 2010). Donaldson (2010) and Banerjee *et al* (2009) studied the role of railroads in trade expansion and income level and income growth, respectively. A concern in this type of analysis is reverse causality because infrastructures are potentially driving the growth trajectory while wealthy countries are more able to finance public investment in infrastructures than poor countries. When controlling for endogeneity, the role of railroads in income level and income growth turns out to be mitigated in Banerjee *et al* (2009).

Second, it contributes to the literature on mobility and HIV in Sub-Saharan Africa. It places new emphasis on the effect of mobile groups such as truck drivers, travelers, and seasonal workers on HIV prevalence in the local populations rather than on HIV in the mobile groups themselves, and uses individual-level data and road proximity as variable of interest. The risk profile of mobile groups in this area has been examined in a series of papers that find a greater likelihood of both HIV infection and HIV-related risky behavior. These analyses have focused on long-distance truck drivers (e.g., Oruboloye *et al* 1993; Rakwar *et al* 1999; Ramjee and Gouws 2002; Morris and Ferguson 2007) and the temporary migrant workers (e.g., Meekers 2000; Adaji Nwokoji and Ajuwon 2004). This paper evaluates the impact of mobility on HIV prevalence in the general population, as does work by Oster (2011) and Tanser *et al* (2000). However this study differs from Oster (2011) - who predicts regional prevalence rates as a function of exports - in its evaluation of the impact of road distance at the individual level. Although Tanser *et al* (2000) also evaluate the relationship between road proximity and HIV in a rural South African setting, they use data from Antenatal Clinic (ANC) Surveillance on the prevalence of HIV among pregnant women who come to these clinics for antenatal health care. They show a correlation between the location of the clinics and the ANC-based prevalence rates of HIV. This analysis should be interpreted carefully because ANCs are not uniformly distributed within countries -their location is strongly determined by proximity to the road network - and because it considers HIV rates only among pregnant women who seek health care.

Two main policy implications can be drawn from our empirical findings. The first concerns the design of public policies to fight the epidemic. Over the last 25 years, public policies to fight AIDS in Africa have hinged on providing information about HIV transmission and subsidizing or providing free condoms. Our results suggest that while informing people about HIV risks and facilitating/providing self-protection may be of some benefit, these measures do not counter other factors that increase risky HIV-related behavior. To curb the spread of the HIV/AIDS epidemic, people may need to receive incentives to use the self-protective measures available to them. The second policy implication concerns investment in road infrastructure, which our findings suggest has heretofore unexplored costs and benefits. Building more roads will increase access to condoms and improve individual knowledge about the risk of infection, but will also raise HIV prevalence rates.

The paper is organized as follows. Section 2 discusses the influence of road infrastructure on HIV/AIDS outcomes. Section 3 describes the data used in the analysis and the empirical strategy. Section 4 presents the empirical results and discusses the identification and threats to validity. Section 5 examines the implications of the findings for the role of ignorance versus intentional risk taking in HIV infection. Section 6 concludes.

2 Theory

Although the direct impact of road infrastructure on HIV/AIDS outcomes has never been explored, reason suggests a number of plausible rationales for an association between the two. By facilitating travel and access to markets, roads may have two competing effects on the risk of HIV infection.

On one hand, road infrastructure facilitates access to goods and services, which may include not only access to information about the risks of HIV/AIDS infection but access to direct protection against infection, namely to condoms. The quality of individuals' knowledge about HIV/AIDS may be increased via proximity to educational and health facilities and access to media such as newspapers, television, and radio - all of which are associated with the presence of road infrastructure. Furthermore people living in proximity to roads are by definition more easily reachable by sensitization groups. People living close to a road are also more likely to have access to shops and health-care facilities that may stock condoms as well. Also the flow of people brought by the presence of the road may confer some anonymity to the purchase of condoms. Therefore, roads may act to increase motivation to avoid infection and reduce the costs associated with protection.

On the other hand, road infrastructure also facilitates physical contact and communication among people, which may accelerate the spread of the epidemic by increasing the pool of potential sexual partners, increasing the proportion of infected potential partners, and diversifying the probability of infection. Those living close to roads will have more contact with mobile groups in particular - truck drivers, migrant workers, travelers, servicemen, traders - groups more likely to be HIV infected and to engage in risky sexual behaviors than local populations. For example, Oruboloye *et al* (1993) found that long-distance truck drivers in Nigeria are more likely to engage in multiple sexual partnerships, including stable partnerships with women who are not commercial sex workers. Meekers (2000) reported similar findings for migratory mine workers in South Africa. Adaji Nwokoji and Ajuwon (2004) found that naval servicemen posted abroad had a higher number of sexual partners, were more likely to have had sex with a female sex worker, and were less likely to have used condoms during their last sexual intercourse with a sex worker than naval personnel stationed locally. In addition, road infrastructure may prompt residents of accessible communities to use the road to visit areas where HIV prevalence may be higher, leading to greater risk. By contrast in remote communities, the HIV prevalence rates tend to be stable and contained, as the rate of in- and out-migration is lower.

An important empirical question is which of these two sets of road infrastructure effects dominate in the reduced-form relationship between road distance and HIV infection. Of particular interest are the competing influences of increased opportunity to engage in risky sexual encounters versus increased access to information on and protection against HIV infection. 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 in the demand for condom use and on extramarital sexual encounters is not straightforward. And although road infrastructure extends potential sexual networks, the increase in the supply may not be followed by an increase in demand for casual sex, especially given the concomitant increase in access to information on inherent risks.

3 Data Description and Estimation Strategy

3.1 Sample Characteristics

The Demographic and Health Surveys (DHS) are collected in several countries across the world using a standardized sampling design and standardized questionnaires that allow for cross-country comparisons in terms of health care, and maternal and child health. 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 aged 15-49 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[1] of the sampled households were selected. In this subsample, all men aged 15-54[2] were eligible to be interviewed if they were either permanent residents or visitors present in the household on the night before the survey.

A module about HIV/AIDS is included in the DHS to assess the knowledge, attitudes, and practices of the general population. 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. In addition to the self-reported sexual behaviors and HIV-status, the new generation of surveys collect geographic locations in the form of latitude and longitude coordinates for the communities (or sampled clusters) where DHS respondents live. A community is defined geographically as a city block in urban areas or as a village in rural areas. One latitude/longitude coordinate is collected for each community in the DHS survey and is essential for the empirical strategy of this study because it enables us to locate each community on a country map and relate it with the road network. For reasons of confidentiality, 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. This issue of measurement error is discussed later in the paper.

From among countries with geocoded DHS data containing the HIV testing component, we select Cameroon (2004), Ethiopia (2005), Ghana (2003), Kenya (2003), Malawi (2004) and Zimbabwe (2005/06) for analysis[3]. We restrict the sample to the usual residents that constitute 97.10% of the total sample because we can not generate the distance to the road for the visitors, for whom we lack residential information. Our final sample contains 86,644 individuals (see Table I) including 53,579 individuals tested for HIV.

Table II presents means and standard deviations of key variables for the entire sample and by country. The average rate of HIV prevalence is 7.9% over the six countries. Ethiopia and Ghana are low-prevalence countries, where about 2% of the tested respondents had HIV, and Malawi and Zimbabwe are high-prevalence, where 12% and 18% of the adults tested positive, respectively.

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

[2]15-59 in Cameroon, Ethiopia, Ghana

[3]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.

The second line reports the proportion of respondents reporting that they have ever been tested for HIV. Note that this proportion is orthogonal to the percentage of respondents tested for HIV during the survey. In all, 15% of the respondents report being previously tested for HIV; in Ethiopia and Ghana, the proportion is less than 10%. Investigating the reasons for these low rates of HIV testing is beyond the scope of the study, but it is worth noting that the vast majority of the population does not know their HIV status, either because they have never been tested, or tested so long ago that the results may no longer be salient, or tested but never received their results.

As a consequence of the sample design, 66% are women, except in Ghana and Zimbabwe where the sample is closer to the gender balance. Respondents are 29 years old on average; 33% are living in urban areas. Although 25% of the entire sample report no formal education, this proportion ranges from 3% in Zimbabwe to 54% in Ethiopia. Over all, 38% have some primary and 34% have some secondary education. As far as religious affiliations, 47% of the respondents are protestant, 18% Catholic and 17% Muslim. In Zimbabwe, more than two-thirds of the respondents are protestants; the highest proportion of Muslims is found in Ethiopia (33% of the respondents).

In the module of questions about HIV/AIDS, respondents are asked about the validity of six statements about HIV-transmission, such as *"Can people reduce their chance of getting the AIDS virus by not having sex at all?"*[4]. For each question, we observe whether the individual answers correctly, incorrectly, or does not know. When generating a measurement for HIV/AIDS knowledge, the main concern is to deal with the "don't know" answers, which in most survey analyses are recoded as either a missing value or a wrong answer. We adopt none of these options for the following reasons. First, while a missing answer is mainly due to a coding error, a nonapplicable question, or a refusal to answer, in our study, the "don't know" answers in this case reveal important information on respondents' level of knowledge about HIV/AIDS, especially their level of ignorance. Second, treating "don't knows" as a wrong answer is misleading because ignorance is different from misunderstanding or misinformation. For example, in bargaining over condom use, someone who

[4] 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?"*.

is ignorant of their protective qualities may not be as reluctant as someone who is convinced that condoms are useless.

To keep as much information as possible, for each question k, we generate a variable $score_{ik}$ equal to 1 if individual i answers question k correctly, -1 if incorrectly, and 0 for "don't know." This formulation sanctions the false statement more than the ignorance, and will be discussed later in the analysis of HIV/AIDS-knowledge. Scores are summed for these six answers, generating a total score from -6 to +6, +6 being the score of an individual who answers all questions correctly. A full 97% of the sampled respondents report that they have heard about AIDS, but the means of contracting and preventing HIV infection are not widely understood. In particular, 24% of the respondents still think that HIV can be transmitted through mosquitos and 19% think that one cannot protect against HIV with a condom. For the whole sample, the average score is 3.6, but the distribution of scores varies across countries, with Kenya and Ethiopia being respectively at the top and the bottom of the distribution.

3.2 Geographical Data

Our empirical analysis relies on measures of the community's distance to the nearest paved road. To construct this variable, we use data from the Digital Chart of the World (DCW) available at the Harvard Geospatial Library. This GIS database gives the worldwide road network from which we extract the network of paved roads for each of the six sampled countries. We restrict our analysis to paved roads because these are the routes used to transport goods and people, and they are passable all the year long. Our measure of interest is computed in two steps. First we use cartographic techniques on ArcGIS to map the road network and the communities on a country map, as depicted in Figure I for Zimbabwe (see online appendix for the other country maps). Second, we compute the straight-line distance in kilometers between the center of the community and its nearest paved road.

Respondents are split into 466 clusters in Cameroon, 529 in Ethiopia, 412 in Ghana, 399 in Kenya, 521 in Malawi, and 398 in Zimbabwe (see Table I) - each cluster with an assigned community road distance. Table III reports the distribution of the respondents across countries according to

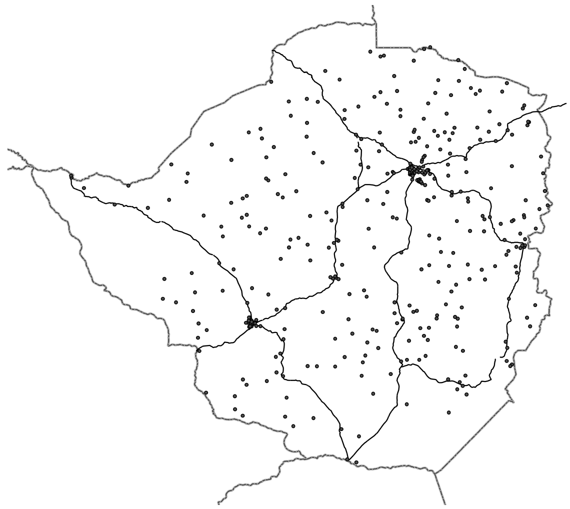

Figure I: Cluster location and paved road network, realized with 2005/06 Zimbabwe DHS

their proximity to the nearest primary road. Over the entire sample, 11% of the respondents (11,375 individuals) live on a paved road, ranging from only 7% in Malawi to 28% in Ethiopia. Respondents live an average of 24.39 kilometers from the nearest paved road, ranging from 20 kilometers away in Cameroon and Malawi to 31 kilometers away in Zimbabwe. Cameroon and Kenya seem to have the most developed road networks since 75% of these respondents live less than 29 and 25 kilometers, respectively, from a paved road. However in Kenya, there is a wide heterogeneity among the 25% of the respondents who live furthest from a road since they are distributed between 25 and 288 kilometers. Malawi appears as the country in which the paved road network is the most dense since the furthest distance to a road is 95 kilometers; this feature is probably due to the narrowness of the country.

3.3 Empirical Strategy

Our main estimation strategy consists in establishing the differences in the risk of HIV-infection between individuals who live close to a paved road and those who do not. The difference between our estimates and a treatment effect is that we use a continuous measure of the intensity of treatment (road distance), and thereby capture more variation in the data. Let HIV_{ijr} be the HIV status of individual i living in community j located in region r. Let $distroad_{jr}$ be a continuous variable for the distance between the center of community j and the nearest paved road. If $distroad_{jr}$ was randomly assigned across communities and people, we could estimate the average treatment effect of road distance by estimating the following equation through a probit model:

$$Pr(HIV_{ijr} = 1) = \phi(\alpha + \beta log(1 + distroad_{jr}) + X_{ijr}^{I\prime}\delta_1 + X_{jr}^{J\prime}\delta_2 + \gamma_r + \varepsilon_{ijr}) \quad (1)$$

where α is the constant, X_{ijr}^{I} the set of individual characteristics and X_{jr}^{J} the set of community-level characteristics, γ_r is the region-specific effect and ε_{ijr} the error term. The region is defined as the intermediate disaggregated level between the community and the country, and corresponds to the official national regions.

Although most of the roads were built before the onset of the HIV/AIDS epidemic, road placement may have been driven by the same observable characteristics that are also driving the spread of the epidemic, such as wealth, trade and urbanization. If road placement were not randomly assigned across the country then the estimated coefficient of road distance, $\hat{\beta}$, would be biased.

To deal with factors that could be determinant in the road placement, we first control for a vector of community-level variables in estimating equation (1). Covariates include whether the community is located in a urban area, the population density, the distance to the nearest city (equal to zero for the urban communities), and the proportion of very rich people. The measure of population density comes from data produced by the Gridded Population of the World (GPW) project of the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University. We use the data for 2005 to compute the population density for each sampled community. The location of the cities, used to compute the straight-line distance between each rural community and its nearest city, comes from the layer "World Cities" in ESRI (2007). The wealth index is drawn from a principal component analysis generated at the country level by the data provider and based on durable goods' ownership. Using this index, respondents are divided into five wealth categories, and we compute for each community, the proportion of people who are in the highest wealth quintile to distinguish the wealthiest communities.

We also deal with potential bias due to the non-random placement of roads by instrumenting the road distance variable using the community land gradient and community terrain ruggedness. The rationale for using land slope as an instrumental variable for road distance is that land gradient affects the cost of building a paved road, suggesting that roads are more likely to be placed in flat than in steep areas. The same type of argument holds for using terrain ruggedness. A few recent papers have used slope measures to predict or instrument for the placement of physical infrastructure. Duflo and Pande (2007) have instrumented dam construction using river gradient across Indian districts, and Dinkelman (2011) has used land gradient as an instrumental variable for the placement of an electrification project in South Africa. Land gradient has also been identified as a strong determinant of the cellular phone coverage in Malawi by Batzilis et al (2010). As suggested in Dinkelman (2011), one concern with using land slope as an instrument for location is that in a rural setting, it may affect agricultural outcomes. In our case the direct impact of gradient on farm productivity would be through a change in wealth, which is controlled for in the estimations. Terrain ruggedness may not only influence the cost of road construction but also act as a good proxy for community isolation.

A second concern is that individuals may sort non-randomly across accessible and remote areas - in particular, respondents may have moved to live closer to a road. The moves may be endogenously related to the existence of the road or to the individual's HIV status. The latter may logically occur if HIV-infected people move to increase access to antiretroviral therapy or avoid stigma on their family. Therefore we must account for the individuals who have moved and also for those who are likely to know their HIV status. In the DHS, respondents are asked whether they were born in their current place of residence and if not, when they arrived in their current place. Because respondents who moved are not asked about their previous residence, we do not know whether it was closer or further away from a road than their current residence. In the absence of this information, we are estimating equation (1) using different sub-samples to rule out the possibility that migrant risk takers are driving our core results.

4 The Impact of Road Distance on HIV-risk

4.1 Average Impact of Road Distance

Table IV presents the estimated coefficients of equation (1) and their robust standard errors clustered at the community level. Column (1) estimates the likelihood of being HIV infected controlling for road distance and regional dummies, while column (2) adds a set of individual sociodemographic characteristics. Both columns show that increasing road distance significantly reduces the individual risk of HIV infection. Although the effect of road distance remains negative and statistically significant, controlling for individual wealth decreases the point estimate slightly (column (3)), suggesting that wealth is correlated with road distance and that wealth influences the risk of infection. Based on durable goods ownership, we find that respondents from the lowest two quintiles are at equal risk of infection, while they are less likely than the richer respondents to be infected with HIV. Depending on the set of conditioning information used in the estimation, the marginal effect of road distance varies between -0.0048 and -0.0076, suggesting that at mean value of all covariates, increasing the distance to a road by one standard deviation, that is about 3.4 kilometers, reduces the risk of infection by between 0.7 and 1.0 percentage point.

Estimations reported in Table V show that the negative and statistically significant relationship between road distance and HIV risk is robust to the inclusion of additional covariates that could be viewed as confounding variables at the community level, such as population density (column (1)), distance to the nearest city (column (2)), community well-being (column (3)) and urbanization (column (4)). Note that the effect of population density is not statistically significant in predicting HIV status when road distance is controlled for. The distance to the nearest city is negatively and significantly related to HIV risk, although the effect is small and does not rule out the role of road distance. Living in a rich community significantly increases the individual risk of HIV infection, as does living in a urban area; however controlling for urbanization does not rule out the road effect. Note that these community-level covariates are very correlated to whether the community is urban or rural, hence in the rest of the paper, we will only control for urban residence along with the individual characteristics. Controlling for individual-level and community-level characteristics, the estimation results suggest that at any covariate mean, the predicted probability of infection is about 4.4% and a one standard deviation increase in road distance reduces the likelihood of being infected with HIV by between 0.5 and 0.7 percentage point approximately.

4.2 Identification

4.2.1 Endogenous road placement

For probit estimates to be consistent requires that road infrastructure be uncorrelated with characteristics that could also drive HIV patterns. This assumption may be violated if, for instance, more trade-intensive or richer areas have both greater road infrastructure and greater HIV prevalence. To address this problem, and on top of controlling for confounding factors such as wealth and urbanization, we exploit the variation in road infrastructure induced by differences in land gradient across communities and, by differences in terrain ruggedness to obtain instrumental variable estimates.

We implement this instrumental variable strategy based on the geography of road infrastructure by constructing measures of land gradient and terrain ruggedness for each community of our sample, using ArcGIS. We construct the measure of land gradient using the SRTM digital elevation map[5]. To generate the terrain ruggedness index, we use the GTOPO30 elevation grid and the formula proposed in Riley *et al* (1999), as does work by Nunn and Puga (2011). We calculate the terrain ruggedness index for all grid cells in the six countries and then extract the exact value of the index for each community.

Estimated coefficients from IV regressions of road distance are displayed in Table VI, and are using exactly the same set of control variables as in the probit estimation displayed in column (4) in Table V. The IV estimates use a linear probability model instead of a probit because the dimension of the matrix makes the estimation impossible while controlling for regional dummies. Likewise, for illustrative purpose on the size of the coefficient, column (1) estimates the effect of road distance through a OLS-Linear Probability Model. Columns (2) and (3) report the IV estimates from a 2SLS specification in which land slope and terrain ruggedness, respectively, are used as instrumental

[5]We use the SRTM (Shuttle Radar Topography Mission) data, at a resolution or cell size of approximately 90 meters

variables. The first-stage estimations reported in column (2a) suggest that communities on higher slopes are consistently less likely to have access to a road since the relationship between road distance and land gradient is found positive and statistically significant. In the same line, column (3a) reveals that communities with highly rugged terrain are located significantly further to the road than their less rugged counterparts. These findings provide supportive evidence that roads are more likely to be built in flat and regular terrain.

Two results from the second-stage estimation displayed in columns (2b) and (3b) are of note. First, IV estimates suggest that after controlling for non-random road placement, road distance remains significantly and negatively related to the likelihood of HIV infection. People living in proximity to a paved road are more likely to be infected than their counterparts living in remote areas. Second, instrumenting road distance using land slope and using terrain ruggedness increases the point estimates in the two cases, suggesting that the risk-reducing effect of road distance is greater. The increase in point estimates suggest that if the roads were randomly assigned across communities, the individuals living in proximity to a paved road would have been at a greater risk of infection.

4.2.2 Endogenous individuals' placement

The issue of non-random placement of individuals is two-fold and may imply reverse causality and selection biases.

The argument of reverse causality would tell us that people who are HIV infected may have decided to move and live close to a road - perhaps to gain access to antiretroviral treatment or to live in a more populous area for reasons of anonymity and HIV-related stigma. In an effort to rule out the possibility that HIV infection has driven the migration decision, we look at whether the individual has ever been tested for HIV and the number of years living in the current place of residence.

The proportion of all respondents who report having ever been tested for HIV is only 15%; and only 10% of respondents have both been tested and received the results of their test. Moreover, among this 10%, some may have been HIV negative, and their results may be no longer salient.

Therefore, one may reasonably assume that most of the respondents do not know their current HIV status. However we still take into account this potential reverse causality bias formally in Panel A, Table VII, where we re-estimate the reference equation (Table V, column (4)), separating the sample in two groups depending on whether the individual has ever been tested for HIV. Empirical findings suggest that road distance reduces the risk of HIV infection for both groups, even after controlling for endogenous road placement. The Wald test of equality between the probit coefficients of each group fails to be rejected, suggesting that the road distance risk-reducing effect is not statistically different across the two groups.

We also attempted to rule out the reverse causality argument by removing from the estimation sample all respondents who lived in their current place of residence less than 10 years. Ten years is used as a threshold because in the absence of antiretroviral treatment, ten years is roughly the median period between HIV infection and death, so it is likely that HIV-positive respondents living 10 years or more in their current residence were infected after they moved - or equivalently, were not infected before they moved. Thus HIV infection would not be a driver in the migration decision for this group. The results confirm the robustness of the reference estimates over the whole sample in the sense that, for this sub-sample, the impact of road distance remains negative and statistically significant in all estimations except when road distance is instrumented by the land slope measure where the estimated coefficient fails to be statistically significant at a conventional level (with a p-value just above 10%)(see Panel B, Table VII).

The second issue of concern is that the results may be driven by a selection bias if more at-risk individuals have a higher likelihood to migrate and to move close to a road. Observable or unobservable variables may influence both the choice of residence location and the risk of infection. Given the lack of good instrumental variable to take into account the driver of migration and of the choice of residence, we estimate separately the effect of road distance for the sub-sample of respondents who were born in their current place of residence and for the sub-sample of respondents who have migrated. The underlying assumption is the reasons why the respondent's parents were living in her birth place are orthogonal to her own risk of infection.

When estimating separately the effect of road distance on the risk of HIV infection for the sub-samples of migrants and non-migrants, it turns out that road distance has a risk-reducing effect for both groups (see Panel C, Table VII). Although the effect fails to be statistically significant for the non-migrants when instrumenting for road distance, there is a priori no reason why the influence of land gradient and terrain ruggedness on road placement would have different effects for the migrants than for the non-migrants. We test for the equality of the point estimates between the two groups in the probit specification, and show that the Wald test of equality is rejected. Hence the effect of road distance on the risk of infection is statistically greater for the non-migrants than for the migrants.

Last, it is worth mentioning that most of the migrants of the sample have moved from a rural area to another rural area, suggesting that the move is related to a household migration. The circumstance of a male migrant moving alone and having potentially many sexual partners resulting from his migration is shown to be a marginal potential. We can identify the group of individuals who may be at the origin of the potential selection bias as those who migrate after reaching 15 years old and before getting married, in the sense that these respondents are those most likely to have initiated the decision to relocate (e.g. to find a job) and to most benefit from the extended set of sexual partners. Panel D re-estimates the benchmark equation removing from the estimation sample this subset of potential selection drivers. The estimation results suggest that the statistically significant negative relationship between road distance and HIV infection holds for respondents born in their current place of residence, and respondents who arrive at their current residence either as children or as married people. The point estimates in the probit (column (1)) and 2SLS (columns (3)-(4)) models are very similar to their respective benchmark point estimates (column (4), Table V and columns (2)-(3), Table VI, respectively), meaning that the selection bias, if any, is small in our analysis.

4.3 Heterogeneous Impact of Road Distance

4.3.1 Gender and Urban Heterogeneities

To examine the treatment effect heterogeneity across gender and urbanization, we estimate the road distance effect separately for women and men, and for urban and rural groups (see Table

VIII).

First, the effect of road distance on the likelihood of HIV-infection is negative and statistically significant for both sexes (see Panel A in Table VIII). The effect of road proximity is greater for males than for females. This difference in the size of the effect might reflect 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. Nevertheless the Wald test of the equality of the two estimated coefficients fails to be rejected, meaning that there is no statistically significant difference in the road distance effect between men and women. Moreover when controlling for endogenous road placement, the risk-reducing effect of road distance turns out to be statistically significant only for women.

Second, performing separate estimations for rural and urban agents (see Panel B), we find that the probit and 2SLS coefficients of road distance are negative for both groups, but are statistically significant only in the probit specification. As far as the size of the effect, the effect appears to be greater for rural agents than for urban agents indicating that rural agents react more to the presence of road than the urban ones. This finding may 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. Nevertheless the Wald test shows that there is no statistically significant difference in the road distance effect between urban and rural people. This empirical finding is supportive of our estimation strategy that consists in including both urban and rural people in the analysis.

4.3.2 The mobility scenario

As mentioned previously the impact of road proximity on the risk of HIV infection may be explained by reducing the distance between people. Not only individuals living close to a road are in touch with the mobile population using this road but they may also use this road to move to other areas where the prevalence rates are different and potentially higher than in their community, both factors increasing their risk of infection.

First to examine the role of individuals' mobility on the observed relationship between road distance and HIV risk, we separate the sample depending on the household ownership of car, motorcycle or bicycle. Before commenting the empirical findings, note that the vast majority of our sample does not own any transportation means. Over the estimation sample, 3%, 6% and 26% of the households own a motorcycle, a bicycle or a car respectively. The proportion varies from one country to another, if we take the bike ownership, only 3% of the Ethiopian households own a bike, while they are 50% in Malawi.

Panels A-C in Table IX reports the estimated coefficients from a probit (column (1)), and the two-stage least squares models using slope (column (2)) or ruggedness (column (3)) as instrumental variable. The results show that for the individuals who own a car, living close or far away from a paved road does not make the difference in their risk of HIV infection. The same is true for those who own a bike or a motorcycle. These results suggest that the agents who are able to be mobile, are not protected from living in a remote area since they are able to move around in such a way that they will bear a similar likelihood of being infected with HIV as the agents living on the road or at least very close to it. This said, the results show however that for the people who do not own a car, a bike or a motorcycle, living far away from a road prevents them from getting HIV. In the case of the bike non ownership, the 2SLS estimate fails to be statistically significant at 10% but the p-value is just above the conventional levels, at 11.5%.

Second, it is plausible that the effect of road distance or road proximity is sensitive to the use of the road. If 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, we should find that a road that is crowded with traffic induces a higher risk of infection than a road that is less used. To examine this scenario, our estimation strategy consists in interacting the road distance variable with a measure of traffic flows (Panel D). We construct a road-specific measure of traffic flows using the trade flows between neighboring countries that come from the Correlates of War project (Barbieri *et al* 2008). The COW data source provides annual import and export data in current US dollars for each pair of trade partners. We take the average of the trade 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 the total amount of trade flows that were transported through it. Note that we do not include measures of internal trade flows assuming that if internal trade flows affect the regions differently, this is captured by the regional dummy variables, that are controlled for in each estimation.

In the probit specification, the risk of infection is found to decrease with the distance to the road and to decrease at a lower speed if the road is widely used to transport goods. In the 2SLS specification, we found that the level of traffic flows increases the risk of infection and that its effect is sensitive to the individuals' location. Our finding confirm Oster (2011) who investigates the relationship between trade openness and HIV prevalence at the regional level and shows that HIV prevalence increases with the level of trade flows both in volume and in value, suggesting that the increased human mobility induced by trade increases the HIV prevalence rates. The coefficient of the interaction term shows that the effect of trade flows decreases with the distance to the road, showing that the effect of trade on HIV infection is sensitive to the individual's place of residence.

One pitfall of this analysis is that we do not observe the prevalence rates of the mobile population following the roads. The increase in the risk of infection induced by the presence of the roads might depend not only on whether the road is widely used to transport goods and people, but also on the prevalence rates of the people who use it[6].

[6] To incorporate this feature, we implement 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 community from the nearest neighboring country, of the prevalence rate on the other side of the frontier and of their interaction term. Prevalence rates come from UNAIDS (2004) and we use the average of the prevalence rates among the adult population in 2001 and 2003. None of these variables were found statistically significant.

4.4 Other threats to the validity of the results

One threat to validity of the results relies on the refusal of being tested for HIV. As mentioned above, except in Ghana and Zimbabwe, not all sampled households are eligible for HIV-testing (see last column of Table I). In each community, the sample of respondents eligible for the HIV-test is selected randomly at the household level, in such a way that all women and men in the households selected for the male survey are eligible. Among the 86,644 individuals surveyed over the six countries, 36,358 households were eligible for the male survey and HIV testing, which stand for 63,007 individuals. In other words, 73% of the sampled respondents were eligible for the HIV-testing. Among these individuals who are eligible for the test, some did not show up, either because they refused to be tested or because they were absent for the test, leading to a refusal rate of 15% over the whole sample used in the paper.

One could argue that the people who are at higher risk of infection could have been more likely to refuse the test because they fear knowing the result. However the survey design rules out this option because DHS makes it clear to their eligible respondents that the test is anonymous and no HIV status will be given after the test. We address the potential selection bias due to the refusal to be tested by studying (i) whether distance to the road affects the likelihood of showing up for the test among the sub-sample of eligible respondents; and (ii) whether high risk people are less likely to accept being tested.

Firstly, Panel A in Table X uses the sample of eligible respondents to predict the probability of showing-up for the test using road distance as a unique control variable. We find that distance to a road increases the acceptation propensity. As we have shown that people living close to a road are more likely to be infected, this additional feature suggests that in the absence of refusal we should have had a higher number of people tested positive in proximity to a road. As a consequence the point estimates of our estimation could be considered as lower bounds for the true estimates of the effect of road distance on the individual risk of HIV infection.

Secondly, one could argue that the likelihood of accepting the test depends on whether people are adopting HIV-related risky or safe practices. To address this threat to validity of the impact of road on the risk of infection, we look at whether the propensity to adopt HIV-related risky behaviors vary depending on whether people accepted to be tested or not. In the very last section of the paper, we estimate the probability of having sex with a usual partner (one spouse or cohabiting partner) and the probability of using a condom during the last sexual intercourse. This bivariate probit estimation allows us to get the propensity of adopting the riskiest practice, that is to have sex with a casual partner without wearing a condom. It turns out that this predicted probability is equal to 10.13% for the individuals who have not been tested while they were supposed to be, to 10.01% for those who were tested and 10.3% for those who were non eligible for the test. This suggests that the individual's risk taking behaviors do not vary across groups and do not influence the propensity of showing-up for the blood sample collection.

A second threat to validity relies on the measurement error that the DHS imposes on the community location in order to preserve anonymity. The error is randomly assigned such that the measurement error is not the main concern. The main concern is rather that it may be the case that the geographical values that have been attributed to one community should have been attributed to another one. To deal with this, we compute the distance between each community and its nearest community and remove from the analysis, the communities that are located at less than 2 km (in urban areas) or 5 km (in rural areas) away from the nearest sampled communities, to rule out any overlap. Panel B in Table X reports the probit and 2SLS estimated coefficients of road distance for the individuals who are living in the remaining communities for which there is not any possible overlap. Results suggest that the effect of road distance remains negative and statistically significant in each estimation.

5 Channels

5.1 Road Distance and the Access to Prevention

In order for road distance to affect HIV risk through the channel of reduced access to prevention, individuals must have less access to HIV/AIDS-related information and to self-protective devices when their communities are connected to a paved road. This would have been counter-intuitive given previous works showing the improved access to markets induced by the presence of roads.

Results presented in Tables XI and XII illustrate that individuals living far away from a road have a lower level of HIV/AIDS-knowledge and a lower access to condoms than their counterparts living in proximity to a road. Robust standard errors are clustered at the community level. Note that the estimations are using all DHS respondents and not only those with non missing HIV status.

5.1.1 HIV/AIDS Knowledge

Results displayed in Table XI illustrate that individuals living in proximity to a road have a greater level of HIV/AIDS-knowledge than their counterparts living in remote areas. 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.

Column (1) controls for standard demographic variables (gender, urban residence, education, age, wealth, marital status, religion), while columns (2) and (3) include additional variables to assess the individual's level of confrontation with the disease by controlling for whether the respondent has ever been tested for HIV and for whether he knows someone who is infected with HIV or has died from AIDS. Note that on average, about 45% of the respondents know someone who has HIV or who died from AIDS and as mentioned earlier, 15% declare having been tested for HIV. First, the results confirm the "model of confrontation" stressed in de Loenzien (2005) as knowing someone infected is found to increase the quality of HIV/AIDS-knowledge 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. This seems efficient since it turns out that someone who has 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 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 road distance is reduced by the introduction of these variables suggesting that the road effect partly captures the effect of the access to media on the quality of knowledge. 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.

The size of the effect is quite small however. Increasing the distance to a road by one standard deviation reduces the level of HIV/AIDS-knowledge by between 0.02 and 0.04 point. Even though the effect of the road is small, results indicate that spatial inequalities in the knowledge about AIDS persist due to the unequal access to information and technology.

Note that results are qualitatively similar when estimating the road impact on different indicators[7] of HIV/AIDS-knowledge as the level of knowledge is always significantly decreasing with the distance to the road (see Online appendix). These results are suggestive that the definition of the knowledge variable we adopt throughout the paper is not a major source of bias in our estimates.

5.1.2 Condom Access

To complete the analysis of the supply of preventive measures, we investigate whether road infrastructure facilitates access to and the ability to buy a (male) condom. Two types of questions

[7]We propose three alternative measurements: (i) we use the same idea as above with the exception that we attribute the same score for ignorance and incorrect knowledge; (ii) the second measurement relies on the principal component analysis method to generate a score of knowledge based on the six initial variables; and (iii) the third measurement, often called "comprehensive knowledge" is a binary variable equal to one if the individual gives correct answer for every question and zero otherwise

are used for this analysis. On one hand, 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 non medical private sector (e.g. shops). On the other hand, the sub-sample of respondents who report knowing where to find condoms are asked about their ability to buy one through the question: *"If you wanted to, could you yourself get a condom?"*.

In Table XII, we estimate alternatively the probability of knowing any place where one could purchase a condom (column (1)), the likelihood of citing at least one place from the public medical sector (column (2)), from the medical private sector (column (3)) or from the non medical private sector (column (4)), and eventually, we estimate the ability to buy a condom (column (5)).

The probit estimates displayed in column (1) suggest that the ability of knowing of at least one place where one can find a condom decreases with the distance to a road. At mean value of all covariates, the predicted probability of knowing one place is only equal to 30%. The marginal effect of road distance is equal to -0.0063 suggesting that at mean values, a one standard deviation increase in road distance decreases the likelihood of citing at least one place by 0.9 percentage point. While the predicted probability of citing at least one place is 67% over the whole sample, it is equal to (i) 77% for someone who lives at less than 5 kilometers away from the nearest road, (ii) 65% for someone who live between 5 and 10 kilometers away from a road, (iii) 61% for someone who live between 10 and 15 kilometers away and (iv) 51% for someone who live at more than 100 kilometers away from a road.

Interesting results are found when the analysis distinguishes the type of places known by the respondents (see columns (2)-(4)). It turns out that the distance to a road increases the likelihood of citing a place from the public sector, and reduces the likelihood of citing a place from either the medical or the non medical private sector. At mean values, a one standard deviation increase in road distance increases the likelihood of citing a place from the public sector by 1.8 percentage point and reduces the likelihood of knowing a place from the private medical and non medical sector by 1.7 and 1.6 percentage point respectively. 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. Column (5) shows that road distance is significantly and negatively associated with the ability to get a condom. The marginal effect being equal to -0.0075, increasing the distance to the road by one standard deviation (that is about 3.4 kilometers) approximately reduces the declared ability to buy a condom by 1.1 percentage point.

Our findings are suggestive that although spatial inequalities are persistent in terms of access to preventive measures, they do not drive the association between HIV infection and road proximity. 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 Road Distance and the Demand for Prevention

To examine whether the observed relationship between proximity to road and HIV infection is driven by differences of risk taking behaviors, we investigate whether on average, 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 casual sex, or both.

In the Demographic and Health Surveys, respondents are asked to self-report the nature of their relation 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. We estimate the choice of condom use and the choice of partner (casual v.s. usual) in a bivariate probit specification, which is supported by the Wald test of correlation between the error terms from the two equations.

Table XIII reports the estimated coefficients from the bivariate probit model. Column (1) considers the choice of the last sexual partner and estimates the probability of having sex with one's spouse (or cohabiting partner). Results suggest that the likelihood of having sex with one's spouse increases with the distance to a road. The marginal effect of road distance is equal to 0.0309 suggesting that a one standard deviation increase in road distance increases the likelihood of having sex with one's spouse by 4.6 percentage points. Thus, people living in proximity to a road are more likely to have had their last sexual intercourse with a casual partner than their counterparts living further away from a road. Results are supportive of the fact that people prefer to have more sex if road proximity makes it possible. We find 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. In column (2), condom use significantly decreases with the distance to the nearest road, so that at mean value, a one standard deviation increase in road distance reduces the expected probability of using a condom by 5.1 percentage points. It seems that the effect of road proximity in enhancing the use of condoms goes beyond the condom access it induces because the effect holds even controlling for whether the individual knows at least one place where one could find a condom. Note that the latter variable turns out to have a unexpected negative effect on condom use, while HIV/AIDS-knowledge has the expected effect on condom use and appears as a pre-requisite for wearing a condom.

Table XIV reports the four joint probabilities. Columns (1) and (2) report the probability of having sex with one's spouse with and without condom respectively. The agents living in proximity to a road are found to be more likely to have sex with their spouse with a condom and less likely to have sex with their spouse without a condom than their counterparts living further away from a road. Columns (3) and (4) deal with extramarital sexual relations and show that the probability of having sex with a casual partner, both with or without using a condom, decreases with the distance to a road. A one standard deviation increase in road distance reduces the likelihood of having sex with a casual partner with condom by 0.16 percentage point and without condom by 0.39 percentage point.

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. We find that access to information increases the demand for condom and the agents seem to choose the preventive measure that hurts their utility the least. It appears that in proximity to a road, condom use is preferred to faithfulness.

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 likelihood of having 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 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. 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. Mobility appears to play a role in the observed relationship in two respects. Firstly, the risk-reducing role of road distance holds mostly for individuals who are constrained in their moves as they do not own a car, a bicycle or a motorcycle. Secondly, 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 traffic flows passing through the road. 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 could 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 although it also increases the access to protection (and hence reduces the cost of protection) is inconsistent with the persistent wisdom that ignorance and lack of access to preventive measures are driving the spread of HIV in Africa. Thus the results support the idea that the incentives to invest in health remain too low in these countries and that the demand for risky sex depends upon the individual'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 features. First access to condoms and to information about the importance of using them have facilitated their use. Second people express their preferences towards the set of available preventive measures and choose the one that hurts their utility the least. The individuals 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. Firstly, our findings are suggestive of persistent disparities in the access to information and to protective devices such that it might be worth considering drawing programs to fight the spread of the epidemic that are specific to accessible and to remote areas. Secondly, results suggest that knowledge and condom availability are necessary but not sufficient to prevent individuals from being HIV infected and that the deliberated risk-taking is one dimension that has been partly ignored when drawing public programs while it should be taken into account. There is a need to provide people more incentives to preserve their health status, and especially, to invest in self-protection. Thirdly, in terms of road investment, we do not promote the autarky or the freeze of investment in communication infrastructure; however our empirical findings are suggestive of the existence of additional costs and benefits of roads that were not pointed out beforehand, since the road network makes individuals more aware of the risk of HIV transmission but also at higher risk of infection.

References

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

[2] Azam, Jean-Paul. 2007. *Trade, Exchange Rate and Growth in Sub-Saharan Africa*. Cambridge: Cambridge University Press.

[3] Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2009. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." Working Paper.

[4] Barbieri, Katherine, Omar Keshk, and Brian Pollins. 2008. "Correlates of War Project Trade Data Set Codebook, Version 2.0. Online: http://correlatesofwar.org.

[5] Batzilis, Dmitrios, Taryn Dinkelman, Emily Oster, Rebecca Thornton, and Deric Zanera. 2010. "New Cellular Networks in Malawi: Correlates of Service Rollout and Network Performance." NBER Working Paper 16616.

[6] Buys, Piet, Uwe Deichmann, and David Wheeler. 2010. "Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa." *Journal of African Economies* ,19 (3): 399-432.

[7] de Loenzien, Myriam. 2005. *Connaissances et attitudes face au VIH/SIDA*. Paris: L'Harmattan, Collection Populations.

[8] Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna. 2008. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." IFPRI Discussion Paper 00840.

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

[10] Dillon, Andrew, Manohar Sharma, Xiaobo Zhang. 2011. "Estimating the Impact of Rural Investments in Nepal." *Food Policy*, in press doi:10.1016/j.foodpol.2010.11.022.

[11] Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review,* forthcoming.

[12] Donalson, Dave. 2010. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." Mimeo. http://econ-www.mit.edu/files/6038.

[13] Duflo, Esther, and Rohini Pande. 2007. "Dams." *The Quarterly Journal of Economics*, 122 (2): 601-646.

[14] ESRI, 2007. ESRI Data and Maps 2007 Data Update, DVD.

[15] Evans, William N., John D. Graham. 1991. "Risk reduction or risk compensation? The case of mandatory safety belt use laws." *Journal of Risk and Uncertainty*, 4: 61-73.

[16] Gachassin, Marie, Boris Rajman, and Gael Raballand. 2010. "The Impact of Roads on Poverty Reduction: A Case Study of Cameroon." Policy Research Working Paper 5209, World Bank.

[17] Gibson, John, and Scott Rozelle. 2003. "Poverty and Access to Roads in Papua New Guinea." *Economic Development and Cultural Change*, 52 (1): 159-185.

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

[19] Jacoby, Hanan G. 2000. "Access to Markets and the Benefits of Rural Roads." *The Economic Journal*, 110 (465): 713-737.

[20] Jacoby, Hanan G., Bart Minten. 2009. "On Measuring the Benefits of Lower Transport Costs." *Journal of Development Economics*, (89): 28-38.

[21] Kazianga, Harounan. 2005. "HIV/AIDS Prevalence and the Demand for Safe Sexual Behavior: Evidence from West Africa." Unpublished Manuscript: http://www.hkazianga.org/Ppapers/KaziangaHIV_Ver3.pdf.

[22] Khandker, Shahidur K., Zaid Bakht, and Gayatri B. Koolwal. 2009. "The Poverty Impact of Rural Roads: Evidence from Bangladesh." *Economic Development and Cultural Change*, 57 (4): 685-722.

[23] Meekers, Dominique. 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.

[24] Morris, Chester N., and Alan G. Ferguson. 2007. "Sexual and Treatment-Seeking Behaviour for Sexually Transmitted Infection in Long-Distance Transport Workers in East Africa." *Sexually Transmitted Infections*, 83: 242-245.

[25] Nunn, Nathan, and Diego Puga. 2011. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics*, forthcoming.

[26] Oruboloye, IO., Pat Caldwell, and John C. Caldwell. 1993. "The role of high-risk occupations in the spread of AIDS: truck drivers and itinerant market women in Nigeria." *International Family Planning Perspectives*, 19: 43-48.

[27] Oster, Emily. 2011. "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa." *Journal of the European Economic Association*, 9 (2), forthcoming.

[28] Peltzman, Sam. 1975. "The Effects of Automobile Safety Regulation." *Journal of Political Economy*, 83 (4): 677-726.

[29] Peterson, Steven, George Hoffer, and Edward Millner. 1994. "Are Drivers of Airbag-Equipped Cars More Aggressive? A test of the offsetting behaviour hypothesis." Journal of Law and Economics, 38, 251-265.

[30] Rakwar, Joel, Ludo Lavreys, Mary Lou Thompson, Denis Jackson, Job Bwayo, Salim Hassanali, Kishorchandra Mandaliya, Jeckoniah Ndlinya-Achola, and Joan Kreiss. 1999. "Cofactors for the Acquisition of HIV-1 among Heterosexual Men: Prospective Cohort Study of Trucking Company Workers in Kenya." *AIDS* 13: 607-614.

[31] Ramjee, Gita, and Eleanor Gouws. 2002. "Prevalence of HIV Among Truck Drivers Visiting Sex Workers in KwaZulu-Natal, South Africa." *Sexually Transmitted Diseases*, 29 (1): 44-49.

[32] Riley, Shawn J., Stephen D. DeGloria, and Robert Elliot. 1999. "A terrain ruggedness index that quantifies topographic heterogeneity." *Intermountain Journal of Sciences*, 5(1-4): 23-27.

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

[34] Straub, Stephane, Charles Vellutini, and Michael Warlters. 2008. "Infrastructures and Economic Growth in East Asia." Policy Research Working Paper n4589. World Bank.

[35] Tanser, Frank, David LeSueur, Geoff Solarsh, and David Wilkinson. 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.

[36] UNAIDS. 2004. *2004 Report on the Global AIDS Epidemic*. 4th global report.

Appendix

Table I: Sample size

	All obs.	Women	nb clusters	nb households women survey	nb households men survey- 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	4,234
Malawi	14,679	11,503	521	13,644	4,580
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 eligible for HIV testing and men survey

Table II: Summary statistics

Variable	**All**	**CMR**	**ETH**	**GHA**	**KEN**	**MWI**	**ZWE**
HIV+	0.079	0.054	0.018	0.021	0.065	0.124	0.179
HIV testing	0.154	0.191	0.074	0.092	0.156	0.149	0.22
know s^{one} HIV+	0.45	0.45	0.14	0.39	0.74	0.66	0.30
HIV/AIDS-knowledge	3.57	3.29	2.95	3.48	4.52	3.43	4.11
	[2.33]	[2.28]	[2.56]	[2.39]	[1.73]	[2.28]	[2.11]
age	28.60	28.55	28.77	30.26	28.53	28.15	27.81
	[10.16]	[10.46]	[10.27]	[10.91]	[9.87]	[9.44]	[9.94]
women	0.6606	0.6659	0.7005	0.5305	0.6946	0.7836	0.55
urban	0.3332	0.4993	0.2994	0.3988	0.33	0.14	0.35
no education	0.2462	0.1755	0.5440	0.2851	0.14	0.2087	0.0316
primary educ	0.3795	0.3997	0.2439	0.1848	0.5313	0.6271	0.3181
secondary educ	0.3378	0.3935	0.1815	0.4898	0.2501	0.1561	0.6101
higher educ	0.0365	0.0313	0.0305	0.0404	0.0816	0.0081	0.0401
catholic	0.1753	0.3892	0.0099	0.1587	0.2383	0.2182	0.1029
protestant	0.4670	0.3539	0.1533	0.5353	0.6044	0.6158	0.6790
muslim	0.1696	0.1720	0.3269	0.1941	0.1209	0.1518	0.0075

Note: Standard deviations are in brackets

Table III: Distance to the nearest primary road in kilometers

Country			percentiles			
	Mean	**Std dev.**	**25th**	**50th**	**75th**	**Max**
All	24.50	32.68	2.44	11.06	35.94	287.83
Cameroon	20.11	28.75	1.33	5.58	28.96	160.42
Ethiopia	26.98	35.53	0	11.82	39.01	192.28
Ghana	25.32	32.94	2.68	10.14	36.64	177.71
Kenya	22.26	37.60	0	10.29	24.78	287.83
Malawi	19.99	22.09	2.73	10.35	32.21	94.89
Zimbabwe	30.88	35.31	2.76	13.82	50.83	172.08

Table IV: Road and HIV-risk- Probit Estimates

	(1)		(2)		(3)	
Road distance†	-0.0698***	(0.009)	-0.0704***	(0.009)	-0.0517***	(0.010)
1 for woman			0.2100***	(0.019)	0.2022***	(0.019)
married			0.4893***	(0.032)	0.5044***	(0.033)
prev. married			1.0224***	(0.037)	1.0402***	(0.038)
age			0.0111***	(0.001)	0.0107***	(0.001)
primary educ			0.1974***	(0.037)	0.1715***	(0.037)
secondary educ			0.2677***	(0.041)	0.2013***	(0.042)
higher educ			0.0690	(0.065)	-0.0325	(0.067)
catholic			0.0403	(0.046)	0.0507	(0.046)
protestant			0.0140	(0.043)	0.0271	(0.043)
other religion			0.1720***	(0.046)	0.1913***	(0.046)
HIV/AIDS-knowledge			0.0035	(0.005)	0.0002	(0.005)
wpoorer					0.0492	(0.038)
wmiddle					0.1524***	(0.038)
wricher					0.2551***	(0.039)
wrichest					0.2637***	(0.045)
constant	-1.3408***	(0.082)	-2.4365***	(0.091)	-2.5919***	(0.095)
Regional FE	yes		yes		yes	
N	53,239		50,842		50,842	
Number of clusters	2,703		2,703		2,703	

Note: Robust standard errors clustered at the community level in parentheses

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

† equals to log(1+distance to the paved nearest road)

Omitted categories: men, muslim, no education, single, poorest

Table V: Road and HIV-risk, controlling for community-level characteristics
Probit Estimates

	(1)	(2)	(3)	(4)
Road distance†	-0.0515***	-0.0438***	-0.0404***	-0.0471***
	(0.010)	(0.011)	(0.010)	(0.010)
Population density	0.0000			
	(0.000)			
Dist to nearest city, log		-0.0007**		
		(0.000)		
% of richest people			0.2989***	
			(0.063)	
1 if urban				0.0761**
				(0.035)
Regional FE	yes	yes	yes	yes
Individual covariates	yes	yes	yes	yes
N	50,824	50,842	50,842	50,842
Number of clusters	2,703	2,703	2,703	2,703

Note: Robust standard errors clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† equals to log(1+distance to the paved nearest road)
Controls include gender, age, marital status, educational attainment, wealth, religion, HIV/AIDS-knowledge, regional dummies

Table VI: Non-random Road Placement
Using land slope and terrain ruggedness as IV

	(1)	(2)		(3)	
	LPM	2SLS		2SLS	
		IV = slope		IV = ruggedness	
		1st stage	2st stage	1st stage	2st stage
		(a)	(b)	(a)	(b)
Road distance†	-0.0063***		-0.0410**		-0.0523**
	(0.001)		(0.021)		(0.026)
Slope		0.0163***		n.a.	
		(0.001)			
Ruggedness		n.a.	0.0006***		
			(0.000)		
Regional FE	yes	yes		yes	
Covariates	yes	yes		yes	
N	50,951	50,529		50,766	

Note: Robust standard errors clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† equals to log(1+distance to the paved nearest road)
Other controls include urban, gender, age, marital status, educational attainment, wealth, religion, HIV/AIDS-knowledge, regional dummies

Table VII: Non-random Individual Placement
Probit, LPM and 2SLS Estimates

Coefficient on road distance	(1) Probit	(2) LPM	(3) 2SLS IV= slope	(4) 2SLS IV=ruggedness
Panel A: 1 if ever been tested for HIV				
Ever been tested	-0.0372*	-0.0070*	-0.1208**	-0.0886*
	(0.020)	(0.004)	(0.048)	(0.046)
Never been tested	-0.0489***	-0.0060***	-0.0242	-0.0433*
	(0.011)	(0.001)	(0.020)	(0.026)
Wald test $H_0 : \beta_{ever} = \beta_{never}$	0.32			
Panel B: Removing the new movers				
Removing the new Movers	-0.0534***	-0.0061***	-0.0302	-0.0452*
	(0.012)	(0.001)	(0.019)	(0.027)
			[0.103]	
Panel C: 1 if born in the current place of residence				
Non migrants	-0.0701***	-0.0073***	-0.0320	-0.0477
	(0.014)	(0.002)	(0.025)	(0.038)
Migrants	-0.0286**	-0.0045***	-0.0558**	-0.0632**
	(0.012)	(0.002)	(0.028)	(0.029)
Wald test $H_0 : \beta_{nonmigrant} = \beta_{migrant}$	5.96**			
Panel D: Removing the group of selection drivers				
Removing Selection drivers	-0.0528***	-0.0072***	-0.0444*	-0.0596**
	(0.011)	(0.001)	(0.023)	(0.030)

Note: Robust standard errors, clustered at the community level in parentheses; P-values in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table VIII: Separated effects of distance to a road on infection
Probit and 2SLS Estimates

Coefficient on road distance	(1) Probit	(2) 2SLS IV= slope	(3) 2SLS IV=ruggedness
Panel A: By gender			
Women	-0.0455***	-0.0614**	-0.1090**
	(0.012)	(0.028)	(0.051)
Men	-0.0510***	-0.0212	-0.0071
	(0.014)	(0.021)	(0.020)
Wald test $H_0 : \beta_{women} = \beta_{men}$	0.11		
Panel B: by location			
Urban	-0.0329**	-0.0163	-0.0300
	(0.016)	(0.026)	(0.019)
Rural	-0.0512***	-0.0527	-0.0537
	(0.012)	(0.036)	(0.043)
Wald test $H_0 : \beta_{urban} = \beta_{rural}$	1.10		

Note: Robust standard errors, clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IX: Mobility Scenario
Probit and 2SLS Estimates

Coefficient on road distance	(1) Probit	(2) 2SLS IV= slope	(3) 2SLS IV=ruggedness
Panel A: whether the household owns a car			
Car	0.0300	-0.0269	-0.0575**
	(0.051)	(0.031)	(0.028)
No car	-0.0534***	-0.0456*	-0.0572*
	(0.011)	(0.025)	(0.033)
Panel B: whether the household owns a bike			
Bike	-0.0290	-0.3201	-0.2585
	(0.020)	(0.390)	(0.226)
No bike	-0.0562***	-0.0281	-0.0481*
	(0.012)	(0.018)	(0.027)
Panel C: whether the household owns a motorcycle			
Motorcycle	-0.0805	-0.0440	-0.1290
	(0.050)	(0.046)	(0.196)
No motorcycle	-0.0476***	-0.0439*	-0.0627*
	(0.011)	(0.023)	(0.033)
Panel D: Interaction with Traffic flows			
Road distance	-0.1300***	-0.0084	-0.0631
	(0.031)	(0.028)	(0.091)
Traffic flows	-0.0031	0.0231*	-0.0048
	(0.019)	(0.012)	(0.041)
Interaction term	0.0161***	-0.0079*	0.0026
	(0.006)	(0.004)	(0.015)

Note: Robust standard errors, clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table X: Robustness checks: Threats to validity

	(1) Probit	(2) 2SLS IV= slope	(3) 2SLS IV=ruggedness
Panel A: Acceptation and distance to road			
Dependent variable: Probability of showing-up for the test			
Road distance	0.1008***		
	(0.009)		
Panel B: Random reallocation of communities			
Dependent variable: HIV-infection			
Road distance	-0.0390***	-0.0602*	-0.0666**
	(0.012)	(0.033)	(0.032)

Note: Robust standard errors, clustered at the community level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table XI: Road Distance and HIV/AIDS-Knowledge
OLS estimates

	(1)		(2)		(3)	
Road distance†	-0.0262**	(0.011)	-0.0238**	(0.011)	-0.0174	(0.011)
1 for woman	-0.5043***	(0.024)	-0.5072***	(0.025)	-0.4041***	(0.026)
married	0.0442*	(0.023)	0.0551**	(0.024)	0.0591**	(0.024)
prev. married	0.1183***	(0.033)	0.0939***	(0.035)	0.1261***	(0.035)
age	0.0038***	(0.001)	0.0021**	(0.001)	0.0018*	(0.001)
1 if urban	0.1882***	(0.034)	0.1581***	(0.034)	0.1383***	(0.034)
primary educ	0.7101***	(0.030)	0.6450***	(0.031)	0.5185***	(0.032)
secondary educ	1.4622***	(0.035)	1.3811***	(0.036)	1.1375***	(0.038)
higher educ	1.7381***	(0.047)	1.6161***	(0.049)	1.3631***	(0.051)
wpoorer	0.2099***	(0.032)	0.1897***	(0.034)	0.1289***	(0.034)
wmiddle	0.3481***	(0.033)	0.3151***	(0.035)	0.2257***	(0.035)
wricher	0.5235***	(0.035)	0.4847***	(0.036)	0.3323***	(0.037)
wrichest	0.8004***	(0.039)	0.7316***	(0.040)	0.5026***	(0.042)
catholic	0.2503***	(0.043)	0.2229***	(0.043)	0.2161***	(0.042)
protestant	0.2641***	(0.042)	0.2408***	(0.041)	0.2420***	(0.041)
other religion	0.1511***	(0.047)	0.0859*	(0.046)	0.0873*	(0.045)
knows s^{one} HIV+			0.2125***	(0.021)	0.1780***	(0.020)
ever tested for HIV			0.1794***	(0.022)	0.1556***	(0.022)
magazines less than once a week					0.3541***	(0.024)
magazines at least once a week					0.2381***	(0.028)
magazines almost every day					0.1605***	(0.039)
radio less than once a week					0.2636***	(0.033)
radio at least once a week					0.3085***	(0.034)
radio almost every day					0.3891***	(0.030)
tv less than once a week					0.1359***	(0.030)
tv at least once a week					0.0113	(0.035)
tv almost every day					0.1670***	(0.031)
constant	2.2177***	(0.110)	2.2531***	(0.111)	2.0737***	(0.111)
Regional FE	yes		yes		yes	
N	81,246		70,115		69,870	
Number of clusters	2,703		2,703		2,703	

Note: Robust standard errors clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† equals to log(1+distance to the paved nearest road)
Omitted categories: rural, men, muslim, no education, single, poorest and; for equation (3): never listen to the radio, never watch tv, never read newspapers.

Table XII: Road Distance and Condom Access
Probit model estimates

	Dependent variable: 1 if the individual knows a place where one could find a condom				
	any place	public sector	private health sector	other private sector	ability to buy
	(1)	(2)	(3)	(4)	(5)
Road distance†	-0.0186**	0.0352***	-0.0388***	-0.0313***	-0.0329***
	(0.007)	(0.010)	(0.010)	(0.008)	(0.009)
1 for woman	-0.5095***	0.0454**	-0.0809***	-0.4535***	-0.7895***
	(0.020)	(0.018)	(0.019)	(0.018)	(0.028)
married	0.3445***	0.3070***	0.1800***	0.0519***	0.4333***
	(0.017)	(0.017)	(0.017)	(0.016)	(0.023)
prev. married	0.3127***	0.2417***	0.1785***	0.1261***	0.4296***
	(0.025)	(0.024)	(0.024)	(0.023)	(0.034)
age	-0.0035***	0.0034***	-0.0012	-0.0081***	-0.0061***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
1 if urban	0.1909***	0.0933***	0.1938***	0.1478***	0.0622**
	(0.027)	(0.030)	(0.028)	(0.027)	(0.027)
primary educ	0.5563***	0.4977***	0.3931***	0.3944***	0.1048***
	(0.019)	(0.020)	(0.024)	(0.021)	(0.027)
secondary educ	0.9599***	0.7342***	0.7344***	0.6124***	0.2866***
	(0.023)	(0.024)	(0.026)	(0.025)	(0.031)
higher educ	1.2614***	0.7209***	1.0640***	0.7791***	0.4629***
	(0.046)	(0.037)	(0.039)	(0.037)	(0.051)
wpoorer	0.0543***	0.0426**	0.0734***	0.0725***	-0.0498
	(0.020)	(0.021)	(0.024)	(0.022)	(0.032)
wmiddle	0.0952***	0.0534**	0.1404***	0.1292***	-0.0232
	(0.021)	(0.023)	(0.026)	(0.022)	(0.031)
wricher	0.1837***	0.0421	0.2640***	0.2186***	0.0120
	(0.023)	(0.026)	(0.030)	(0.024)	(0.032)
wrichest	0.4580***	0.0825**	0.5253***	0.4458***	0.1086***
	(0.028)	(0.032)	(0.033)	(0.028)	(0.037)
catholic	0.0769***	0.0521*	-0.0189	0.0328	0.0291
	(0.029)	(0.031)	(0.033)	(0.031)	(0.036)
protestant	0.0846***	0.0668**	-0.0278	0.0223	-0.0040
	(0.026)	(0.029)	(0.031)	(0.029)	(0.032)
other religion	0.0419	0.0306	-0.0196	0.1116***	0.1335***
	(0.028)	(0.031)	(0.033)	(0.030)	(0.042)
HIV/AIDS-knowledge	0.0921***	0.0567***	0.0666***	0.0672***	0.0484***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
constant	-0.8024***	-2.7058***	-1.9190***	-0.5187***	0.3965***
	(0.090)	(0.102)	(0.114)	(0.094)	(0.122)
regional FE	yes	yes	yes	yes	yes
N	76,152	76,152	76,152	76,152	42,747
Number of clusters	2,703	2,703	2,703	2,703	2,703

Note: Robust standard errors, clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† equals to log(1+distance to the paved nearest road)
Omitted categories: rural, men, muslim, no education, single, poorest

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

	(1)		(2)	
	1 if sex with spouse		1 if condom use	
Road distance†	0.0309**	(0.012)	-0.0346***	(0.010)
know where to find condom			-0.4413***	(0.037)
1 for woman	0.3325***	(0.036)	-0.5009***	(0.023)
married	4.4203***	(0.078)	-1.3011***	(0.028)
prev. married	1.8459***	(0.080)	-0.2720***	(0.038)
age	0.0181***	(0.002)	-0.0147***	(0.001)
1 if urban	-0.1212***	(0.037)	0.0826***	(0.029)
primary educ	-0.1248**	(0.049)	0.2736***	(0.043)
secondary educ	-0.2243***	(0.053)	0.5446***	(0.047)
higher educ	-0.1571**	(0.075)	0.6953***	(0.061)
wpoorer	0.0233	(0.047)	0.0750*	(0.041)
wmiddle	-0.0577	(0.048)	0.1092***	(0.040)
wricher	-0.1334***	(0.049)	0.2660***	(0.040)
wrichest	-0.1019*	(0.055)	0.3733***	(0.046)
catholic	-0.1577***	(0.059)	0.0734*	(0.044)
protestant	-0.0450	(0.056)	0.0039	(0.042)
other religion	-0.2012***	(0.063)	0.0135	(0.049)
HIV/AIDS-knowledge	-0.0206***	(0.006)	0.0479***	(0.005)
constant	-3.1648***	(0.195)	0.1693	(0.119)
Regional FE	yes			
Wald test of $\rho = 0$	chi2(1)=488.28***			
N	48,876			
Number of clusters	2,703			

Note: Robust standard errors clustered at the community level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
† equals to log(1+distance to the paved nearest road)
Omitted categories: rural, men, muslim, no education, single, poorest

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]	
Road distance†	-.0023***	(.000)	.0060***	(.002)	-.0011***	(.000)	-.0026**	(.001)
woman	-.0402***	(.003)	.0843***	(.005)	-.0182***	(.001)	-.0258***	(.004)
married	.0265***	(.001)	.9455***	(.003)	-.2747***	(.008)	-.6974***	(.009)
prev. married	-.0083***	(.003)	.0831***	(.004)	-.0139***	(.001)	-.0609***	(.003)
age	-.0009***	(.000)	.0031***	(.000)	-.0006***	(.000)	-.0016***	(.000)
1 if urban	.0047*	(.002)	-.0199***	(.005)	.0037***	(.001)	.0114***	(.004)
primary educ	.021***	(.004)	-.0364***	(.007)	.0075***	(.001)	.0078	(.005)
secondary educ	.0479***	(.005)	-.0766***	(.008)	.0165***	(.002)	.0123**	(.006)
higher educ	.0907***	(.012)	-.1119***	(.015)	.0232***	(.004)	-.0021	(.007)
wpoorer	.0068*	(.004)	-.0040	(.006)	.0010	(.001)	-.0038	(.005)
wmiddle	.0084**	(.004)	-.0155*	(.007)	.0032**	(.001)	.0040	(.005)
wricher	.0220***	(.004)	-.0391***	(.007)	.0083***	(.002)	.0088	(.005)
wrichest	.0341***	(.005)	-.0469***	(.008)	.0099***	(.002)	.0029	(.006)
catholic	.0033	(.004)	-.0238***	(.009)	.0043**	(.002)	.0162**	(.007)
protestant	-.0004	(.003)	-.0050	(.007)	.0008	(.001)	.0046	(.006)
other religion	-.0025	(.004)	-.0245**	(.010)	.0038**	(.002)	.0231***	(.008)
HIV/AIDS-knowledge	.0036***	(.000)	-.0060***	(.001)	.0012***	(.000)	.0013**	(.001)
know where to find condom	-.0428***	(.004)	.0428***	(.004)	-.0087***	(.001)	.0087***	(.001)

Note: Robust standard errors clustered at the community level in parentheses

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

† equals to log(1+distance to the paved nearest road)

y_1 equals 1 if the respondent's last partner is one's spouse, 0 otherwise

y_2 equals 1 if the respondent reports condom use, 0 otherwise

Omitted categories: rural, men, muslim, no education, single, poorest