Women's Position in Ancestral Societies and Female HIV: The Long-Term Effect of Matrilineality in Sub-Saharan Africa *

JOB MARKET PAPER

Jordan Loper†

March 30, 2021

Most Recent Version Here

Abstract

Can contemporary female HIV rates be traced back to women's position in ancestral societies ? In matrilineal kinship organizations, lineage and inheritance are traced through female members and children integrate the kin group of their mother rather than their father. Ethnographic accounts suggest that in matrilineal kinship structures, women benefit from greater autonomy and spousal cooperation is reduced. I test the hypothesis that, by affecting women's sexual and contraceptive behaviours, ancestral matrilineality has a causal impact on the prevalence of female HIV. Using variation in ethnic groups' ancestral kinship organizations within Sub-Saharan African countries, I find that females originating from ancestrally matrilineal ethnic groups are today more likely to be infected by HIV. This finding is robust to the inclusion of subnational fixed effects, as well as a large set of cultural, historical, geographical, and environmental factors. I find consistent results using a number of alternative estimation strategies, including a geographic regression discontinuity design at ethnic boundaries and an instrumental variable strategy. Matrilineal females' riskier sexual and contraceptive behaviours constitute the main explanatory mechanisms. These results call for policies moving beyond the "one-size-fits-all" strategy and taking local cultural contexts into account.

Keywords: Kinship systems, matrilineality, cultural persistence, HIV, sexual behaviour, gender

JEL Classification: D13, D91, I12, I15, J12, N37, Z13

*I thank Alberto Alesina, Isaiah Andrews, Duman Bahrami-Rad, Jean-Marie Baland, Anke Becker, Nicolas Berman, Mathieu Couttenier, Melissa Dell, Yannick Dupraz, Victor Gay, Catherine Guirkinger, Joseph Henrich, Asim Khwaja, Jacob Moscona, Lorenzo Rotunno, Mohamed Saleh, Marc Sangnier, Heather Sarsons, Jonathan Schulz, Paul Seabright, Liyang Sun, Michèle Tertilt, Edoardo Teso, Alessandra Voena, Romain Wacziarg and participants at various seminars and conferences for helpful comments and suggestions. I also especially thank Nathan Nunn and Roberta Ziparo for their warm support and guidance. This work was supported by the French National Research Agency Grant ANR-17-EURE-0020. Part of this paper was written while I was visiting Harvard University and I thank them for their hospitality.

†University of Lyon, ENS Lyon, GATE Lyon/St-Etienne (email: jordan.loper@ens-lyon.fr)

1 Introduction

Every four minutes three young women become infected with HIV in the world. In 2018, the number of women in the world living with HIV amounted to 18.8 millions (UNAIDS, 2018b). With 14.5 millions of women living with HIV[1], representing about 80% of the worldwide female HIV positive population, Sub-Saharan Africa is the most affected region in the world.[2] Strikingly, it is the only place in the world where more women than men live with HIV: adult female HIV positive population was about 1.5 larger than male HIV positive population[3], and adult females represented about 57% of new adult infections in 2017.[4] This phenomenon has been referred to "Feminization of HIV" in Sub-Saharan Africa. Although this gender difference is widely acknowedged and partly driven by biological factors making women more suceptible to contract the virus than men, little research has sought to explain the variation in female HIV rates across the Sub-Saharan African continent.[5] While most of research has considered sexual and contraceptive behaviours to understand variations in HIV prevalence, one long-term determinant of such behaviours has been overlooked so far: *ethnicity* and related *transmitted cultural traits*. Understanding such cultural channel should help design policies that are more tailored to local cultural contexts in Sub-Saharan Africa, a continent marked by great ethnical diversity.

Researchers are increasingly coming to understand the role of transmitted cultural traits on contemporary individuals' behaviour (Giuliano and Nunn, 2019). Ethnic groups' ancestral kinship organizations, which are fundamental social institutions determining group membership and social obligations in Sub-Saharan Africa, have been recently emphasized as deep-rooted determinants of individuals' behaviour (Moscona et al., 2020). Recent evidences have also highlighted the long-term infuence of ancestral kinship organizations on gender norms and related gender-specific behaviours. Building on the anthropological literature, Lowes (2018a) provides experimental evidence that *matrilineality* undermines spousal cooperation within household. While in *patrilineal* kinship organizations children integrate their father's kin group and inheritance can only be passed on to children of male group members, in *matrilineal* kinship organizations group membership and inheritance are traced through female members. In other words, matrilineal males do not transmit to their biological children but to their sister's children. Further, in *patrilineal* societies a wife is effectively incorporated into the lineage of her husband because she is not relevant to her kin group for determination of lineage or inheritance, thereby reducing her ability to rely on her own kin group in the case of separation or conflict. In contrast, in *matrilineal* societies husbands and wives maintain strong allegiances with their own kin group, allowing matrilineal wifes to benefit from an improved outside option relative to their patrilineal counterparts, thereby reducing matrilineal men's authority over their spouses (Lowes, 2018a). Along these lines, one may expect matrilineality to also affect women's sexual and contraceptive behaviours.

In this paper, I examine ancestral matrilineality as a deep-rooted determinant of contemporary female HIV prevalence in Sub-Saharan Africa.[6] To do so, I use a representative random sample of individuals reporting their ethnicity and who were tested for HIV in 32 DHS (*Demographic Health Surveys*), across 18 Sub-Saharan African countries. Using information on individual's ethnicity, I link individuals to their ancestral ethnic group in the

[1] 11.2 millions in "Eastern and Southern Africa" and 3.3 millions in "West and Central Africa".
[2] These numbers are from `http://aidsinfo.unaids.org/`
[3] 9.5 millions males lived with HIV in Sub-Saharan Africa.
[4] Among the 1.02 millions newly infected adult individuals in Sub-Saharan Africa in 2017, 580,000 were females.
[5] Anderson (2018) and Bertocchi and Dimico (2019) constitute recent exceptions, exploring the infuence of legal systems and of historical exposure to the slave trade respectively.
[6] Henceforth, "ancestral" means dating back to before colonization.

Ethnographic Atlas, a worldwide anthropological database containing ethnographic information on cultural aspects and ways of life of ethnic groups prior to industrilization and colonial contact. This allows me to estimate the probability of testing positive for HIV as a function of whether an individual originates from an ancestrally matrilineal ethnic group. I find a higher probability of testing positive for HIV for females originating from ancestrally matrilineal ethnic groups.

Causally assessing the long-term impact of a cultural trait is an empirical challenge. Concern remains that the variation in ethnic group's ancestral kinship organizations exploited here may be correlated with important unobservables which are driving my main result, and has nothing to do with the channel of causality argued for in this paper. In my main identication strategy I exploit the variation in ethnic groups' ancestral kinship organizations within subnational regions in Sub-Saharan African countries. This enables me to include subnational region-survey fixed effects, and therefore control for a large set of national and subnational legal, institutional and economic confounding factors. The results are very similar when I control for an extensive set of covariates, including individual covariates which are plausible drivers of sexual and contraceptive behaviours; ethnic-group historical covariates capturing long-term confounding factors affecting contemporary gender norms; and village-level geographic covariates capturing geographical channels potentially shaping contemporary variation in HIV prevalence in Sub-Saharan Africa.[7] Finally, I show that my main results are not driven by differential selection into HIV testing, nor reflect differences in general health status, but are specific to sexually transmitted diseases.

Despite the care provided in controlling for a large range of observables, the OLS results might still suffer from the presence of omitted factors, including those that are unobservable to the researcher. Given this, I go one step further by formally testing for omitted bias, estimating Oster (2017) bias-adjusted lower bound coefficients. This exercise provide no evidence of an omitted bias in my OLS estimates.

Reverse causality may still remain a concern if matrilineality was an adaptative response in environments associated with more promiscuous sexual behaviours. To deal with this potential endogeneity issue, I exploit data on the GPS location of DHS villages as well as the digitized Murdock's map of ancestral ethnic groups in Africa (Figure 2) to implement alternative identification strategies. Computing the measure of distance between DHS villages and ancestrally matrilineal geographic areas, I conduct a geographic regression discontinuity (RD) design strategy at ethnic boundaries. Restricting attention to individuals living in DHS villages located close to an ancestral matrilineal ethnic boundary, I estimate the effect of living in a village located on the matrilineal side of the ancestral border, while controlling for geographic location/distance running variables. Under the assumption that unobervables vary smoothly over space, this empirical strategy allows me to estimate the effect of the variation in cultural trait on individuals living in similar environments. I find that RD estimates are qualitatively similar to OLS (though smaller in magnitudes, due to presumably blurry ancestral ethnic boundaries and potential spillovers at the border). The findings are robust to different bandwidths and to different methods of controlling for the

[7] Individual covariates are computed from the DHS and consist of age, marital status, education, employment status, urban status, number of children and religion. Ethnic-group historical covariates are computed from the *Ethnographic Atlas* and consist of women's historical participation in agriculture and related agricultural practices such as plough and pastoralism, ancestral marital norms such as bride price and polygyny, ancestral settlement complexity, ancestral presence of clans, ancestral juridictional institutions, ancestral modes of subsistence and year of observation of the ethnic group. Village-level geographic covariates are mainly computed from the DHS and consist of latitude, longitude, altitude, nightlight composite, population density, distance to lake or coastline, distance to nearest international border, distance to urban center, distance to and presence of nearby active mine, malaria incidence, vegetation index, length of growing season, and ethnic fractionalization and polarization.

two-dimensional running variables.[8]

To explain such long-term effect of ancestral matrilineality on contemporary female HIV in Sub-Saharan Africa, I provide evidence that, benefiting from a higher social status, bargaining power and subsequent higher sexual autonomy, matrilineal females adopt riskier sexual behaviour which are more conducive to HIV. I also show that women originating from ancestrally matrilineal ethnic groups are more likely to be HIV positive while having a HIV negative husband, suggesting extramarital channels of transmission of the virus. These results can be rationalized by evolutionary psychology theories, as will be seen in my conceptual framework in subsection 2.3. In addition, I underline that matrilineal females' higher contraception-related decision-making power translates into them substituting condom with long-term contraceptive methods. This means that, incidentally, matrilineal females substitute more protective contraceptive methods with less protective ones. Nevertheless, I also find evidence that when they have internalized the risk of transmission of the virus, matrilineal individuals are more likely to adopt condom as a contraceptive method. This last result calls for policies aiming at raising awareness about the risk associated with promiscuous behaviours, to fight against the spread of HIV for this population at high risk. Finally, I provide evidence that discard differences in access to condom and in sexual debuts as other explaining mechanisms. Indeed, I find that matrilineal females have in fact an easier access to condom and begin their sexual life later.

The findings of this paper contribute to a better understanding of the historical determinants of female HIV in Sub-Saharan Africa. A nascent literature has documented the effect of legal systems inherited from colonization (Anderson, 2018) as well as the adoption of polygynous practices following exposition to the slave trade (Bertocchi and Dimico, 2019).[9] These studies have advocated for empowering women as the main policy approach.[10] The findings of this paper complement these studies in several ways. First, I show that contemporary variation in female HIV rate in Sub-Saharan Africa can be traced back to ancestral (i.e. *before* colonization) kinship organizations. I therefore explore a more deep-rooted factor. Further, I provide evidence that the long-term effect of ancestral matrilineality holds true within common law countries, and for non-polygynous couples. In addition, I highlight different routes of infection to explain the higher prevalence of HIV for matrilineal females, namely extramarital routes. However, I also show that matrilineal individuals are more likely to use condom when they have properly internalized the risk of transmission of the virus by being aware of their seropositive status. Finally, I draw new conclusions and *complementary* policy recommendation: the findings of this paper call for information campaigns targeting *already empowered* women and raising awareness about the risk associated with promiscuous sexual behaviours. Further, the results of this paper call for policies moving beyond the "one-size-fits-all" strategy and taking local cultural contexts into account.

This paper also contributes to the literature exploring the contemporary determinants of the spread of HIV in

[8] In further robustness checks I also exploit information on GPS location of DHS villages and ethnic groups' boundaries to (1) perform an IV strategy instrumenting individual's ethnic group's ancestral matrilineality with the distance between the location of DHS villages and the nearest ancestral matrilineal ethnic boundary; and (2) estimate the average treatment over treated effect of being located on an ancestrally matrilineal geographic area on DHS villages' proportion of HIV positive females, matching villages with their nearest neighbor in an ancestrally non-matrilineal geographic area based on a large array of geographic observables. These exercises provide additional support to my main OLS findings.

[9] Cagé and Rueda (Forthcoming) also show that contemporary geographical variation in HIV across Sub-Saharan Africa was influenced by Protestant and Catholic missions and their health investments in the early 20th century.

[10] Anderson (2018) documents that female's property rights being weaker in common law countries, women suffer from lower intra-household bargaining power and related ability to impose safe sexual practices to their husbands. Ultimately, these women are mainly infected by their husbands and suffer from more HIV than women in civil law countries. Bertocchi and Dimico (2019) document that historical slave trade has fostered current polygynous practices, which are associated with more female's marital dissatisfaction. Thus, they find that females in polygynous union are more likely to adopt riskier sexual behaviour. In turn, this increases their likelihood to contract and transmit the virus, through the husband, to their faithful co-wives.

Africa, such as differences in sexual behaviour (Oster, 2005); exports and movement of people (Oster, 2012b); mine workers' migration (Corno and De Walque, 2012); or sexual debut (Case and Paxson, 2013). While I explore sexual and contraceptive behaviours as main mechanisms explaining the higher rates of HIV found for matrilineal female populations, I show that the long-term effect of ancestral matrilineality on contemporary female HIV is robust to movement of people and mine workers' migration alternative channels. Further, I provide evidence that age at sexual debuts can be discarded as an explanatory mechanism of the effect of ancestral matrilineality.

This paper further contributes to the literature aimed at understanding the influence of HIV risk perception on sexual and contraceptive behaviour. Experimental evidences have been brought that risk preferences (Björkman Nyqvist et al., 2018) and learning of HIV status (Thornton, 2008) matter in shaping such behaviours. Along these lines, some studies have explored how the adoption of risky sexual behaviour is influenced by expectation about own and partner's HIV status, transmission rates, as well as perceived impact of HIV on survival (Paula et al., 2014; Delavande and Kohler, 2015).[11] The result I provide in the paper on the heterogeneity in condom use by perception of the risk of transmitting the virus is fully in line with the results of this literature.

This paper also relates to the literature emphasizing cultural factors shaping contraception use, crucial for limiting the spread of HIV. In Sub-Saharan African context, fidelity norm and reproduction norm have been emphasized as two of the fundamental elements that guide spouses' condom use behaviour (Cordero-Coma and Breen, 2012; Chimbiri (2007)).[12] In the context of Bangladesh, Islam et al. (2009) underline that matrilineal *Garo* women contraceptive behaviour differs from their patrilineal *Bengali* counterparts, in that their current use of contraceptives is higher than at national level, but their use of condom is lower than at national level. I find a similar result when I explore differences in contraceptive methods used by matrilineal and patrilineal individuals.

Finally, this papers speaks to the literature exploring the long-term determinants of contemporary gender outcomes.[13] It contributes to previous studies that document several historical determinants such as pre-industrial agricultural practices (Alesina et al., 2013; Becker, 2018); pre-colonial customs about marriage patterns and living arrangements (Alesina et al., 2016); or slave trade (Teso, 2019). I contribute to this literature by showing that women originating from ancestrally matrilineal ethnic groups in Sub-Saharan Africa benefit from higher marriage outside option, intrahousehold bargaining power and sexual autonomy. This paper also links to the burgeoning literature assessing the influence of ethnic norms on women's well-being. Some studies have explored the effect of ethnic norms *per se*, such as matrilineality (Lowes, 2018a) or the practice of bride price (Lowes and Nunn, 2017). Other studies have assessed the influence of ethnic norms on the differential impact of nation-wide policies aimed at improving women's well-being (Ashraf et al., 2020b; Bargain et al., 2020). In this paper, I show that ancestral matrilineality has an ambiguous effect on women's well-being. Indeed, although I find that matrilineal women have a higher sexual freedom, I also find that they are more at risk of being HIV positive.

The remainder of the paper is organized as follows. In section 2, I provide an overview of ancestral matrilineality as well as contemporary female HIV in Sub-Saharan Africa, and I expose my conceptual framework. Then, I provide

[11] Oster (2012a) investigates the lack of behavioural change despite the high prevalence of HIV in Sub-Saharan Africa. She finds reduction in risky sexual behaviour only in areas with higher life expectancy. Consistent with optimizing behaviour, she explains that high rates of non-HIV mortality suppress behavioural response. However, she does not find evidence of greater behavioural change in areas with higher knowledge of the epidemic.

[12] Focusing on rural Malawi, Chimbiri (2007) documents that condom use is negligible inside marriage, and that initiating a discussion about condom use for preventing infection in marriage is like bringing an "intruder" into the domestic space.

[13] See Giuliano (2017) for a review.

a description of the data and I discuss my empirical strategy in section 3. My main results, robustness checks and alternative identification strategies are presented in section 4. I then explore the mechanisms in section 5, and I provide concluding thoughts in section 7.

2 Context and Conceptual Framework

2.1 Ancestral Matrilineality in Sub-Saharan Africa

Kinship relations are important in the context of Sub-Saharan Africa, where kinship groups form a basic political unit in which members recognize each other as kin and often have certain obligations toward each other, such as land sharing, contribution to bride price payments for lineage members, provision of financial support (school fees, burial payments, etc.) (Fox, 1934; Lowes, 2018a). In matrilineal kinship system, individuals trace lineage and descent through women. As such, while biologically related to family of both their mother's side and their father's side, individuals are considered kin only if they share a common female ancestor (Lowes, 2018a).

Figure 1 (a) is from Lowes (2018a) and illustrates the structure of matrilineal kinship systems. As she explains: "In the diagram, men are represented by triangles and women are represented by circles. Membership in the same matrilineal group is denoted with red. Children are in the same matrilineal group as their mothers. Likewise, a mother is in the same matrilineal group as her male and female siblings. In many matrilineal societies, the mother's brother has an important role relative to his sister's children. His inheritance and lineage will be traced through his sister's children, and he has obligation to financially support her children. Importantly, husband and wife do not share the same lineage". Even once married, the wife remains in her lineage of origin: for all married couples one spouse is red and the other spouse is blue. Consequently, as noted by anthropologists, "husbands are less able to mistreat their spouses in matrilineal systems since these latter have greater support from their kin groups." (Lowes, 2018a).

Figure 1 (b) is also from Lowes (2018a) and presents the structure of patrilineal kinship. As she explains: "Now, children are in the same group as their father, as denoted in blue. In a patrilineal society, rather than maintaining strong ties with her own lineage, a woman is effectively incorporated into the lineage of her husband upon marriage. This is because once she is married, she is not relevant for determining descent and inheritance for her lineage. This is illustrated in the patrilineal kinship diagram by the married women denoted in grey, while the unmarried daughter shares the same color as her father." (Lowes, 2018a).

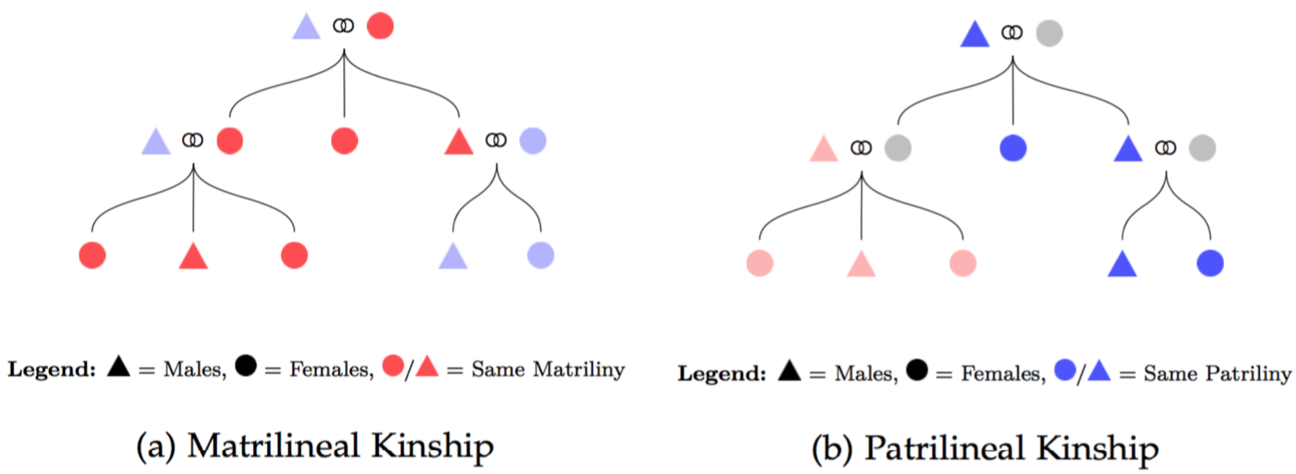
Historically, matrilineal kinship systems are correlated with other cultural traits, which have been shown to have long-term impact on gender roles and economic development.[14] For example, Lowes (2018a) shows that, in Africa[15], matrilineality is negatelively correlated with the practice of bride price, the use of the plough as well as animal husbandry. As reported in Table A.1 in appendix, I find similar evidence looking at my main sample's descriptive statistics. Therefore, as described in subsection 3.3, it will be crucial to control for all these historical correlates.

15 percent of the 527 Sub-Saharan societies recorded in the Ethnographic Atlas are matrilineal; while 70 per-

[14]Interested reader may find an exhaustive overview of the origins of matrilineal kinship systems and related women's empowerment in Lowes (2018a) appendix.

[15]Using data from the *Ethnographic Atlas*.

Figure 1: Diagrams of kinship systems (source: Lowes, 2018a)



cent are patrilineal. Furthermore, as noted in Lowes (2018a), the vast majority of these matrilineal societies are distributed across the center of Africa in the so called *"Matrilineal Belt"*; which intersects present day Angola, Republic of Congo, DRC, Gabon, Malawi, Mozambique, Namibia, Tanzania and Zambia. Figure 2 illustrates the matrilineal belt across Africa, with matrilineal groups denoted in blue, patrilineal groups denoted in green, and bilateral and other groups in beige[16]. In my final sample, about 67% of individuals originating from an ancestrally matrilineal ethnic groups are indeed living in one of the countries crossed by the matrilineal belt. Interestingly, the highest female HIV rates in Sub-Saharan Africa are also found in the countries crossed by the matrilineal belt. In fact, ones talk about an *"AIDS Belt"*. I describe it in more detail in the next subsection.

2.2 Female HIV in Sub-Saharan Africa

In 2018, 14.5 millions of women lived with HIV in Sub-Saharan Africa, representing about 80% of the worldwide female HIV positive population (UNAIDS, 2018b). Uniquely, it is the only place in the world where more women than men live with HIV: adult female HIV positive population is about 1.5 larger than male HIV positive population[17], and adult females represented about 57% of new adult infections in 2017.[18] This phenomenon has been referred to "Feminization of HIV" in Sub-Saharan Africa.[19]

Interestingly, these numbers hide substantial within Sub-Saharan Africa variation in female HIV rates. Indeed, among the 14.5 millions of women living with HIV in Sub-Saharan Africa in 2018, 11.2 millions lived in "Eastern and Southern Africa", against 3.3 millions in "West and Central Africa". The geographical variation in female HIV rates in Sub-Saharan Africa, computed from observations in my final sample using DHS data and linking individuals to their ancestral ethnicity, is represented in Figure 3[20] and shows high prevalence rates in Eastern and Southern

[16]This map is based on Murdock's map of ethnic group's ancestral boundaries. Note that Murdock's classification of ethnic groups slightly differs from the Ethnographic Atlas (EA)'s classification. I therefore use Michalopoulos et al. (2019) and Teso (2019) mappings of Murdock-EA ethnic groups.

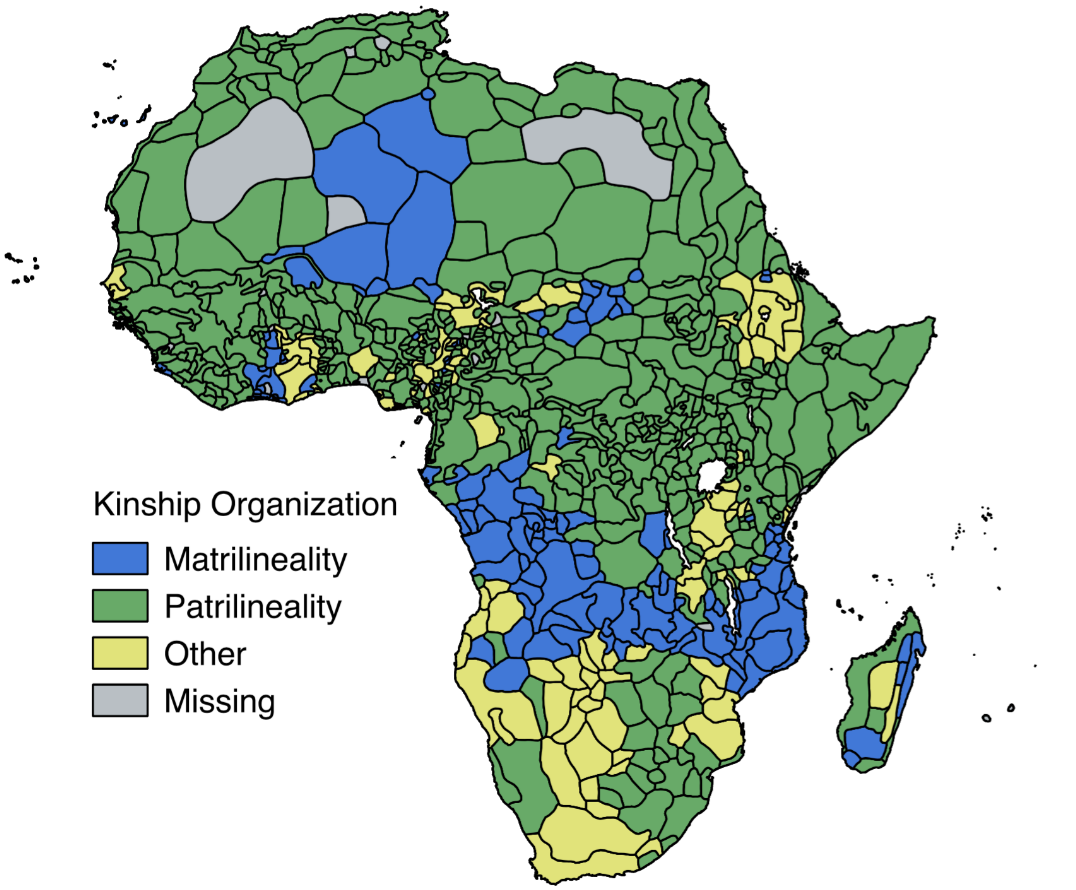
[17]9.5 millions males live with HIV in Sub-Saharan Africa.

[18]Among the 1.02 millions newly infected adult individuals in Sub-Saharan Africa in 2017, 580,000 were females.

[19]These numbers are from `http://aidsinfo.unaids.org/`

[20]This map is based on Murdock's map of ethnic group's ancestral boundaries. Note that Murdock's classification of ethnic groups slightly differs from the Ethnographic Atlas (EA)'s classification. I therefore use Michalopoulos et al. (2019) and Teso (2019) mappings of Murdock-EA ethnic groups.

Figure 2: Ancestral Ethnic Group Boundaries and Matrilineal Belt



Africa, the so-called *"AIDS Belt"*.[21] Zambia and Malawi, two of the countries with highest female HIV rates are also two of the countries with the highest proportions of individuals originating from matrilineal ethnic groups.[22]

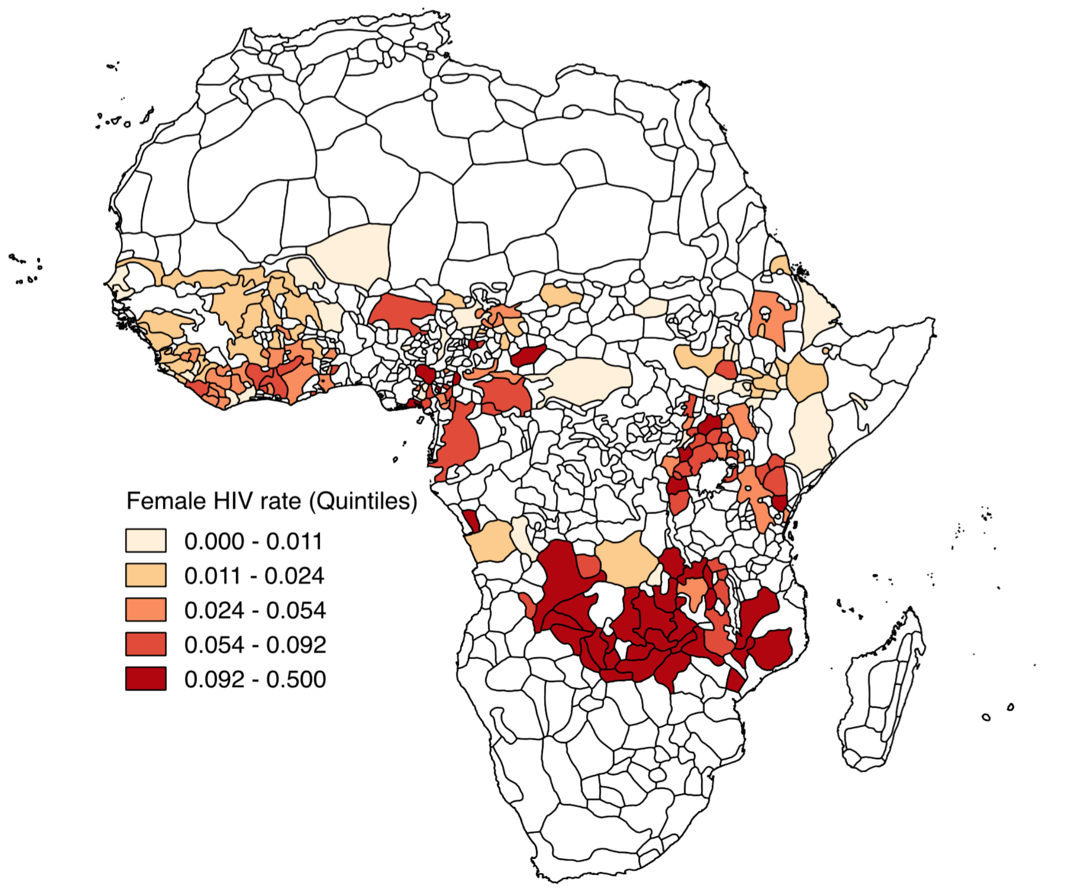
As detailed in section 1, several factors have been emphasized to explain variation in female HIV rates within Sub-Saharan Africa. Institutions have highlighted gender inequality and disempowerment as key barriers to progress against the HIV epidemic (UNAIDS, 2018a), and urged for effort to address these issues. A widespread conjecture is that strengthening women's property and inheritance rights will prevent the spread of HIV/AIDS by promoting women's economic security and empowerment. Anderson (2018) provides the first empirical evidence of a causal relationship between female bargaining power and female HIV infection rates in Sub-Saharan Africa. Exploiting variation in legal origins of Sub-Saharan countries, she finds that HIV prevalence is today higher for women living in common law countries, where code of law is associated with weaker female property rights, as compared to women living in civil law countries. She proposes women's lower intrahousehold barganing power and ability to impose safe sexual pratices to their husband as main explanatory mechanisms.

[21] Interested reader might have a look at `https://www.hsph.harvard.edu/news/magazine/spr08circumcisionmap/`; and/or `https://www.prb.org/thestatusofthehivaidsepidemicinsubsaharanafrica/`

[22] In my final sample of matrilineal versus patrilineal females, proportions of matrilineal females are about 87% in Malawi and about 60% in Zambia; female HIV rates are about 12% in Malawi and about 14% in Zambia. These numbers are computed using the following DHS surveys: Malawi (2004, 2010, 2014) and Zambia (2007, 2013). See subsection 3.1 for more details. This compares to the proportion of about 15% of matrilineal females, and the average female HIV rate of about 5% in my full final sample of matrilineal versus patrilineal females.

However, this mechanism does not explain the geographical correlation between the so-called *"Matrilineal Belt"* and the so-called *"AIDS Belt"* in Africa: the highest rates of female HIV can be found in ancestrally matrilineal geographical areas, correlated with higher women's status and sexual autonomy. I explore this puzzle in this paper. To do so, I begin with a description of my conceptual framework in the next subsection.

Figure 3: Ancestral Ethnic Group Boundaries and Contemporary Female HIV Rates (Final sample)



2.3 Conceptual Framework

2.3.1 Gender Differences in Sexual Strategies

Humans, like other sexually reproducing species, do not choose mate randomly. According to the latest theories from evolved psychologists, our mating is strategic, and the sexual strategies we developed were shaped in the (very) long run by natural selection through our ability to survive and reproduce successfully. As Buss (2016) states: *"Those in our evolutionary past who failed to mate successfully failed to become our ancestors. All of us descend from a long and unbroken line of ancestors who competed successfully for desirable mates, attracted mates who were reproductively valuable, retained mates long enough to reproduce, fended off interested rivals, and solved the problems that could have impeded reproductive success. We carry in us the sexual legacy of those success stories."*

Further, underlying each sexual strategy[23] are psychological adaptations (e.g. preferences for a mate, feelings of love, desire for sex, sexual jealousy, etc.), that are sensitive to the information or cues from the external world.

These theories provide an interesting framework to understand why males and females exhibit on average marked different sexual preferences. According to numerous psychological studies[24], males exhibit a stronger preference for casual relationships and sexual variety; while females tend to be more choosy in their mate as well as more looking for long-term committed relationships. Indeed, producing million of sperms, wich are replenished at a rate of roughly 12 millions per hour (Buss, 2016), males can expect at most to reproduce as many times as there are available fertile females willing to have sex. Additionally, males only bear minimal initial parental investment in case of fecondation. Consequently, preference for sexual variety and casual sexual relationships is a psychological trait that has been relatively more developed in male evolved sexual psychology. On the contrary, producing only a fixed limited amount of reproductive cells (i.e. only 400 ova mature to the point where they are capable of being fertilized), females face a limited number of reproductive opportunities. In addition, bearing the larger share of the initial parental investment (i.e. gestating, bearing, nursing, nurturing and protecting a child), females provide extraordinarily valuable, but limited, reproductive resources. Therefore, because females in our evolutionary past risked enormous investment from having sex, evolution favored females who were highly selective about their mates. In particular, evolution favored females capable of reproducing with long-term committed and reliable provider mates, ensuring the survival and reproduction success of their offsprings (Diamond, 1998).

Nevertheless, while the reproductive benefits of casual relationships for males are large and direct, women may also reap benefits from short-term mating, according to the latests evolutionary psychology theories. As a matter of fact, even though having access to more sperm would not increase a woman's reproductive success, through casual sex it is possible for a woman to gain superior genes that are passed on to their children. This is the so-called "Better Genes Theory"[25] (Buss, 2016; Greiling and Buss, 2000). As such, a women might try to secure the investment of a lower-ranking man by marrying him, for example, while simultaneously securing the genes of a higher-ranking man by mating with him casually (the maing marketplace rending far easier for a woman to get a men with better genes to have sex with her than to get him to marry her) (Buss, 2016). One version of this "Better Genes Theory" has been called the "Sexy Son Hypothesis", according to which women prefer to have casual sex with men who are attractive to other women because they will have sons who possess the same charming characteristics and therefore will enjoy greater mating success in the next generation.

However, adopting such extra-pair mating strategies may come at a cost. In particular, an unfaithful married woman risks the withdrawal of resources by her husband, reputational damages as well as getting pregnant and bearing a child without the benefits of an investing partner. To sum up, short-term mating poses hazards, but has powerful benefits as well, and women have evolved psychological mechanisms to select circumstances in which the cost of short-term mating are minimized and the benefits increased.

Matrilineal kinship organizations, relative to patrilineal ones, constitute environments in which extra-pair mating sexual relationships should be beneficial for women. Indeed, as detailed in subsection 2.1, in matrilineal kinship

[23] *"Sexual strategies do not require conscious planning or awareness. [...] Indeed, just as pianist's sudden awarenes of her hands may impede performance, most human sexual strategies are most successfully carried out without the awareness of the actor."* (Buss, 2016)

[24] See Buss (2016) for an extensive review.

[25] I refer reader interested in alternative hypothesis on women's extra-pair mating function to Buss (2016) and Greiling and Buss (2000), which provide extensive discussions.

organizations males' biological children do not integrate their lineage but the lineage of their mother. In addition, matrilineal males do not transmit their wealth to their biological children but to their sisters' children. In other words, females in matrilineal societies should have lower expectations regarding their mate's propensity to engage in long-term committed relationships to maximize the reproductive success of their offsprings. Consequently, for their reproductive success they should expect relatively greater benefits from gaining access to better genes, through extra-pairing mating. Furthermore, contrary to patrilineal societies where females integrate their husband's lineage once married, matrilineal females remain in their lineage of origin after marriage and keep strong bounds with their family. In other words, matrilineal females benefit from higher marriage outside option, relative to their patrilineal counterparts, but also from relatively lower opportunity cost associated with extra-pair mating sexual strategy. Finally, being inherently valued more through their position in their societies, matrilineal females should also benefit from a greater sexual autonomy than their patrilineal counterparts, and a greater ability to pursue their preferred sexual strategy.

If, according to my conceptual framework, ancestral matrilineality fostered females to seek for better genes through sexual variety, matrilineality should be correlated with greater genetic diversity as well as father diversification. This is exactly what I find in Table A.2 in appendix. As reported in the two first columns, I find a positive country-level association between the estimated proportion of countries' citizens with ancestors that had matrilineal inheritence rule (using data from Giuliano and Nunn, 2018) and countries' indexes of genetic diversity (computed by Ashraf and Galor, 2013b), robust when including continent fixed effects as well as a large array of controls, including distance to Addis Ababa, the cradle of humankind. Further, as reported in the two last columns, running individual-level regressions following my main identification strategy on my final sample from DHS (described in more detail in subsection 3.1), I find that women originating from ancestrally matrilineal ethnic groups are significantly more likely to diversify the fathers of their biological children.

Alternatively, the adoption of matrilineal kinship structures may have been an adaptative response to environments with low paternal certainty. Indeed, evolutionary anthropologists explain the existence of matrilineal societies as the result of an evolutionary process that created institutions suitable for the ecological and social environment (Lowes, 2018a). One hypothesis is that matrilineality would have emerged in environments with low paternal certainty because, while it is difficult to confirm paternity, maternity is easily observable. Thus, an inheritance system in which property passes from the mother's brother to her sons may be optimal since the brother knows he is related to his sister, but cannot verify that he is related to his children (Fortunato, 2012). However, as noted by Lowes (2018a), this model alone would require that paternity certainty be below .268, a value that is unrealistically low even for matrilineal societies. Still, developping a more complex model Holden et al. (2003) argue that daughter-biased investment (leading to matrilineality) may be adaptive when the marginal benefit of investing in sons (relative to daughters) is not sufficient to offset by the risk of non-paternity of the son's children.[26] In sum,

[26]The evolutionary anthropology literature provides alternative hypothesis on the ecological origins of matrilineality. Aberle (1961) posits that matrilineality would be more beneficial with certain types of production such as hoe agriculture, while patrilineality would be more beneficial with hunting, which requires skill development and male cooperation. Holden and Mace (2003) argue that with the rise of moveable heritable wealth, such as cows, the marginal benefits of investing in sons increases, leading to the demise of matrilineal societies. Pastoralism, and the spread of cattle in particular, would have therefore led to the loss of matrilineality against the adoption of patrilineality and mixed descent rules. BenYishay et al. (2017) find that coral reef density predicts the prevalence of matrilineality in the Solomon Islands. Further, they find that reef density explains as much as 10% of the variation in inheritance rules across villages in the Solomon Islands, suggesting an adaptation of institutions to ecological conditions. Interested reader may find a more exhaustive overview of the literature on the origins of matrilineal kinship systems in Lowes (2018a) appendix.

ancestral matrilineality would be associated with ancestrally higher paternal uncertainty and more promiscuous sexual behaviour and, ultimately, with more promiscuous contemporaneous sexual behaviours.

All in all, my conceptual framework points toward a main prediction, that I empirically test in this paper: I expect to find that, as of today, matrilineal females adopt sexual behaviours that are more promiscuous than their patrilineal counterparts, and therefore suffer from more HIV.

2.3.2 Women's Empowerment and Contraceptive Use

As discussed above, I expect matrilineal women to be more empowered than their patrilineal counterparts, which may also affect their contraceptive behaviour. Indeed, when it comes to the choice of contraceptive use within a couple, intra-household bargaining matters. As such, and because men's preferences may constrain household adoption (Ashraf et al., 2020a), women's empowerment acts a driver of contraceptive use (Anderson, 2018, Cassidy et al., 2018, Kibira et al., 2014). However, what mainly matters to restrain the spread of HIV is not the use of contraceptives *per se*, but the *type* of contraceptive technology used. This latter is also affected by women's empowerment. Indeed, bearing the responsability of contraception and reproduction, more empowered women may be more likely to substitute short-term contraceptive methods such as male condom with long-term contraceptive methods (e.g. intra-uterine device (IUD), female sterilization, implants, pills or injections) giving them long-term control over their fertility (Islam et al., 2009, Palamuleni and Adebowale, 2014, Samari, 2018).

As detailed in subsection 2.1 and reported in Lowes (2018a), due to their central position in their kinship structure and the support they receive from their relatives, matrilineal women have a higher social status than their patrilineal counterparts. Therefore, I expect them to to benefit from a higher intra-household bargaining power, and to be consequently more likely to substitute male condom with long-term contraceptive methods. By doing so, they would substitute more protective contraceptive methods with less protective ones, which would contribute to explain why they suffer from more HIV.

3 Data and Empirical Strategy

To study the long-term impact of ancestral matrilineality on female HIV, I match contemporary individual-level data from the *Demographic and Health Surveys (DHS)* with historical ethnic group-level data from the *Ethnographic Atlas (EA)*. This section describes the data and the empirical strategy.

3.1 Contemporary Data

Data on HIV infection come from the DHS, which have been conducted in sub-Saharan African countries since the 1990's. The DHS household surveys typically interview a nationally representative sample of between 10,000 to 20,000 women (aged 15-49) and men (aged 15-59). By collecting blood for HIV testing from representative samples of the population, the DHS Program provides nationally representative estimates of HIV prevalence rates. The testing is simple: the interviewer collects dried blood spots (DBS) on filter paper from a finger prick and the filter paper is transported to a laboratory for testing. The testing is anonymous, voluntary, and non-informative to respondents. The average response rate is extremely high; 93 percent for women (slightly lower for men).

I restrict my sample to DHS surveys containing both HIV testing and individual ethnicity information, as well as GPS data. I further restrict my sample to individuals originating from ethnic groups which are either ancestrally matrilineal, or patrilineal. This leaves me with a sample of 159,560 women across 18 Sub-Saharan African countries (i.e. 32 country-surveys[27]). The proportion of men tested for HIV is lower and I have a sample of 120,580 men. As a main outcome, I construct an individual-level indicator variable, *HIV*, that takes value one if the respondent is seropositive according to the DHS HIV Test.

On average, 4.8 percent of women in my initial sample are HIV positive (this compares to 3.1 percent of men). The average HIV infection rate of women originating from ancestrally matrilineal ethnic groups is approximately 11.4 percent. It is close to one-fourth, at 3.7 percent, for women originating from ancestrally patrilineal ethnic groups. This compares to 7.8 percent for matrilineal males vs. 2.1 percent for patrilineal males. Though only correlational, these numbers provide a first evidence of the higher prevalence of HIV among individuals originating from ancestrally matrilineal group, women in particular.

Further, DHS surveys present questions that are useful measures of female sexual autononomy, actual sexual behaviour, contraception use, acknowledgment about HIV risks and protective methods, and sexual debuts, which I investigate in section 6.

Finally, I also exploit information on geographic covariates, computed at the village level from numerous data sources, and provided by the DHS. Additionally, I exploit data on the location of large-scale active mines in Africa, provided by S&P Global Market Intelligence[28] and used in Berman et al. (2019), and I compute measures of presence of active mines near DHS villages. Village-level geographic controls are described in subsection 3.3.

3.2 Historical Data

Data on pre-colonial ethnic groups' traits come from the *Ethnographic Atlas (EA, 1967)*, a worldwide ethnicity-level database constructed by George Peter Murdock. This database covers 1,265 ethnic groups in the world[29], and contains detailed ethnographic information on cultural aspects and ways of life of the portrayed ethnic groups, *prior to industrialization and colonial contact*[30], such as kinship and family organization, settlement patterns, political organization, institutional complexity, historical mode of subsistence, etc.

To match individual-level contemporary data with historical data, I use information provided in DHS on individual's ethnicity. However, the classification of respondents' ethnic groups in DHS does not always coincide with the *Ethnographic Atlas*' classification, requiring a matching between the two datasets.[31] Therefore, I first follow Michalopoulos et al. (2019) matching, which enables me to do the matching for most of the individuals in my sample. I then follow Teso (2019) matching[32] to match some remaining individuals. Finally, I build on online sources[33]

[27] Burkina-Faso (2003, 2010); Cameroon (2004, 2011); Chad (2014); Congo Democratic Republic (2007, 2013); Ethiopia (2005, 2011, 2016); Gabon (2012); Ghana (2003, 2014); Guinea (2005, 2012); Ivory Coast (2011); Kenya (2003, 2008); Liberia (2013); Malawi (2004, 2010, 2014); Mali (2006, 2012); Senegal (2005, 2010); Sierra Leone (2008, 2013); Togo (2013); Uganda (2011); Zambia (2007, 2013).

[28] Interested reader may find their website here: `https://www.spglobal.com/marketintelligence/en/campaigns/metals-mining`

[29] The majority of the ethnicities sampled are in Africa.

[30] "For the parts of the world without a written history, the information is from the earliest observers of these cultures. For some cultures, the first recorded information is from the 20th century, but even for these cultures, the data capture as much as possible the characteristics of the ethnic group prior to European contact." (Alesina et al., 2013)

[31] For many of the groups, the matching is straightforward as the name used in the DHS is the same or very similar to the one used in the *Ethnographic Atlas (EA)*. When the name of an ethnic group in DHS is not found in EA's classification, this is typically because an alternative group's name is used.

[32] Teso (2019) matching procedure largely builds on Michalopoulos et al. (2019) matching procedure.

[33] One of the most useful sources of information on alternative ethnic groups' names is the Joshua Project website (`http://www.`

to match ethnic groups not matched in previous procedures. I end up with a sample of 349,895 individuals in DHS matched with an ancestral ethnic group in the *Ethnographic Atlas (EA)*.

I then discard individuals originating from ancestral ethnic groups with inheritance rule and kinship organization which were neither matrilineal nor patrilineal (such as bilaterality of ambilinearity), or with missing information in the EA, representing 69,755 out of 349,895 individuals in my sample. I therefore restrict my sample to individuals originating from an either ancestrally matrilineal or patrilineal ethnic group, and I end up with a sample of 280,140 individuals (159,560 women and 120,580 men).

Finally, I construct my main explanatory variable, *Matrilineality*, as an indicator of ethnic-group's ancestral matrilineality, from EA information on ethnic group's inheritance rule. In my analysis, I additionally use a wide array of historical controls varying at the ethnic group level and computed from the EA. I describe these controls in the next subsection.

3.3 OLS Empirical Strategy

I explore the gender-specific long term effect of ancestral matrilineality on HIV by estimating the following equation:

$$\begin{aligned} y_{ievrt} = \alpha + \beta_1 Matrilineality_e + \beta_2 Female_i + \beta_3 Female_i \times Matrilineality_e + \\ \mathbf{X'_{ievrt}}\Delta + \mathbf{X'_{ert}}\Omega + \mathbf{X'_{vrt}}\Pi + \lambda_{rt} + \varepsilon_{ievrt} \end{aligned} \tag{1}$$

with y_{ievrt} denoting an individual-level outcome (for example an indicator for whether the individual is HIV positive) for individual i from ethnic group e, living in village v in within-country DHS region r, and surveyed at year t. $Matrilineality_e$ is an indicator for whether an individual originates from an ancestrally matrilineal ethnic group (an ancestrally patrilineal ethnic group otherwise); $Female_i$ is an indicator for whether an individual is a female. β_1 is intended to capture a "Matrilineal effect"; and β_2 is intended to capture a "Gender effect" on HIV.[34] β_3 is my main coefficient of interest and captures the effect of ancestral matrilineality on female HIV once the "Matrilineal effect" and the "Gender effect" have been controlled for. The inclusion of this interaction term is motivated by my conceptual framework, according to which originating from an ethnic group with an ancestral matrilineal kinship organization should significantly influence contemporary sexual behaviour for female individuals specifically. This hypothesis is further motivated by the first descriptive statistics reported in subsection 3.1 on heterogeneity of contemporary HIV rates by gender and individual's ethnic group's ancestral kinship organization.

$\mathbf{X'_{ievrt}}$ represents a set of individual-level covariates which includes indicators of marital status; a dummy for whether an individual is in a polygynous union; number of children; age; age squared; a dummy for whether an individual lives in an urban location; education in number of years; a dummy for whether an individual is currently working; wealth index indicators and religion indicators.

$\mathbf{X'_{ert}}$ represents a set of ethnic group-level ancestral covariates which includes intensity of women's historical

`joshuaproject.net/`). For most of the unmatched ethnicities, the respondent lists her nationality as ethnicity.

[34] According to the medical literature, females are more likely than males to be infected when exposed to HIV, due to physiological and anatomical reasons, such as larger surface area of mucosal HIV exposure (Yi et al., 2013).

participation in agriculture; ancestral polygyny; ancestral bride price; ancestral plough use; ancestral pastoralism; ancestral presence of clans; indicators of ancestral settlement patterns; indicators of ancestral juridictional hierarchies beyond local communities; ancestral reliance on hunting; ancestral reliance on fishing; ancestral reliance on gathering; ancestral reliance on animal husbandry; ancestral reliance on agriculture; ancestral presence of large domesticated animals; indicators of intensity of ancestral agriculture; and year of observation of the ethnic group in the *Ethnographic Atlas*.

$\mathbf{X'_{vrt}}$ represents a set of village-level geographic covariates which includes latitude; longitude; altitude; nightlight composite; population density (2010); distance to lake or coastline; distance to nearest international border; average time (minutes) required to reach a high-density urban center (2015); malaria incidence (2010); vegetation index; indicators for the length of the growing season; distance to the nearest active mine; a dummy indicating the presence of an active mine within 1000 km; an index of ethnic fractionalization; and an index of ethnic polarization. Summary statistics of all my control variables are reported in Table A.1 in appendix.

Crucially, I also include (within-country) DHS region-survey (year) fixed effects, λ_{rt}, to take into account time trends, as well as unobserved country-level and within-country level institutional, economic and geographic factors that could potentially affect contemporary HIV prevalence and also be correlated with the geographical distribution of ancestral matrilineality. Doing so allows me to assess the influence of matrilineal versus patrilineal ancestral kinship organizations on individuals located in the same institutional, economic, and geographic environment. This is the epidemiological approach of the cultural economics literature (Fernández, 2011).

Finally, since variation in the main explanatory variable occurs at the ethnic group level, observations of outcomes of individuals of the same ethnic group may not be independent. Consequently, in order to account for potential within-group correlation of the residuals (ε_{ievrt}), throughout the analysis standard errors are clustered at the ethnic group level.

A crucial concern for the causal interpretation of the OLS estimates is the possible presence of omitted variables that are correlated with both contemporary HIV prevalence and with ancestral matrilineality of ethnic groups. For instance, if ancestrally matrilineal ethnic groups were more likely to have social organizations and institutions[35], as well as modes of production more conducive of equal gender norm[36] or of the spread of the virus, this would translate in an estimate of β_3 that is biased upward. The ethnicity-level controls are meant to alleviate these concerns. Additionally, I include ethnic group's year of observation in *Ethnographic Atlas* to alleviate the concern that some groups were portrayed later than others and might therefore have been more developed, and hence potentially more gender equal, for example.

Likewise, β_3 might be biased upward if ancestral matrilineality was correlated with geographic factors that are conducive of HIV. For example, geographic characteristics such as the type of vegetation or the altitude may be correlated with the spread of the virus. I directly control for that, and I also control for malaria incidence, a proxy measure of virus spread. Further, following Corno and De Walque (2012), mine workers' international migration is an other driver of the spread of HIV. I therefore control for the village's distance to the nearest active mine, as well as for the presence of an active mine within 1000km. In the same vein, Oster (2012b) highlights exports and road networks, and subsequent increase in movements of people and sexual contacts, as an other factor of HIV infection.

[35] In this vein, Ashraf et al. (2020b) document the positive role of ethnic groups' practice of bride price on female education.

[36] In this vein, Alesina et al. (2013) document the negative role of ancestral plough use on contemporary female empowerment.

Controlling notably for distance to international borders as well as average time to reach a high-density urban center as a proxy allows me to alleviate these concerns. An other long-term determinant of female HIV in Sub-Saharan Africa put forward by Cagé and Rueda (Forthcoming) is the geography of Protestant and Catholic missions in the early 20th century, as well as their health investments. In the same vein, Teso (2019) shows the long term effect of the slave trade in Sub-Saharan Africa on contemporary gender norms. Again, my OLS estimates would be biased if ancestral matrilineality was correlated with such factors. The inclusion of numerous geographic covariates, such as latitude, longitude as well as distance to lake or coastline and average time to reach a high-density urban center, is meant to alleviate these concerns.

Along the same lines, β_3 might also be biased upward if ancestral matrilineality was correlated with ethnic fractionalization and/or ethnic polarization. Indeed, Tequame (2012) provides evidence that these latter are associated with higher information asymmetry and lower social sanctions within communities, which are more conducive to risky sexual behaviour and therefore HIV. Computing indexes of ethnic fractionalization and polarization at village level and including them in my regressions is aimed at controlling for such potential omitted bias.

Ethnic groups' economic prosperity may be an other possible omitted variable, potentially correlated with both ancestral matrilineality and contemporary HIV prevalance. To account for this, I control for village's population density (2010) and nightlight composite, two proxies of economic prosperity. The inclusion of historical ethnic groups' modes of production and institutional controls are also meant to capture historical economic prosperity. Again, including ethnic group's year of observation as a control allows me to alleviate the concern that some groups were portrayed later than others and might therefore have been more developed. Related to this concern, ethnic groups' access to contemporary health infrastructures could be an other possible omitted variable: controlling notably for average time (minutes) required to reach a high-density urban center (2015) is meant to control for this.

Finally, institutional factors may also drive female HIV. Anderson (2018) underlines that legal origins of Sub-Saharan African countries significantly determine current female HIV rates: she namely finds that female HIV rates are significantly higher in common law countries, relative to civil law countries. The inclusion of (within-country) DHS region x survey (year) fixed effects is notably meant to capture such country (and within-country) institutional factors, in addition of time trends.

4 The Long-Term Effect of Ancestral Matrilineality on Female HIV

4.1 Main Results - OLS Estimates

Table 1 presents the OLS estimates of the effect of ancestral matrilineality on contemporary female HIV, once gender and ancestral matrilineality have been controlled for. While in column 1 I do not include any control, I include (within-country) region-survey (year) fixed effects in column 2. The coefficient of the interacted variable $Female \times Matrilineality$ is positive and statistically significant (row 3), and is not affected by the inclusion of region-survey fixed effects. Further, following the inclusion, of individual controls to account for differences in socio-demographic composition of matrilineal and patrilineal individuals, the main estimate of interest remains unchanged in column 3. To alleviate omitted variable concerns detailed in subsection 3.3, I subsequently include ethnic group's controls in my regressions (column 4), and village level geographic controls (column 5). The coefficient of the interacted variable $Female \times Matrilineality$ remains very consistent and of large magnitude across the specifications. Women originating from an ancestrally matrilineal ethnic group are 1.6 to 2.2 percentage points more likely to be HIV positive than their patrilineal counterparts. This effect is large in magnitude, corresponding to 70% to 105% of the average HIV prevalence (2.1% to 2.3%) among patrilineal males (control group) in my full regression sample. This result supports the main prediction of my conceptual framework.

Interestingly, I also find a consistent and statistically significant positive estimate of being a female on the likelihood of being HIV positive (row 2). This is an additional evidence of the well-documented "Feminization of HIV in Sub-Saharan Africa" discussed in section 1. As highlighted in Greenwood et al. (2019) and in the medical literature, for physiological and anatomical reasons, women are at higher risk than men to contract the virus when exposed to it (Yi et al., 2013).

Importantly, the "Matrilineal effect" (row 1) becomes non-significant and very close to zero, once region-survey fixed effects are included. Ancestral matrilineality in Sub-Saharan Africa being essentially geographically located in countries of the so-called "Matrilineal Belt", this suggests that the unconditional positive significant effect of $Matrilineality$ found in column 1 may in fact capture the effect of other country and within-country level factors. Among them, legal systems and codes of law could be plausible candidates. Indeed, Anderson (2018) highlights that female HIV rates are signicantly higher in common law countries. Since most of countries of the "Matrilineal Belt" are common law countries, the "Matrilineal effect" found in column 1 might in fact be a "Common Law Countries effect" (I discuss it in more details in the following subsection). However, this effect disappears once (within-country) region-survey fixed effects are included, while my main coefficient of interest remains consistent, positive and statistically significant at the 1% level.

All in all, the absence of significance of the "Matrilineal effect" as well as the positive significant effect of my main variable of interest ($Female \times Matrilineality$) is confirmed by the analysis made by gender subsamples, and presented in columns 6 and 7 of Table 1: I find that women originating from an ancestral matrilineal ethnic group are 1.2 percentage points (column 6) significantly more likely to be HIV positive (representing 31% of the average HIV prevalence of patrilineal females in my regression sample); while I do not find any significant effect for their male counterparts (column 7).

This gender-specific effect can be rationalized by the interaction of two phenomena. First, according to my conceptual framework, ancestral matrilineality has a long-lasting effect specific on female's contemporary sexual behaviour. Second, the medical literature highlights that, for physiological and anatomical reasons (e.g. larger surface area of mucosal HIV exposure; increased mucosal expression of the HIV co-receptor CCR5; and a greater probability of virus exposure on the rectal mucosa among other factors), women are at higher risk than men to contract the virus when exposed to it (Yi et al., 2013).[37] In the end, the fact that matrilineal women adopt riskier sexual behaviours than their patrilineal counterparts combined with the higher suceptibility of women to contract the virus when exposed to it lead to the gender-specific long-lasting effect of matrilineality on HIV.[38]

[37] Building on several medical studies, Greenwood et al. (2019) assume that women are 75% more likely to get infected than males for physiological and anatomical reasons.

[38] I illustrate this in section 6, performing a simulation based on an epidemiological model assuming behavioural differences between matrilineal and patrilineal women as well as gender differences in HIV susceptibility.

Table 1: **The Effect of Ancestral Matrilineality on Female HIV (OLS)**

	HIV						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Matrilineality	0.058***	0.000	0.001	-0.000	-0.001	0.012*	-0.004
	(0.011)	(0.004)	(0.005)	(0.006)	(0.005)	(0.007)	(0.005)
Female	0.016***	0.012***	0.009***	0.009***	0.008***		
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)		
Female × Matrilineality	0.020***	0.022***	0.019***	0.016***	0.017***		
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)		
Ind. Controls			Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls				Yes	Yes	Yes	Yes
Village-Geographic Controls					Yes	Yes	Yes
Region-survey FE		Yes	Yes	Yes	Yes	Yes	Yes
Gender	Both	Both	Both	Both	Both	Female	Male
Observations	280,140	280,140	273,417	193,991	182,312	105,964	76,348
Adj. R-squared	0.018	0.050	0.078	0.076	0.078	0.087	0.065
Clusters	172	172	172	100	100	100	82
Mean Dep. Var. (Patri. Males)	0.021	0.021	0.021	0.023	0.023	0.039	0.023

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group (it therefore excludes individuals originating from ethnic groups with alternate inheritance rules (ambilineality, bilinearity, duolinearity, etc.)). "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. **"HIV"** is a dummy indicating whether an individual is HIV positive (from DHS HIV Tests). *The Individual Controls* are computed from DHS and include: indicators of marital status; a dummy for polygynous union; number of children; age; age squared; a dummy for living in an urban location; education in number of years; a dummy for currently working; wealth index indicators and religion indicators. *The (Ancestral) Ethnic Group Controls* are computed from the Ethnographic Atlas (EA) and include: intensity of women's historical participation in agriculture; ancestral polygyny; ancestral bride price; ancestral plough use; ancestral pastoralism; ancestral presence of clans; indicators of ancestral settlement patterns; indicators of ancestral juridictional hierarchies beyond local communities; ancestral reliance on hunting; ancestral reliance on fishing; ancestral reliance on gathering; ancestral reliance on animal husbandry; ancestral reliance on agriculture; ancestral presence of large domesticated animals; indicators of intensity of ancestral agriculture; and year of observation of the ethnic group. *The Village-Geographic Controls* are computed at the village level and include: latitude; longitude; altitude; nightlight composite; population density (2010); distance to lake or coastline; distance to nearest international border; average time (minutes) required to reach a high-density urban center (2015); distance (in km) to nearest active mine; a dummy indicating whether an active mine is located within a distance of 1000 km max. of the village; malaria incidence (2010); vegetation index; indicators for the length of the growing season; an index of ethnic fractionalization; and an index of ethnic polarization. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. Sample in column 6 consists of female individuals. Sample in column 7 consists of male individuals. ** p<0.10,** p<0.05,*** p<0.01*

4.2 Robustness Checks

4.2.1 Selection Analysis

The blood test being not compulsory, selection might arise in the sample. However, the DHS program reports that the average response rate, for those who are eligible for the test, is extremely high and that a comparison between the characteristics of those who agreed to be tested and those who refused testing shows minimal bias.[39] Moreover, it is reasonable to expect that any selection would cause a downward bias since infected individuals should be less keen to be tested.[40] As a result, if female HIV infection is positively affected by ancestral matrilineality, more underreporting among females originating from ancestrally matrilineal ethnic groups should be expected. In order

[39]See https://dhsprogram.com/topics/HIV-Corner/hiv-prev/index.cfm

[40]This is confirmed by Mishra et al. (2006), who find that the rate of HIV infection among individuals not tested for HIV is systematically larger than the rate among those not tested.

to test this hypothesis, I test the potential effect of ancestral matrilineality on the probability of underreporting HIV infection, which in practice amounts to a refusal to consent to the blood test and non take-up of it.

Further, given the persistence of mistrust in medicine in regions where colonial medical campaigns were established (Lowes and Montero, 2018), lack of consent in regions close to missions could pose a potential threat to the estimation. The inclusion of (within-country)-survey fixed effects as well as numerous village-level geographical controls allows me to rule out such potential threat.

Column 1 and column 2 of Table A.3 report estimates on the probability that DHS respondant consent to take HIV test, and on the probability that she actually takes it. The estimate of $Female \times Matrilineal$ is non significant and very close to zero in both cases, suggesting that potential selection can be ruled out (in fact, descriptive statistics show that about 95% of patrilineal females against 93% of matrilineal females consent to HIV test in my final sample, and these numbers are similar for actual take-up.)

While I do not find evidence of selection into DHS HIV testing between matrilineal and patrilineal women, selective attrition due to HIV mortality might affect my estimates. The direction of this bias is unclear. On one hand, if we assume that seropositive individuals have a lower life expectancy, the higher HIV prevalence among matrilineal women should bias downward my main results.[41] On the other hand, using country-level data from UNAIDS, I find a positive association between matrilineality and antiretroviral treatment (ART) coverage (see my discussion on ART in subsection 5.4), and a negative association between matrilineality and HIV mortality.[42] While time-varying country-level variations are explicitly controlled for in my regressions through the inclusion of (within-country) region-survey (year) fixed effects, these country-level correlations suggest that bias due to selective mortality could also go in the other way.

4.2.2 Other Health Outcomes as a Falsification Test

HIV is a highly infectious, largely sexually transmitted disease. Using DHS data on level of anemia in respondents, a not sexually transmitted disease[43], as well as BMI and Rohrer index as other objective measures of health status[44], I perform falsification tests by estimating the long-term effect of ancestral matrilineality on such health outcomes.[45] This allows me to rule out the possibility that differences in HIV rates found previously hide more general differences in economic development and/or health infrastructures, and therefore differences in overall health status. Results are reported in column 3 to 5 in Table A.3. I find that ancestral matrilineality is not associated with an increase in the prevalence of anemia, neither with Body Mass Index (BMI)[46] or Rohrer index[47] of female individuals, corroborating the hypothesis that sexual behaviour is the actual driver of my results on HIV.

[41] The average time from infection to the outbreak of symptoms is equal to 10 years; and the average time from the outbreak of symptoms to death is 2 years (Greenwood et al., 2019).

[42] By lowering the viral load, ART makes the seropositive person taking the drugs feel healthier and live longer, notably.

[43] Bertocchi and Dimico (2019) select anemia for their falsification test because of its relevance, given the association between anemia and malaria, another vast-scale health problem in Africa.

[44] This measures are only available for female respondants in DHS.

[45] I measure the severity of anemia with a dummy variable taking value one when an individual is diagnosed with either mild, moderate, or severe anemia, and zero otherwise.

[46] Body Mass Index (BMI) is computed as follows: $mass/height^2$

[47] Rohrer index is computed as follows: $mass/height^3$

4.2.3 Robustness to Alternative Channels

The previous results demonstrate that ancestral matrineality is a long-term determinant of contemporary prevalence of HIV among female populations in Sub-Saharan Africa. However, other long-term determinants of female HIV in Sub-Saharan Africa have also been recently highlighted in the literature. I discuss here the robustness of my main findings to these competing channels.

Common Law vs. Civil Law Countries. Anderson (2018) has recently highlighted the legal origins of female HIV in Sub-Saharan Africa. In particular, exploiting variation in legal origins of formerly colonized countries and the fact that common law is associated with weaker female marital property laws as compared to civil law, she finds higher prevalence of female HIV in common law countries. Lower bargaining power of women in these countries and therefore lower ability to impose safe sexual pratices to their husbands constitute her main mechanism. Interestingly, in her identification strategy she exploits geographical variation in common law vs. civil law countries *within ethnic group*, by including ethnic group fixed effects, and therefore rule out any ethnical effect. In some sense, my identification strategy in symmetric since I exploit ethnic norm (and related kinship organization) variation within *(within-country) region - survey (year)*, therefore allowing me to rule out any legal/institutional effect.

As a further robustness check, I also perform an heterogeneous analysis, estimating Equation 1 on the subsample of individuals residing in common law countries versus individuals residing in civil law countries.[48] The two first columns of Table A.4 reports the result of these estimations and provide evidence of an heterogeneity: the effect of ancestral matrilineality on female HIV holds true within common law countries only. In fact, looking at the magnitude, it seems that the average effect found in subsection 4.1 is mostly driven by individuals residing in common law countries. However, this is not a surprising result since most of the variation in ancestral inheritance norm is found within common law countries.[49]

Polygyny. Bertocchi and Dimico (2019) have recently highlighted contemporary polygyny as an other driver of female HIV in Sub-Saharan Africa. Females' riskier sexual behaviour triggered by the absence of the husband, and subsequent multiplicative virus transmission by the husband to his other wives constitute their main mechanism. The inclusion of a dummy for actual polygynous union in my individual-level controls, as well as a dummy for ancestral polygyny norm in my ethnic-group controls are meant to capture this alternative transmission channel.

Again, I also peform an heterogeneous analysis as an additional robustness check, by estimating Equation 1 on the subsamples of individuals who are not currently in a polygynous union versus individuals who are currently in a polygynous union. The results are reported in column 3 and 4 of Table A.4 and indicate an heterogeneity in the long-term effect of ancestral matrilineality on female HIV in Sub-Saharan Africa: these effect holds true for non-poygynous individuals only. Similarly to the previous heterogenous analysis, the magnitude of the effect found for non-polygynous individuals suggest that these latter are in fact mainly driving the average effect found in subsection 4.1. However this is not a surprising result since most of the variation in ancestral inheritance norm is found among non-polygynyous females.[50]

[48]Following Anderson (2018), I use the dataset from La Porta et al. (2008) to identifiy common law and civil law countries.

[49]In my final sample, in common law countries about 29% of females originate from an ancestrally matrilineal ethnic group vs. about 5% in civil law countries.

[50]In my final sample, about 17% of non-polygynous females originate from an ancestrally matrilineal ethnic group vs. about 7% of

Ancestral Matrilocality. Matrilocality, an ethnic-group traditional norm according to which a married couple is supposed to live with or close to the wife's family, is highly correlated with matrilineality in Sub-Saharan Africa. In my main regression sample, about 92% of matrilocal individuals are matrilineal and about 94% of non-matrilocal individuals are non-matrilineal. Hence, ancestral matrilocality is hardly separable from ancestral matrilineality in my conceptual framework, and can be seen as an additional dimension of women's position in ancestral matrilineal societies. For these reasons, I decide to not control for it in my preferred specifications. Nevertheless, as reported in Table A.5 in appendix, I perform a robustness check exercise to check that ancestral matrilocality is not driving my main results.

Several lessons can be drawn from this exercise. First, looking at column 1, my main findings are unchanged when I additionally control for ancestral matrilocality. Unsurprisingly, the adjusted r-squared remains also unchanged. Second, looking at column 2, my main findings remain valid for non-matrilocal individuals, who constitute the majority of my regression sample. All in all, these results suggest that it is very unlikely that my main findings are driven by differences between ancestrally matrilineal and patrilineal ethnic groups' adoption of a matrilocal traditional norm.

Geographic Channels. Several geographic factors of the spread of HIV have been highlighted in the literature. Corno and De Walque (2012) show that mine workers' international migration is a driver of the spread of HIV. Along the same line, Oster (2012b) highlights exports and road networks, and subsequent increase in movements of people and sexual contacts, as an other factor of HIV infection. An other long-term determinant of female HIV in Sub-Saharan Africa put forward by Cagé and Rueda (Forthcoming) is is the geography of Protestant and Catholic missions in the early 20th century, as well as their health investments. In the same vein, Teso (2019) shows that the slave trade in Sub-Saharan Africa had long-term effects on contemporary gender norms, and Bertocchi and Dimico (2019) underline that the slave trade was a driver of actual polygyny in Sub-Saharan Africa.

In my main specification, and in addition of (within-country) DHS survey region fixed effects which should capture most of these geographical variations, I also include a host of village-level geographical controls, including latitude, longitude, altitude, nightlight composite, population density (2010), distance to lake or coastline, distance to nearest international border, average time (minutes) required to reach a high-density urban center (2015), distance to nearest active mine, an indicator for the presence of an active mine within 1000km of the village, malaria incidence (2010), vegetation index, and indicators for the length of the growing season (see subsection 3.3). My main results being robust and almost unchanged following the inclusion of these controls (see Table 1, column 5), I can reasonably rule out the possibility that such competing geographic channels are driving my main results.

Ethnic Fractionalization and Polarization. Ethno-linguistic diversity has been emphasized by Tequame (2012) as a driver of risky sexual behaviour and subsequent HIV. As a matter of fact, as she argues, one mechanism might be that social sanction due to risky behaviour is less costly in heterogeneous societies rather than homogeneous ones. A second mechanism might be that information spreads more easily in homogeneous than in heterogeneous societies, because the former are more likely to have the same language, culture and networks. Since to be subject

polygynous females.

to social sanction individuals should be detected as having risky sexual behaviuor, individuals who want to keep risky sexual behaviour secret might find heterogeneous societies more favorable. In addition, social interactions, which might differ by the degree of ethnical homogeneity, provide information about the level of HIV/AIDS at community level, including information on infectious status and risky behaviour of partners. Therefore, to account for the possibility that ancestral matrilineality might be associated with different degrees of ethnical heterogeneity, I control for an index of ethnic fractionalization and an index of ethnic polarization, which I compute at the village level using information on individual's ethnicity, following Montalvo and Reynal-Querol (2005) formulas.[51]

Economic Development. Differences in economic development may trigger differences in HIV rates. For example, legal systems and institutions are well-know driver of economic development (see La Porta et al., 2008 for a review). Such differences are controlled for both at the country and within-country level with the inclusion of (within-country) region fixed effects in my regressions. I also control for potential differences at the village-level by including, as proxies of village-level economic development, village's nighlight composite as well as village's population density. I finally control for such diferences at the ethnic group level by controlling for ancestral settlement patterns; ancestral juridictional hierarchies beyond local communities; ancestral reliance on hunting; ancestral reliance on fishing; ancestral reliance on gathering; ancestral reliance on animal husbandry; ancestral reliance on agriculture; ancestral presence of large domesticated animals; intensity of ancestral agriculture; and year of observation of the ethnic group (to alleviate the concern that some groups were portrayed later than others and might therefore have been more developed).

I find robust and almost unchanged estimates of the long-term effect of ancestral matrilineality on female HIV once these controls are included (see column 5 in Table 1), thus alleviating the concern that differences in economic development may drive my results.

Controlling for gender-specific alternative channels. So far, my OLS estimates are robust to a wide range of alternative channels that may spuriously drive my results on HIV. As a further robustness check, I check that my results are not driven by gender-specific alternative channels. In other words, I further check that alternative channels discussed so far are not affecting female populations differently. To do so, I additionally control for my set of controls and fixed effects interacted with a "Female" dummy. Results from this robustness exercise are reported in Table A.6. While I find that the effect of being a female is affected by the inclusion of this new host of gender-specific controls, my main result on the long-lasting effect of matrilineality on female populations remains unaffected.

Controlling for additional observables. Despite the emphasis put so far in controlling for numerous alternative channels, based on observables computed at either the individual, Ethnographic Atlas (EA) ethnic groups or DHS village level, I intend here to control more directly for alternative channels discussed above, adding covariates computed at the Murdock's ethnic group level, and based on Nunn (2010), Nunn and Wantchekon (2011) and Teso (2019) datasets. The limitation of this exercise is that I cannot match DHS ethnic groups with Murdock's ethnic groups as extensively as I did when I matched DHS ethnic groups with EA ethnic groups. Therefore, adding these new covariates will restrict my sample size. However, the value of this exercise is to assess the robustness and the

[51] These formulas and their interpretations are reported in subsubsection A.2.3 in appendix.

stability of my estimates when explicitly controlling for additional alternative channels. In this way, I more explicitly control for the slave trade alternative channel by controling for the logarithm of 1 plus the number of slaves taken in the transatlantic and in the indian slave trade from the Murdock ancestral ethnic group, divided by the area of the land historically inhabited by the ethnic group. Then, I aim to control for ethnic group's contact with colonizers during colonization by computing a dummy for whether a European explorer's route traveler crossed the historical land of the ethnic group, and a dummy for whether part of the railway network built by the Europeans was on the land of the ethnic group. Further, I also control for the differential effects of the different types of religious missions (Cagé and Rueda, Forthcoming) by including in my regression the number of catholic, protestant and British and Foreig Bible Society (BFBS) missions per square kilometer of an ethnic group's land during the colonial period. Additionally, I directly control for the minimum distance of the centroid of the land historically inhabited by the ethnic group from the routes of the Saharan trade, and from the closest city of the Saharan trade. Finally, I also estimate my regressions by controlling for the number of cities with more than 20,000 inhabitants that were present in 1400 on the land inhabited by the ethnic group, and for the number of conflicts between 1400 and 1700 in the area inhabited by the ethnic group, based on Besley and Reynal-Querol (2014) original dataset.

In the same spirit, I also control for additional covariates, computed at the DHS village level, which likewise reduce my sample size but allow me to assess the robustness of my estimates. More specifically, using answers from male respondants in DHS[52], I compute village's proportion of male circumcised as well as village's proportion of males who report having paid for sex in the last 12 months. These two covariates are intended to control for circumcision and prostitution[53], which have both been emphasized in the literature as an impediment and a driver of HIV, respectively.

Table 2 reports the OLS estimates of this additional robustness exercise and underlines that my OLS estimates are remarkably consistent when controlling for all of these covariates both separately and simultaneously. Thus, according to my fully-controlled OLS regression reported in column 9, I still find that females are more likely to suffer from HIV than their male counterpart. Further, I also still find that female originating from an ancestrally matrilineal ethnic group are 1.7 percentage point more likely to suffer from HIV than their patrilineal counterparts, an effect of very similar magnitude than the effect I found in my main specification, in column 5 of Table 1.

[52]The reduction in sample size when adding these controls is notably explained by the fact that only females are interviewed in some DHS villages.

[53]Unfortunately for this study, female respondants have not been interrogated in DHS about prostitution.

Table 2: **Robustness of OLS Estimates to Additional Controls**

	HIV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Matrilineality	0.001	-0.002	0.002	0.001	0.001	0.002	-0.007	-0.007	-0.013
	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.010)
Female	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***	0.009***	0.008***	0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.017***	0.017***	0.017***	0.016***	0.016***	0.016***	0.017***	0.017***	0.017***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Murdock Ethnic Groups Historical Controls:									
ln(1 + Total slave exports/area)	Yes								Yes
Contact explorer route		Yes							Yes
Colonial railway		Yes							Yes
Catholic missions/area			Yes						Yes
Protestant missions/area			Yes						Yes
BFBS missions/area			Yes						Yes
Distance Saharan route				Yes					Yes
Distance Saharan node				Yes					Yes
Cities in 1400					Yes				Yes
Precolonial conflicts (1400-1700)						Yes			Yes
DHS Villages Contemporary Controls:									
Village's prop. of males circumcised							Yes		Yes
Village's prop. of males who paid for sex								Yes	Yes
Observations	167,485	167,485	167,485	169,510	169,510	169,510	115,620	136,586	106,363
Adj. R-squared	0.080	0.080	0.080	0.080	0.080	0.080	0.090	0.086	0.092
Clusters	83	83	83	84	84	84	73	74	58
Mean Dep. Var. (Patri. Males)	0.024	0.024	0.024	0.024	0.024	0.024	0.026	0.023	0.026

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group (it therefore excludes individuals originating from ethnic groups with alternate inheritance rules (ambilineality, bilinearity, duolinearity, etc.)). **"HIV"**, "Matrilineality", "Female", **"Female × Matrilineality"**, "Ind. Controls", "Ethnic Group Controls", "Village-Geographic Controls", and "Region-survey FE" are defined in Table 1. "ln(1 + Total slave exports/area)" is the logarithm of 1 plus the number of slaves taken from the respondent's ethnic group in the transatlantic and/or the indian slave trade divided by the area of land historically inhabited by the group. "Contact explorer route" is a dummy taking value one if an European explorer route crossed the land of the ethnic group. "Colonial railway" is a dummy taking value one if a part of the railway network built by the Europeans was on the land of the ethnic group. "Catholic missions", "Protestant missions" and "BFBS missions" are the number of catholic, protestant, British and Foreign Bible Society (BFBS) missions per square kilometer of an ethnic group's land during the colonial period, respectively. "Distance Saharan route" and "Distance Saharan node" are the minimum distance of the centroid of the land historically inhabited by the ethnic group from the routes of the Saharan trade and from the closest city of the Saharan trade, respectively. "Cities in 1400" is the number of cities with more than 20,000 inhabitants that were present in 1400 on the land inhabited by the ethnic group. "Precolonial conflicts (1400-1700)" is the number of conflicts between 1400 and 1700 in the area inhabited by the ethnic group. "Village's prop. of males circumcised" is the within DHS village's proportion of male respondants circumcised. "Village's prop. of males who paid for sex" is the within DHS village's proportion of male respondants who report having paid for sex in the last 12 months. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

4.2.4 Assessing Selection on Unobservables

Despite my attempts to control for a large set of observable factors, at both the individual, ethnic group and village level, the estimates reported in Table 1 may still be biased by unobservable factors correlated with both ancestral matrilineality and females' contemporary sexual and contraceptive behaviours, and subsequently HIV. A priori, the direction of the potential omitted variable bias is not clear. Consider for instance the possibility that ethnic groups where women historically benefited from more sexual freedom adopted matrilinelity as a resulting adaptative kinship organization. As such, some evolutionary anthropologists explain the existence of matrilineal societies as the result of an evolutionary process that created institutions suitable for the ecological and social environment.[54] In particular, matrilineality may be advantageous in environments with low paternal certainty since while it is difficult to confirm paternity, maternity is easily observable. Thus, an inheritance system in which property passes from the mother's brother to her sons may be optimal since the brother knows he is related to his sister, but cannot verify that he is related to his children (Fortunato, 2012). This could drive the OLS estimates away from zero. On the other side, mating process might be inherently more assortative in matrilineal societies, relative to patrilineal societies, yielding matrilineal females suffering from relatively lower marital dissatisfaction. If that was the case, this might lead to matrilineal females being relatively less likely to adopt risky sexual behaviours[55], and therefore drive the OLS estimates towards zero. In this subsection I consequently assess the likelihood that the OLS estimates might be biased by unobservables.

Coefficients Ratio Tests (Altonji et al., 2005). I start by assessing the sensitivity of the OLS estimates to controlling for observable characteristics. To do so, I first employ the strategy adapted by Nunn and Wantchekon (2011) from Altonji et al. (2005) that allows to determine how much stronger selection on unobservables would have to be compared to selection on observables to fully explain away my results. To perform this test, I calculate the ratio $\hat{\beta_F}/(\hat{\beta_R} - \hat{\beta_F})$, where $\hat{\beta_F}$ is my coefficient of interest from a regression that includes my full set of controls, while $\hat{\beta_R}$ is my coefficient of interest from a regression that includes a restricted set of controls. The intuition behind the formula is straightforward. First, consider why the ratio is decreasing in $(\hat{\beta_R} - \hat{\beta_F})$. The smaller is the difference between $\hat{\beta_R}$ and $\hat{\beta_F}$, the less the estimate is affected by selection on observables, and the stronger selection on unobservables needs to be (relative to observables) to explain away the entire effect. Next, consider the intuition behind $\hat{\beta_F}$ in the numerator. The larger $\hat{\beta_F}$, the greater is the effect that needs to be explained away by selection on unobservables, and therefore the higher is the ratio (Nunn and Wantchekon, 2011).

The results are reported in columns 1 of Table A.7, where each row reports result for different set of restricted covariates. This yields four ratios that range from -509.27 to 167.42. In some cases, the coefficient in the fully-controlled model is larger than that of the uncontrolled model, giving a negative ratio. In general, these ratios are far greater than 1 in absolute value, and therefore suggest that the influence of unobservable characteristics would have to be far greater than the influence of observable characteristics to fully account for my OLS findings.

Minimum Coefficient Lower Bound (Oster, 2017). Further, I also use the method from Oster (2017) to

[54] Lowes (2018a) appendix provides an extensive overview of the hypothesized origins of matrilineal kinship systems.

[55] In fact Bertocchi and Dimico (2019) show that, due to marital dissatisfaction, women in polygynous union in Sub-Saharan Africa adopt riskier sexual behaviours which are more conducive to HIV.

calculate a bias-adjusted lower bound of my coefficient of interest. Oster shows that if one assumes that observables and unobservables have the same explanatory power in the outcome variable, then the following is a consistent estimator: $\beta^* = \hat{\beta_F} - (\hat{\beta_R} - \hat{\beta_F}) \times ((R^2_{Max} - \hat{R^2_F})/(\hat{R^2_F} - \hat{R^2_R}))$, where $\hat{\beta_R}$ and $\hat{\beta_F}$ are defined as above; $\hat{R^2_F}$ is the R-squared from the fully-controlled regression; $\hat{R^2_R}$ is the R-squared from the restricted regression; and R^2_{Max} is the R-squared from a regression that includes all observable and unobservable controls. Although in theory the maximum possible value of R^2_{Max} is one, as underlined by González and Miguel (2015), in the real world, where there is significant measurement error, the value of R^2_{Max} should be much lower than one. In fact, by definition, $R^2_{Max} \in [\hat{R^2_F}; 1]$. Oster (2017) provides some insights on how R^2_{Max} should be chosen, showing that $R^2_{Max} = 1$ may lead to over-adjustment in many cases. I follow her procedure and present bias-adjusted lower bound coefficients for $R^2_{Max} = min(\mathbf{1.3}\hat{R^2_F}; 1)$[56], $R^2_{Max} = min(\mathbf{1.5}\hat{R^2_F}; 1)$, and $R^2_{Max} = min(\mathbf{2}\hat{R^2_F}; 1)$ in column 2, 3 and 4 respectively (I also report bias-adjusted lower bound coefficients for $R^2_{Max} = 1$ in column 5 for informational purpose).

All bias-adjusted lower bound estimates from this exercise are reported in Table A.7 and remain positive and, taken at face value, still imply a sizeable estimated effect of ancestral matrilineality on female HIV, of same order of magnitude than previously found in my OLS regressions. Further, it is worth noting that the full set of these biased-adjusted lower bound estimates (column 2 to 5) falls within the 99.5% confidence interval of my fully-controlled OLS estimate (column 7), which suggests that the size of the estimate from the OLS regression with full controls is similar to the bias-adjusted estimates. All in all, these tests suggest that my fully-controlled OLS estimates are very unlikely to be affected by omitted variable bias, and therefore support a causal interpretation of my OLS findings.

4.3 Alternative Identification Strategies

Though my OLS estimates are robust to a large set of observables that could be potential confounders, and are unlikely to be affected by omitted variable bias, reverse causality may remain a concern. For example, matrilineality might have been an adaptative response in environments with low paternal certainty (see discussion in subsubsection 4.2.4). If environments with low paternal certainty were associated with more promiscuous sexual behaviours, OLS estimates would be biased away from zero.

In order to provide further support to the causal interpretation of my main individual-level OLS findings, I implement three alternative identification strategies. These strategies have the common feature to allow me to estimate the effect of the variation in cultural trait within similar environments. To do so, I exploit data on the GPS location of DHS villages and the digitized Murdock's map of ancestral ethnic groups in Africa.

4.3.1 Accounting for Unobservables: Geographic RD Estimates

As a first alternative identification strategy, I undertake a geographic regression discontinuity analysis. More precisely, I examine and compare individuals living in villages geographically close, but where some villages are located within the ancestral boundaries of an ancestrally matrilineal ethnic group and the other within the ancestral boundaries of an ancestrally non-matrilineal (i.e. patrilineal or other) ethnic group. In this framework, the ancestral matrilineal ethnic boundary is the delineation created by the ancestral borders of ethnic groups that practiced

[56] Oster (2017) suggests using $R^2_{Max} = 1.3\hat{R^2_F}$ as a cutoff to test for the stabillity of a treatment effect consistent with randomized treatment.

matrilineal descent alongside groups that practiced patrilineal or alternate descent (based on digitized Murdock's map of ancestral ethnic groups in Africa, see Figure 2). The intuition behind this specification is that the ancestral matrilineal ethnic boundary is determined by the ancestral borders of multiple matrilineal and non-matrilineal ethnic groups. The boundaries between these multiple ethnic groups are arbitrary, and along the border geographic areas are quite similar.[57]

Therefore, my strategy is to use a regression discontinuity (RD) estimation method that restricts the sample to individuals living in villages that are sufficiently close to the ancestral matrilineal ethnic boundary and estimate the causal effect of living in a village located on the matrilineal side on female HIV. Importantly, I do not estimate the effect of originating from an ancestrally matrilineal ethnic group anymore, but the effect of living in a village located in an ancestrally matrilineal geographic area. The benefit of this strategy is that it accounts for unobservable factors that vary smoothly across space. Therefore, as long as the determinants of unobservable traits (e.g. geography, history, idiosyncratic shocks, state presence etc.) vary smoothly, the unobservable traits will be accounted for by the RD strategy. Further, in order to get the more conservative estimates, I directly control for the large array of individual, ethnic group's historical and village-level geographic control, as well as (within-country) region × survey (year) fixed effects described in the previous sections. More specifically, my RD specification takes the following form:

$$\begin{aligned} y_{ievrt} = \alpha + \beta_1 Matrilineality_v + \beta_2 Female_i + \beta_3 Female_i \times Matrilineality_v + \\ \mathbf{X'_{ievrt}}\Delta + \mathbf{X'_{ert}}\Omega + \mathbf{X'_{vrt}}\Pi + f(location_v) + \lambda_{rt} + \varepsilon_{ievrt} \end{aligned} \tag{2}$$

with y_{ievrt} denoting an individual-level outcome (indicator for whether the individual is HIV positive) for individual i from ethnic group e, living in village v in within-country DHS region r, and surveyed at year t. $Matrilineality_v$ is an indicator for whether an individual lives in a village located on an ancestrally matrilineal geographic area; $Female_i$ is an indicator for whether an individual is a female. β_1 is intended to capture a "Matrilineal effect"; and β_2 is intended to capture a "Gender effect" on HIV. The coefficient of interest β_3 captures the effect of living in a village located in an ancestrally matrilineal geographic area on female HIV once the "Matrilineal effect" and the "Gender effect" have been controlled for. $f(location_v)$ denotes a a RD polynomial that controls for a smooth function of the geographic location of DHS villages. In my specifications I alternatively use the minimum distance to the nearest ancestral matrilineal ethnic boundary (in km) and the gps coordinates (latitude and longitude) of the village as running variables. Further, I use several functional forms of the polynomial, using polynomials of different orders, and alternatively estimating them separately on each side of the boundary ("flexible polynomials"). $\mathbf{X'_{ievrt}}$; $\mathbf{X'_{ert}}$, and $\mathbf{X'_{vrt}}$ represent a set of individual-level, ethnic group-level ancestral, and village-level geographic covariates respectively, which are defined in subsection 3.3. λ_{rt} denotes (within-country) DHS region-survey (year) fixed effects. Standard errors ε_{ievrt} are clustered at the ethnic group level, and the sample is restricted to individuals living in villages that are within a certain distance to the ancestral matrilineal ethnic boundary, either 100, 150, or 200 kilometers.

[57]Note that the ancestral matrilineal ethnic boundary does not coincide with any actual border.

Validating the Assignment of Matrilineal Individuals. The boundaries used for my RD estimates are from Murdock (1959, see Figure 2), a source that has been used previously in a number of studies that use a similar RD approach (see Moscona et al. (2020) for a recent and related example). However, an important assumption when using the ethnic boundaries is that they accurately reflect true discontinuities of ethnic affiliation today. This is particularly important since, in reality, one may not observe clear borders between ethnic groups, and instead only a gradual change of the mix of ethnicities over space. Further, an additional assumption that I have made so far is that the matching between individual's self-reported ethnicity in DHS and ancestral ethnic groups in Ethnographic Atlas is accurate.[58] Therefore, I now check the validity of my use of Murdock's ethnic boundaries by examining how individual's ancestral matrilineality varies at ancestral matrilineal ethnic boundaries. This is shown in Figure A.4 in appendix, which reports the bivariate relationship between distance from the ancestral matrilineal ethnic border and individual's ethnic group's ancestral matrilineality. The y-axis displays the fraction of the population in a 5 km bin that reports that they are a member of an ancestrally matrilineal ethnic group, and the x-axis is distance in kilometers from the border. A positive distance indicates a location within the ancestral territory of an ancestrally matrilineal ethnic group and a negative distance indicates a location within the ancestral territory of an ancestrally non-matrilineal (i.e. patrilineal or other) ethnic group. Reassuringly, I find that there is a discontinuous change at the border in the fraction of the population that report being member of an ancestrally matrilineal ethnic group.[59]

Geographic RD Estimates. Before turning to my estimates I first examine the raw data of the RD sample. Figure A.5 in appendix shows a bin scatterplots of the predicted HIV rate for females living in villages located within 150 km of the ancestral matrilineal ethnic border, using a flexible third-order RD polynomial conditioned on region (within-country) $\times$ survey (time) FE and estimated separately on each side of the border. Positive values, on the horizontal axis, reflect 5 km bins in ancestrally matrilineal geographic areas and negative values reflect 5 km bins in ancestrally non-matrilineal geographic areas. Even in the raw data, a discontinuity at the border is apparent: a discontinuous increase in female HIV rate on the matrilineal side of the border can be observed. I next turn to my more formal RD estimates.

Table 3 reports the geographic RD estimates for different bandwidths: 100 km (columns 1 and 2), 150 km (columns 3 and 4), and 200 km (columns 5 and 6); different running variables: minimum distance in km between DHS village and nearest ancestral matrilineal ethnic boundary (columns 1, 3 and 5), as well as village's latitude and longitude (columns 2, 4 and 6); and different polynomial specifications: linear polynomial, flexible linear polynomial, quadratic polynomial, flexible quadratic polynomial, cubic polynomial and flexible cubic polynomial (with "flexible" standing for polynomials estimated separately at each side of the boundary). Several lessons can be drawn from these estimates. First, I find a remarkably consistent and significantly positive "Gender effect", consistent with the "feminization of HIV" in Sub-Saharan Africa, already extensively discussed in the literature and in section 4. Second, and more importantly, I find a positive estimate of $Female \times Matrilineality$, significant when analysing the samples of individuals living in villages located within 150 or 200 km of the nearest ancestral matrilineal ethnic

[58] This is important since individual's ethnic group's ancestral matrilineality variable is based on kinship organization of ancestral ethnic groups reported in the Ethnographic Atlas.

[59] It is however important to note that, while information on individual's ethnic group's ancestral matrilineality is based on the matching between individual's self-reported ethnicity in DHS and ancestral ethnic groups in Ethnographic Atlas, these latter slightly differ from ethnic groups classification in Murdock's Map of ancestral ethnic groups (1959). This might partially explain why the discontinuous increase very close to the boundary is of limited size.

border. This result corroborates my OLS finding that females originating from ancestrally matrilineal ethnic groups suffer from significantly more HIV today.

Noteworthy, the magnitudes of my geographic RD estimates of $Female \times Matrilineality$ are lower than the magnitudes of my OLS estimates. A first explanation stems from the fact that in this analysis $Matrilineality$ indicates whether an individual lives in a village located within the ancestral boundaries of an ancestrally matrilineal ethnic group, instead of an individual's ethnic group's ancestral matrilineality. In addition, although one other explanation for this could be a potential bias from unobservables present in my OLS estimates (which is however unlikely according to the tests I performed in subsubsection 4.2.3), the difference in magnitudes might also be explained by the fact that within an ancestral matrilineal ethnic group's territory, and close to the border, only a fraction of the population is today likely to belong to an ancestrally matrilinal ethnic group. As shown in Figure A.4, close to the border on the matrilineal side approximately 60% to 40% of the population does not belong to an ancestrally matrilineal ethnic group. This suggests that the magnitude of the RD estimates could be biased downwards by this amount as well. In addition, ancestral ethnic groups boundaries are susceptibly blurry, and spillovers from matrilineal group to patrilineal group may arise close to the boundary, especially when it comes to sexual relationships and sexually transmitted diseases. Higher ethnic fractionalization and/or polarization at the boundary may help explain such spillovers since, as underlined by Tequame (2012), riskier sexual behaviours are more easy to conceal in fragmented societies. Alternatively, migration at the boundary could also increase sexual mixing between individuals originating from ancestrally matrilineal ethnic groups and those originating from ancestrally patrilineal ethnic groups. All in all, such spillovers would cause estimated effects at the border to be muted, and partly explain why the estimated effects fall above conventional significance levels when restricting the sample to individuals living very close to the border (i.e. within 100km).[60]

4.3.2 Instrumenting for Ancestral Matrilineality

As a second alternative identification strategy, I perform an instrumental variable (IV) strategy, instrumenting individual's ethnic group's ancestral matrilineality ($Matrilineality$ regressor) with a measure of the minimum distance (in km) between the individual's DHS village to the nearest ancestral matrilineal ethnic boundary (based on digitized Murdock's map of ancestral ethnic groups in Africa, see Figure 2). Also, I instrument my main regressor of interest $Female \times Matrilineality$ with a variable interacting $Female$ dummy with this measure of distance.[61]

The relevance of this instrument is *a priori* straightforward and validated by the bivariate relationship between distance from ancestral matrilineal ethnic border and individual's ethnic group's ancestral matrilineality (see subsubsection 4.3.1 and Figure A.4). The critical issue is whether the distance between an individual's location and ancestral matrilineal ethnic border is uncorrelated with factors, other than individual's ethnic group's ancestral matrilineality, that may have affected individual's sexual and contraceptive behaviour and therefore HIV susceptibility. Therefore, there remain a number of other reasons why the exclusion restriction may not be satisfied.

[60] The reduction in sample size is also likely at play in explaining why the effect is less precisely estimated.

[61] I assign positive values to this measure of distance for DHS villages located within the boundaries of a matrilineal ancestral ethnic groups, while I assign negative values for villages located within the boundaries of a non-matrilineal (i.e. patrilineal or other) ancestral ethnic group. Therefore, a distance of +100 means that a village is located within an ancestrally matrilineal area, 100 km away from the nearest matrilineal/non-matrilineal ancestral ethnic boundary; while a distance of -100 means that a village is located within an ancestrally non-matrilineal area (i.e. patrilineal or other), 100 km away from the nearest matrilineal/non-matrilineal ancestral ethnic boundary.

Table 3: **Geographic RD Estimates**

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep. Var.: HIV					
Distance from matrilineal boundary:	100km		150km		200km	
Running variable:	*Distance*	*Lat./Long.*	*Distance*	*Lat./Long.*	*Distance*	*Lat./Long.*
	Panel A: Linear Polynomial					
Matrilineality	-0.000	0.000	-0.001	-0.001	-0.002	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel B: Flexible Linear Polynomial					
Matrilineality	0.005	0.002	0.000	0.001	-0.001	0.000
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel C: Quadratic Polynomial					
Matrilineality	0.000	0.001	-0.001	0.000	-0.001	-0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel D: Flexible Quadratic Polynomial					
Matrilineality	0.005	0.001	0.005	0.001	0.001	0.001
	(0.007)	(0.009)	(0.006)	(0.008)	(0.005)	(0.008)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel E: Cubic Polynomial					
Matrilineality	-0.000	0.001	-0.001	0.000	-0.001	-0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel F: Flexible Cubic Polynomial					
Matrilineality	0.007	-0.005	0.000	-0.007	0.005	-0.008
	(0.010)	(0.011)	(0.007)	(0.010)	(0.006)	(0.010)
Female	0.011***	0.011***	0.010***	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Matrilineality	0.008	0.008	0.009**	0.009**	0.009**	0.009**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Baseline Controls + FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48,947	48,947	71,333	71,333	89,982	89,982
Clusters	68	68	73	73	76	76
Mean Dep. Var.	0.034	0.034	0.033	0.033	0.032	0.032

Notes: Geographic RD estimates based on Equation 2 are reported with standard errors clustered at the ethnicity level in brackets, for different bandwidths, running variables and polynomial specifications. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group (it therefore excludes individuals originating from ethnic groups with alternate inheritance rules (ambilineality, bilinearity, duolinearity, etc.)). "Matrilineality" indicates (dummy) whether an individual lives in a village located within the ancestral boundaries of an ancestrally matrilineal ethnic group (based on Murdock's map of ancestral ethnic groups, see Figure 2). **"HIV"** and "Female" are defined as in Table 1. "Baseline Controls + FE" are the "Ind. Controls", "Ethnic Group Controls", "Village-Geographic Controls", and "Region-survey FE" defined in Table 1. ** p<0.10,** p<0.05,*** p<0.01*

First, distance between contemporary individual's location and ancestral matrilineal ethnic boundary may be correlated with geographic characteristics (e.g. vegetation, altitude, remoteness, etc.) which might affect the spread of HIV. Second, this contemporary measure of distance may also be correlated with the geographical distribution of economic activities, infrastructures, transportation networks, etc. and therefore reflects migration patterns that appeared in the meantime (i.e. since pre-colonial period) and which might have affected contemporary HIV rates. To alleviate such concerns I include in my IV regressions region (within-country) × survey (time) fixed effects, and I directly control for such alternative channels with a large array of ethnic group's historical controls and village-level geographic controls, which are already detailed in subsection 3.3 and subsubsection 4.2.3.

Table 4 presents estimates of my instrumental variable strategy. The OLS estimate of regressing my measure of distance on individual's ethnic group's ancestral matrilineality, as well as F-Stat of the test that the coefficient of the instrument is null are reported in column 2. They suggest that my instrument is a strong predictor of individual's ethnic group's ancestral matrilineality. Similar conclusions can be drawn from the OLS estimate and F-Stat of regressing $Female \times Matrilineality$ on the variable interacting the $Female$ dummy with my instrument, reported in column 3. More importantly, the IV-2SLS estimate reported in column 1 confirms my OLS result and indicate that women originating from ancestrally matrilineal ethnic groups significantly suffer from more HIV today.

Noteworthy, instrumenting two potentially endogenous regressors with two instruments and assuming clustered standard errors, I follow Andrews (2018) and compute weak instrument-robust 95% confidence interval of my IV estimate.[62] This latter is reported in square brackets in Table 4 and suggests a non-null positive IV estimate of $Female \times Matrilineality$.[63]

4.3.3 Nearest Neighbor Matching

As a final alternative identification strategy to individual-level OLS regressions, I use nearest neighbor matching to compare each DHS village located in an ancestrally matrilineal area to the DHS village located in an ancestrally non-matrilineal (i.e. patrilineal or other) area[64] that is the most similar in terms of geographic characteristics. The matching average treatment over treated effects are reported in Table 5, with village pairs being matched[65] on the full range of village geographic controls of Equation 1. As reported in column 1, villages located in ancestrally matrilineal areas are characterized by average female HIV rates that are 2.9 percentage points higher than their nearest neighbor village located in an ancestrally non-matrilineal area. This effect is significant at 1% and of large magnitude, representing about 78% of the average village's female HIV rate of villages located in an ancestrally non-matrilineal area. Interestingly, no such effect is found on village's proportion of HIV positives males[66], consistent with my hypothesis that the effect of ancestral matrilineality on contemporary HIV rates is specific to female populations.

[62] I use *"twostepweakiv"*, a Stata package developed by and presented in Sun (2018). I compute two-step weak instrument-robust 95% confidence interval, projected on $Female \times Matrilineality$ regressor.

[63] Calculating 95% two-step weak instrument-robust CI based on Andrews (2018) and Sun (2018), I find a 12% distortion cutoff, which suggests that my instruments are strong instruments, in the sense that size distortions are below 12% for conventional 95% confidence intervals.

[64] Based on Murdock's map of ancestral ethnic groups (see Figure 2).

[65] I use nearest neighbor matching based on Mahalanobis distance. Estimates are corrected for bias due to matching on multiple continuous variables, based on Abadie and Imbens (2006) and Abadie and Imbens (2011).

[66] The smaller sample size is explained by the fact that in some DHS villages only females were interviewed.

Table 4: **IV Estimates of the Effect of Ancestral Matrilineality on Female HIV**

	(1)	(2)	(3)
	IV-2SLS	**OLS (First Stages)**	
	HIV	**Matrilineality**	**Female × Matri.**
Matrilineality	-0.010		
	(0.036)		
Female	0.006***	Yes	Yes
	(0.002)		
Female × Matrilineality	0.029***		
	(0.010)		
	[0.013;0.069]		
Distance to matri. boundary		0.0003***	
		(0.0001)	
Female × Distance to matri. boundary			0.0008***
			(0.0002)
Ind. Controls	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes
Observations	179,312	179,312	179,312
Adj. R-squared	0.032	0.835	0.617
Clusters	100	100	100
Mean Dep. Var. (Patri. Males)	0.023	0.158	0.085
F-Stat (test coeff. = 0)		10.60	10.04

Notes: IV-2SLS estimates are reported in column 1; first stage OLS estimates are reported in column 2 and 3; with standard errors clustered at the ethnicity level in brackets. Weak instrument-robust 95% confidence interval (Andrews, 2018, Sun, 2018) is reported in square brackets (calculating 95% two-step weak instrument-robust CI, the distortion cutoff is 12%, indicating strong instruments). The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group (it therefore excludes individuals originating from ethnic groups with alternate inheritance rules (ambilineality, bilinearity, duolinearity, etc.)). "Distance to matri. boundary" is computed on QGIS as the minimum distance between DHS village and the nearest ancestral matrilineal ethnic boundary (based on Murdock's map of ancestral ethnic group boundaries, see Figure 2). This takes negative values for DHS villages located within boundaries of non-matrilineal ancestral ethnic groups. **"HIV"**, **"Matrilineality"**, "Female", **"Female × Matrilineality"**, "Ind. Controls", "Ethnic Group Controls", "Village-Geographic Controls", and "Region-survey FE" are defined in Table 1. * *p<0.10,** p<0.05,*** p<0.01*

5 Mechanisms

The previous section has uncovered a robust relationship between ancestral matrilineality and the contemporary spread of the HIV epidemic among female individuals in Sub-Saharan Africa. According to my conceptual framework, the legacy of ancestral matrilineality on women's sexual preferences and ability to implement them has led matrilineal women to adopt sexual and contraceptive behaviours that favor a higher rate of transmission of HIV today. In this section, I turn to a more direct investigation of these channels, by exploring the empirical relationship between ancestral matrilineality and female's sexual autonomy, sexual and contraceptive behaviour

5.1 Female Sexual Autonomy

First, I explore whether women originating from ancestrally matrilineal ethnic groups benefit from a higher social status, and consequent higher bargaining power, which would allow them to benefit from a higher sexual

Table 5: **Nearest Neighbor Matching (ATT)**

	(1) Village's proportion of HIV positive females	(2) Village's proportion of HIV positive males
Ancestrally matrilineal area	0.029***	0.007
	(0.011)	(0.011)
Observations	13,176	12,710
Mean Dep. Var. (Villages in non-matri. area)	0.037	0.023

Notes: This table reports the average treatment effect on the treated (ATT) on the proportion of village's HIV positive females in column 1, and males in column 2, between DHS villages located on an ancestrally matrilineal area and DHS villages located on an ancestrally non-matrilineal area (patrilineal or other), using nearest neighbor matching. Villages are matched using the Mahalanobis distance function based on all *Village-Geographic Controls* detailed in Table 1. Estimates are corrected for bias due to matching on multiple continuous variables (Abadie and Imbens (2006); Abadie and Imbens (2011)). Abadie and Imbens robust standard errors are reported in brackets. * *p<0.10,** p<0.05,*** p<0.01*

autonomy. First, building on the household economics literature highlighting the role of marriage outside option on intrahousehold bargaining power (Baland and Ziparo, 2018; Bargain et al., 2020), I create a dummy equals to one if a female is currently divorced. Then, following Anderson (2018), I use information provided in DHS on land and house ownership, as a measure of female property rights in case of divorce (restricting my sample to divorced women), and I create a dummy which is equal to one if a divorced female reports owning a house and/or a land. Finally, I construct a dummy equals to one if there is at least one wife's parent in the household.[67] Working with these outcomes allow me to check whether women originating from ancestrally matrilineal ethnic groups benefits from better marriage outside options and, consequently, higher bargaining power within marriage.

To further explore how this translates into female sexual autonomy, I emphasize a second set of outcomes which pertain to women's decision-making regarding contraception. In particular, I use answer to the question asked to women in union currently using contraception about who decides on the use of contraception, and I create a dummy equals to one if a woman reports being decision-maker. Other questions in DHS directly cover sex negotiations and autonomy and asks whether a woman could get a male condom herself, whether a wife is justified to ask husband to use condom if he has a STI (Sexually Trasmitted Infection); and whether it is justified for a women to refuse sex with her husband if he has another women. For each of these questions, I create a dummy equals to 1 if a woman answers affirmatively.

For the analysis in this subsection, I focus on females and estimate the following equation:

$$y_{ievrt} = \alpha + \beta_1 Matrilineality_e + \mathbf{X'_{ievrt}}\Delta + \mathbf{X'_{ert}}\Omega + \mathbf{X'_{vrt}}\Pi + \lambda_{rt} + \varepsilon_{ievrt} \quad (3)$$

where $\mathbf{X'_{ievrt}}$, $\mathbf{X'_{ert}}$, $\mathbf{X'_{vrt}}$, λ_{rt} and ε_{ievrt} are individual-level, ethnic group-level and village-level controls and (within-country) DHS region-survey (year) fixed effects, defined as in subsection 3.3. Standard errors are clustered at the ethnic group level. My coefficient of interest is now β_1 and captures the long-term effect of ancestral

[67] Bargain et al. (2020) show the influence of the presence of wife's family in the household on wife's intrahousehold decision-making, in the Indonesian context.

matrilineality on female populations.[68]

In Table 6, I report results from estimating Equation 3 on these outcomes. I find that ancestral matrilineality is indeed significantly positively associated with all these dimensions of female's social status, bargaining power and sexual autonomy. More specifically, according to the estimates reported in column 1, matrilineal women are 1.5 percentage points significantly more likely to be divorced, meaning twice more likely to be divorced than patrilineal females.[69] Also, according to the estimate reported in column 2, I find that divorced women originating from ancestrally matrilineal ethnic group are 23.5 percentage points (representing about 65% of the mean probability of the patrilineal females in my sample) more likely to own a house and/or a land, as compared to their divorced patrilineal counterparts. I additionally find in column 3 that matrilineal females are 0.08% more likely to have at least one parent present in their household, an effect of large magnitude given that only 2% of patrilineal females benefit from the presence of at least one of their parents in their household. This result is consistent with the correlation underlined in subsubsection 4.2.3 between ancestral matrilineality and ancestral matrilocality. Further, estimates reported in column 3 to 6 indicate, in the same order, that matrilineal women are 4.3 percentage points more likely to be decision-maker regarding the use of contraception; 2.6 percentage points more likely to answer that they could get (male) condom themselves; 3 percentage points more likely to find justified for a woman to ask her husband to use a condom if he has a STI; and 2.8 percentage points more likely to find justified for a woman to refuse sex with her husband if he has another women.

All in all, in line with my conceptual framework, I can rule out Anderson (2018) mechanism, according to which HIV prevalence should be higher for less empowered females since they should be less able to impose safe sexual practices to their husbands. I propose two main other mechanisms in the two next subsections. First, I explore whether the higher sexual autonomy of women originating from ancestrally matrilineal ethnic groups translates into them implementing their preferred sexual strategy, and thereby translates into them adopting riskier sexual behaviours. Further, I explore how matrilineal females' higher decision-making power regarding contraception translates into contraceptive behaviours that are *incidentally* less protective.

[68] Remember that I restrict my sample to individuals originating from an either ancestrally matrilineal or an ancestrally patrilineal ethnic group. Therefore, $Matrilineality = 0$ means that the individual originates from an ancestrally patrilineal ethnic group (comparison group).

[69] The mean proportion of divorced patrilineal females in my regression sample is 1.5%.

Table 6: **Ancestral Matrilineality and Female Sexual Autonomy (OLS)**

	Divorced	Own house and/or land (divorced)	Wife's Parents in HH	Decide Contraception	Can get condom	Wife justified ask condom if husband has STI	Wife justified refuse sex if husband has other women
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Matrilineality	0.015***	0.235**	0.008***	0.043***	0.026**	0.030***	0.028**
	(0.004)	(0.108)	(0.003)	(0.010)	(0.013)	(0.011)	(0.013)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	105,964	1,191	66,902	13,709	56,270	91,420	95,604
Adj. R-squared	0.041	0.316	0.026	0.054	0.123	0.157	0.139
Clusters	100	54	100	79	99	100	99
Mean Dep. Var. (Patri.)	0.015	0.359	0.020	0.871	0.632	0.772	0.622

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is a woman originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Divorced" is a dummy indicating whether an individual is currently divorced. "Wife's Parents in HH" is a dummy indicating the presence of at least one wife's parents in household. "Own house and/or land (divorced)" is a dummy indicating whether an individual owns a house and/or a land (divorced females sample). "Decide contraception" is a dummy indicating whether an individual is decision-maker regarding contraception. "Can get condom" is a dummy indicating whether an individual can get herself a male condom. "Wife justified ask condom is husband has STI" is a dummy indicating whether an individual find justified for a wife to ask her husband to use a condom if he has STI. "Wife justified refuse sex if husband has other women" is a dummy indicating whether a respondant find justified for a wife to refuse sex with her husband if he has other women. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10,** p<0.05,*** p<0.01*

5.2 Sexual Behaviour

Reproductive Evolved Psychology and Sexual Behaviour. In order to capture sexual behaviours that are prominent risk factors of HIV contagion, I follow the well-established medical and economic literature (e.g. Bertocchi and Dimico, 2019), and I create a dummy equals to one if an individual had any sexual activity in the last 4 weeks, a dummy equals to one if an individual reports having any extramarital affair in the last 12 months, and the number of extramarital partners in the last 12 months that she reports (I focus on formally married individuals for these two latter outcomes).

Several lessons can be drawn from Table 7, which presents results from estimating Equation 1 on these sexual behaviour outcomes. To begin with, according to the estimates reported in the second row, sub-saharan african women are significantly and sizably less likely to report adopting risky sexual behaviours than their male counterparts. This result is consistent with the evolutionary psychology theory I build on in my conceptual framework, as well as with numerous psychological studies (Buss, 2016) according to which, relative to females, males exhibit a stronger preference for casual relationships and sexual variety. However I also find that, among females, women originating from ancestrally matrilineal ethnic groups have significantly riskier sexual behaviours, that are more conducive to HIV. More precisely, estimates reported in the third row indicate that matrilineal married women are 4.1 percentage points more likely to have had a sexual activity in the last 4 weeks (column 2). Interestingly, I find qualitatively similar results when working on individuals being in any marital status, while controlling for this latter (column 1). Further, I find that matrilineal married females are 2.6 percentage points (representing about 20% of the patrilineal married males' mean probability) more likely to report any extramarital affair in the last 12 months. In addition, estimate from column 4 indicates that married women originating from an ancestrally matrilineal ethnic group had 0.068 more extramarital partners in the last 12 months, an effect of large magnitude representing about 33% of the mean number of extramarital partners of patrilineal married males. Finally, it is worth noting that according to the estimates reported in the first row, and contrary to their female counterparts, matrilineal males do not have an overall signficantly different propensity to adopt risky sexual behaviour than their patrilineal counterparts. All these results are fully consistent with my conceptual framework and highlight mechanisms at play in explaining the long-term effect on ancestral matrilineality of contemporary female HIV in Sub-Saharan Africa.

Routes of Infection. Building on my conceptual framework and the results on sexual behaviours highlighted above, I expect matrilineal females to be mainly infected through extramarital routes of infection. This is exactly what I find in the analysis on couple's serodiscordancy status, performed in Table A.8.

Restricting my attention to non-polygynous formally married couples with both wife and husbands tested for HIV in DHS and both of them originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group, I find that when the wife originates from an ancestrally matrilineal ethnic group, she is about twice more likely (i.e. 1.3 percentage points) to be HIV positive while having a HIV negative husband, relative to wives originating from ancestrally patrilineal ethnic groups. Importantly, this result remains marginally significant when I also include individual's, (ancestral) ethnic group's and village-geographic controls computed for the husband, in addition of controlling for whether the wife and the husband originate from different ancestral ethnic groups (*"Mixed ethnicity"*), as well as controlling for whether the wife and the husband originate from ethnic groups with different ancestral kinship organizations (matrilineal vs. patrilineal) (*"Mixed matrilineality"*). All in all, these results indicate that, in line with their riskier sexual behaviour outside the domestic sphere, matrilineal women are significantly more likely to be infected by HIV through extramarital channels.

Moreover, results in the two last columns indicate that matrilineal couples are relatively less likely to be HIV+ seroconcordant. This finding may be interpreted as an evidence that the mechanisms highlighted by Anderson (2018) is also at play, namely that patrilineal females are more likely to be infected by their husband due to their lower sexual autonomy. Alternatively, this finding may also be explained by differential couple formation/dissolution dynamics between matrilineal and patrilineal groups (e.g. a matrilineal female may be relatively more likely to divorce[70] if her husband is HIV positive. In this case, the couple is dropped from my sample of couples, leading to non-random attrition).

While DHS data do not allow me to directly empirically test for the effect of matrilineality on gender-related social behaviour, it is worth noting that a recent literature has highlighted the influence of matrilineality on such gender related behaviour. In particular, Lowes (2018a) provides experimental evidence from Democratic Republic of Congo (DRC) that spouses from matrilineal ethnic groups cooperate less with each other than their patrilineal counterparts. Further, Lowes (2018b) provides, in a similar context, experimental evidence that matrilineality closes the gender gap in risk-preference, with matrilineal women having a higher preference for risk than their patrilineal counterparts. These differences in contemporary behaviours could be additional factors shaping riskier sexual behaviour found here, and within marriage contraceptive behaviour explored in the next subsection.

5.3 Contraceptive Behaviour

Condom versus Long-Term Contraceptive Methods. The results highlighted in subsection 5.1 indicated that matrilineal females were more likely to be decision-maker regarding contraception, suggesting that matrilineal females may also differ from their patrilineal counterparts in their contraceptive behaviour. This is what I investigate in this subsection.

To do so, I exploit information in DHS on current contraception method used by respondant to create (1) a dummy for whether the respondant reports using condom as her current contraceptive method; (2) a dummy for whether the respondant reports using a long-term contraceptive method (i.e. "pill", "IUD", "injection", "female sterilization", "implants/norplant", or "lactational amenorrhea (LAM)"); (3) and a dummy for whether the respondant reports using *any* contraceptive method. Incidentally, long-term contraceptive methods are less *protective* methods against HIV, relative to condom (Anderson, 2018).

Table 8 presents results from estimating Equation 1 on these contraceptive use outcomes. These results reveal that, while not having a significantly different contraceptive behaviour at the extensive margin (i.e. the estimate of $Female \times Matrilineality$ on the probability of using any contraception method is not significant (column 3)), matrilineal women significantly differ from their patrilineal counterparts in the contraceptive methods they use. More specifically, matrilineal women are 3.1 percentage points less likely to report using condom as a current contraceptive method, an effect representing about 22% of the mean probability that patrilineal males report using condom as a contraceptive method (column 1).[71] This very low level of condom use within marriage in Sub-Saharan

[70] See my result on divorce in Table 6.

[71] Note that this mean probability is even lower for females, with only 3.1% (not reported) of females in my final sample reporting

Table 7: **Ancestral Matrilineality and Sexual Behaviour (OLS)**

	All Sample	Formally Married Individuals		
	Sexual activity	Sexual activity	Infidelity	Nb. of extramarital partners
	(1)	(2)	(3)	(4)
Matrilineality	-0.005	-0.017	-0.006	-0.038
	(0.011)	(0.016)	(0.010)	(0.024)
Female	-0.079***	-0.094***	-0.127***	-0.196***
	(0.006)	(0.007)	(0.014)	(0.027)
Female × Matrilineality	0.032*	0.041**	0.026*	0.068**
	(0.017)	(0.020)	(0.015)	(0.027)
Ind. Controls	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes
Observations	181,108	104,483	104,489	104,489
Adj. R-squared	0.310	0.087	0.109	0.022
Clusters	100	99	99	99
Mean Dep. Var. (Patri. Males)	0.537	0.768	0.136	0.204

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Sexual activity" is a dummy indicating whether an individual reports having had any sexual activity in the last 4 weeks. "Infidelity" is a dummy indicating whether an individual reports having had any extramarital partner in the last 12 months. "Nb. of extramarital partners" is the number of extramarital partners in the last 12 months reported by an individual. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. Samples in columns 2 to 4 consist of formally married individuals. ** p<0.10, ** p<0.05, *** p<0.01*

Africa is documented in the literature.[72] On the contrary, I also find in column 2 that matrilineal women are 4.7 percentage points more likely than their patrilineal counterparts to report using a long-term contraceptive method, representing about 60% of the mean of patrilineal males.

In sum, it seems that, having a higher status and being more likely to be decision-maker regarding contraception (see subsection 5.1), matrilineal women are more likely to bear the responsability of contraception, and substitute condom with long-term contraceptive methods. Doing so, matrilineal females *incidentally* substitute protective contraceptive methods with less protective ones. This is in line with Islam et al. (2009), who find similar patterns in the context of Bangladesh between matrilineal *Garo* and patrilineal *Bengali* women. Indeed, they find that while matrilineal *Garo* women's use of contraceptive is higher than the national level, condom use among *Garo* women is lower than at national level. On the contrary, *Garo* women's contraceptive methods are highly skewed towards female long-term methods, in comparison to national levels. In line with my findings, they argue that this may be due to the matrilineal nature of the *Garo* society where wives take most of the family planning responsibilities and rely less on their husbands.

condom as a current contraceptive method.

[72] Chimbiri (2007) calls condom an "intruder in marriage".

An other driver of such contraceptive behaviour may be the desired fertility of matrilineal females.[73] Indeed, as reported in Table A.9 in appendix, I find a higher gender discrepancy in desired fertility for the matrilineal group. In other words, I find that matrilineal females are relatively more likely to desire more children than their patrilineal counterparts, while the opposite is true for male individuals. Along this line, I find a similar result regarding actual fertility. These results may be fully rationnalized in my conceptual framework: matrilineal male's biological children will never integrate his lineage in addition of not inheriting from him. On the contrary, as highlighted in Lowes (2018a), by integrating her lineage and allowing her to benefit from support from her brothers, children constitute assets for a matrilineal woman, which may drive her fertility preferences.

All in all, differences in contraceptive behaviours highlighted in this subsection, and in particular matrilineal females' lower propensity to use condom, constitute an additional mechanism that help explaining the highest contemporary rates of HIV found within matrilineal female populations.

Table 8: **Ancestral Matrilineality and Contraception (OLS)**

	Condom	Long-term contraceptive method	Any contraceptive method
	(1)	(2)	(3)
Matrilineality	0.018*	-0.017	0.005
	(0.010)	(0.011)	(0.014)
Female	-0.106***	0.046***	-0.058***
	(0.017)	(0.008)	(0.019)
Female × Matrilineality	-0.031*	0.047***	0.017
	(0.018)	(0.014)	(0.020)
Ind. Controls	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes
Observations	147,175	146,185	147,175
Adj. R-squared	0.155	0.155	0.156
Clusters	84	84	84
Mean Dep. Var. (Patri. Males)	0.138	0.078	0.245

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Condom" is a dummy indicating whether an individual reports "condom (male)" as her current contraceptive method. "Long-term contraceptive method" is a dummy indicating whether an individual reports a long-term contraceptive method (i.e. "pill", "IUD", "injection", "female sterilization", "implants/norplant", or "lactational amenorrhea (LAM)")). "Any contraceptive method" is a dummy indicating whether an individual reports currently using any contraception method. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10, ** p<0.05, *** p<0.01*

[73] Note that I do control for the number of children in my regressions.

Condom Use and Internalized Risk. My previous finding that matrilineal females have a lower use of condom may hide substantial heterogeneity. I investigate in Table A.10 in appendix whether matrilineal individuals adopt different condom use behaviour when they have properly internalized the risk of HIV transmission. More precisely, I first investigate heterogeneity by perception of the condom as a mean of reducing HIV transmission. I find that the negative effect of matrilineality on condom use that I previously found is null for individuals who have a correct belief about the role of condom. Moreover, I find that when they are seropositive, matrilineal individuals are in fact more likely to report using condom as a contraceptive method. Even more strikingly, I find that this heterogeneity is mainly driven by individuals who have ever been tested for HIV before DHS, and who are therefore presumably aware of their serostatus.

All in all, it seems that, when they have correctly internalized the risk of transmitting the virus, matrilineal females take advantage of their higher ability to impose safe sexual practices. However, this result has to be interpreted cautiously as some biases may arise here. For example, individuals who are HIV positive in my sample may systematically differ from HIV negative individuals in unobservables that are not fully accounted for in my regressions (think about unobserved sexual promiscuity, selection into HIV testing or selective mortality). The same limitations arise when I compare individuals who have been tested in the past with those who have never been tested.

Still, the heterogeneities I uncover are in line with Thornton (2008) who provides experimental evidences from Malawi that HIV-positive individuals learning their status are more likely to purchase condoms, while learning HIV-negative status has no effect. Further, these findings are encouraging as they suggest that policies targeting empowered women should raise their awareness about the actual riskiness of promiscuous sexual behaviour in order to induce behavioural changes restraining the spread of HIV.

5.4 Discarded Mechanisms

Acknowledgment of HIV Risks and Access to Condom. To explore whether matrilineal females' adoption of riskier sexual and contraceptive behaviours may result from a lower acknowledgment of HIV risks and access to condom, I create several outcomes from information in DHS on acknowledgment of HIV risks and role of condom in reducing these risks. I create a dummy indicating whether an individual has ever heard about AIDS (*"Heard of AIDS"*); a dummy indicating whether an individual has ever heard about any STI (*"Heard of STI"*); a dummy indicating whether an individual thinks that always using condom during sex reduces chance of getting HIV (*"Think condom reduces HIV"*); and a dummy indicating whether an individual thinks that having only one sexual partner reduces chance of getting HIV (*"Think having one partner reduces HIV"*). Further, I also create outcomes related to access to condom, namely a dummy indicating whether an individual knows a source to get male condoms (*"Know a source to get condom"*); as well as a dummy indicating whether an individual can get herself a male condom (*"Can get condom"*).

Results from regressing Equation 1 on these outcomes are reported in Table A.11. Two main conclusions can be drawn from this analysis: first, I do not find any evidence that matrilineal females suffer from a lower acknowledgment of risks than their patrilineal counterparts (in fact, the vast majority of both matrilineal and patrilineal females acknowledge such risks[74]). Consequently, I can rule out differential acknowledgment of HIV risks as a factor explaining matrilineal females' adoption of more promiscuous sexual behaviour. Second, matrilineal females' lower use of condom cannot be explained by a lack of access to condom. If any, I conversely find that, in line with their higher sexual autonomy, matrilineal women have an overall significantly easier access to condom.

These results, as well as those found in the previous subsection on condom use behaviour suggest that this is not HIV awareness *per se* that induces behavioural changes, but rather the *perception* of the risk of being HIV positive, and transmitting the virus afterwards. This is lin ine with the findings of the literature exploring the influence of HIV risk perception on sexual and contraceptive behaviours (Oster, 2012a, Paula et al., 2014, Delavande and Kohler, 2015), which is detailed in the introduction.

Sexual Debuts. For biological reasons, young women constitute a population at high risk of contracting HIV when exposed to it (Yi et al., 2013). As such, increase in early and/or premarital sexual activity is a driver of the spread of the virus (Oster, 2005, Case and Paxson, 2013). To test whether this could be a driver of matrilineal females' highest rate of HIV, I exploit information in DHS on reported age at first sexual intercourse, as well as age at first marriage. Interestingly and as reported in Table A.12 in appendix, I find that matrilineal females start their sexual life later than their patrilineal counterparts, suggesting that my main result is in fact mitigated by this mechanism. Further, my results on age at first marriage also suggest a relatively lower age gap between matrilineal spouses, which additionally mitigates my main findings on female sexual behaviour and HIV. As a matter of fact, a higher age gap with her husband may constitute a source of dissatisfaction for a married wife, potentially encouraging her to engage into extramarital relationships (Bertocchi and Dimico, 2019). Furthermore, if one assumes that an older husband has a higher likelihood of being HIV positive (because of a longer sexual life) and is consequently more likely to infect his wife, this result may partially explain why, contrary to Anderson (2018), I mainly capture a female's *extramarital* route of infection.

Antiretroviral Therapy (ART). The introduction of the Antiretroviral Therapy (ART) has been a major policy in Africa to fight against the spread of HIV. However, evidences of the efficiency of ART are mixed, since it has two opposite effects on the prevalence of the virus. On one hand, ART is expected to lower the prevalence of HIV by lowering the viral load of a serepositive person and makes her feel healthier, live longer, and be less likely to pass the virus on. Greenwood et al., 2019 call it an *"Equilibrium Effect"*. On the other hand, ART may lead to what Greenwood et al., 2019 call a *"Behavioural Effect"*: since the existence of ART makes life with HIV more tolerable, this may lead healthy-feeling infected people to engage in riskier behaviour (see De Walque et al., 2012 for evidences from Mozambique; see Cohen et al., 2009 and Friedman, 2018 for evidences from Kenya). Moreover, benefiting from a longer life-expectancy, HIV-infected people have more time to pass the virus on. All in all, according to Greenwood et al., 2019, which effect dominates depends on the fraction of people treated.[75]

Unfortunately DHS does not provide data on ART. I am therefore not able to more formally explore ART as a mechanism of my main result on HIV. Using country-level data from UNAIDS on ART coverage in the 18 countries

[74] In my sample, about 98% of matrilineal females and about 94% of patrilineal females have ever heard of HIV.

[75] According to their model, the *"Equilibrium Effect"* dominates when large proportions of the population are treated, while the *"Behavioural Effect"* dominates when only small proportions of the population are treated.

covered in this paper, I find a significant positive association between the proportion of countries' citizens having a matrilineal ancestor and ART coverage (not reported).[76] While illustrative, this finding does not affect my results as time-varying country-level ART coverage is explicitly controlled for in my regressions through the inclusion of (within-country) region-survey (year) fixed effects.

6 An Epidemiological Approach

In this section, I adopt an epidemiological approach to conduct a simulation exercise. The aim of this simulation exercise is not to embrace the full complexity of the world, but rather to provide some insights on how the differences in sexual behaviours found in the previous section translate into different gender-specific HIV rates dynamics, between ancestrally matrilineal and ancestrally patrilineal societies.

6.1 The Model

To illustrate the differential effects of being matrilineal or patrilineal on female and male HIV propagation rates, through the two main routes of infection highlighted in the empirical results, namely extra-couple casual sexual relationships vs. intra-couple long-term sexual relationships, I adopt a compartmental SI ("Susceptible to Infected") epidemic model, that I adapt from Worden and Porco (2017) and where I add heterogeneity by gender.

I consider a closed population composed of female and male individuals (denoted with subscript f and m respectively). Further, each of these gendered populations are composed of a fraction of "susceptible" (i.e. not infected) individuals, and a fraction of "infected" individuals, denoted $s_{f;m}$ and $i_{f;m}$ respectively.

Sexual behaviours are represented by the following parameters: $\mathbb{1}$ is an indicator of a long-term committed relationship; $\alpha_{f;m}$ is the number of extramarital partners; $\mu_{f;m}$ is the number of sexual intercourses with the long-term committed partner; $\tau_{f;m}$ is the number of sexual intercourses with each extramarital partner. Further, the probability that a male is infected per each sexual intercourse with a HIV positive female is $\delta_m = \gamma_m \times \rho_m$, namely the product of the probability that he does not use condom during sexual intercourse with the transmission risk for one-time male unprotected sex (assuming for simplicity that condom is totally efficient against HIV transmission). The probability that a female is infected per each sexual intercourse with a HIV positive male is $\delta_f = \gamma_f \times (\rho_m \times \omega_f)$, with ω_f capturing gender difference in biological susceptibility of contracting the virus when exposed to it.[77] Finally, I assume random mixing, namely that the probability that an individual faces a HIV positive sexual encounter (should it be her long-term committed partner or an extramarital casual partner) is equal to the proportion of infected individuals in the population of the other sex (i.e. i_m for females' probability and i_f for males' probability).

Female propagation rates (λ_f) can thereby be modeled as follows:

[76] UNAIDS data portal can be found using the following link: `https://aidsinfo.unaids.org/`. Data on matrilineality are from Giuliano and Nunn, 2018.

[77] The medical literature highlights that females have a relatively higher susceptibility of contracting HIV when exposed to it, for biological reasons (i.e. larger surface area of mucosal HIV exposure; increased mucosal expression of the HIV co-receptor CCR5; and a greater probability of virus exposure on the rectal mucosa among other factors (Yi et al., 2013)).

$$\lambda_f = \underbrace{\mathbb{1}\mu_f\delta_f i_m}_{\substack{\text{Intra-couple}\\ \text{infection}}} + \underbrace{\alpha_f\tau_f\delta_f i_m}_{\substack{\text{Extra-couple}\\ \text{infection}}} \tag{4}$$

$\Leftrightarrow$

$$\lambda_f = (\underbrace{\mathbb{1}\mu_f}_{\substack{\text{Intra-couple}\\ \text{intercourses}}} + \underbrace{\alpha_f\tau_f}_{\substack{\text{Extra-couple}\\ \text{intercourses}}})\delta_f i_m \tag{5}$$

Symmetrically, male's propagation rate (λ_m) can be modeled as follows:

$$\lambda_m = (\mathbb{1}\mu_m + \alpha_m\tau_m)\delta_m i_f \tag{6}$$

From Equation 5 and Equation 6, gender difference ($\lambda_f - \lambda_f$) in propagation rates can be modeled as follows:

$$\lambda_f - \lambda_m = (\mathbb{1}\mu_f + \alpha_f\tau_f)\delta_f i_m - (\mathbb{1}\mu_m + \alpha_m\tau_m)\delta_m i_f \tag{7}$$

$\Leftrightarrow$

$$\lambda_f - \lambda_m = \underbrace{\mathbb{1}(\mu_f\delta_f i_m - \mu_m\delta_m i_f)}_{\substack{\text{Gender difference}\\ \text{in intra-couple}\\ \text{infection}}} + \underbrace{\alpha_f\tau_f\delta_f i_m}_{\substack{\text{Female}\\ \text{extra-couple}\\ \text{infection}}} - \underbrace{\alpha_m\tau_m\delta_M i_f}_{\substack{\text{Male}\\ \text{extra-couple}\\ \text{infection}}} \tag{8}$$

Interestingly, if one assumes that $\mu_f = \mu_m$, gender difference in intra-couple infection reduces simply to the gender difference in biological susceptibility of contracting the virus when exposed to it, and the difference in the probability that the husband versus the wife is infected.

It is worth noting that, for simplicity, I assume exogenous and time invariant sexual behavioural parameters ($\mathbb{1}; \alpha_{f;m}; \mu_{f;m}; \tau_{f;m}$ and $\lambda_{f;m}$). Further, I assume random mixing. In other words, I assume that the probability that and individual face a HIV positive partner is equal to the proportion of HIV positive individuals in the pool of the opposite sex ($i_{f;m}$). Additionally, I also assume that this probability is the same for a long-term committed partner or an extramarital partner. Moreover, I assume time-invariant contagiousness. Those simplifying assumptions might be relaxed if one is interested in embracing with greater care the full complexity of the world, which is beyond the scope of this simulation exercise.

6.2 The Effect of Matrilineality: Comparative Statics

Females propagation rates (λ_f). From Equation 5, it is straightforward to see that λ_f is increasing in all females' sexual behavioural parameters. Further, according to my conceptual framework, as well as my empirical results on sexual behaviour and contraceptive use, the value of these parameters should be higher for matrilineal females, relative to their patrilineal counterparts, leading to a higher HIV propagation rate for matrilineal female populations. Further, λ_f is also increasing in i_m. Therefore, assuming that i_f and i_m coevolve in the same direction, I expect a self-reinforcing effect for matrilineal females: since matrilineal females are characterized by higher infection rates, more matrilineal males should be contaminated to, leading in turn to even more contaminated matrilineal females. All in all, this leads to the following prediction:

Prediction 1: $\lambda_f^M > \lambda_f^P$,[78] the HIV propagation rate is higher for matrilineal females, relative to their patrilineal counterparts.

Males propagation rates (λ_m). From Equation 6, it is straightforward to see that λ_m is increasing in all males' sexual behavioural parameters. However, since ancestral matrilineality does not yield significant sexual behavioural differences among males populations according to my conceptual framework, as well as my empirical results, the value of these parameters should be the same for both matrilineal and patrilineal males. Nevertheless, λ_m is also increasing in i_f, which should be higher for matrilineal populations (i.e. infection rate is higher for matrilineal females than their patrilineal counterparts). Consequently, relative to patrilineal male populations, I expect the HIV propagation rate of matrilineal male populations to be only weakly higher. This leads to the following prediction:

Prediction 2: $\lambda_m^M \geq \lambda_m^P$, the HIV propagation rate is weakly higher for matrilineal males, relative to their patrilineal counterparts.

Gender differences in propagation rates ($\lambda_f - \lambda_m$). From Equation 8, it is straightforward to see that $\lambda_f - \lambda_m$ is increasing in females extramarital promiscuous sexual behaviour, and decreasing in males extramarital promiscuous sexual behaviour. As discussed earlier, and in line with my conceptual framework as well as my empirical findings, matrilineal females adopt more promiscuous extramarital sexual behaviours, while matrilineal males adopt as promiscuous extramarital sexual behaviours as their patrilineal counterparts. This leads to the following prediction:

Prediction 3: $\lambda_f^M - \lambda_m^M > \lambda_f^P - \lambda_m^P$, the gender difference in HIV propagation rates is higher for matrilineal populations.

6.3 Simulation

Building on my SI compartmental model of infection discussed earlier, I perform a numerical simulation of the gender-specific HIV rates dynamics in matrilineal versus patrilineal contexts. These HIV rates dynamics can be represented by the following system of equations:

$$\begin{cases} s_{f;t+1} &= s_{f;t} - (\mathbb{1}\mu_f + \alpha_f\tau_f)\delta_f i_{m;t} s_{f;t} \\ i_{f;t+1} &= i_{f;t} + (\mathbb{1}\mu_f + \alpha_f\tau_f)\delta_f i_{m;t} s_{f;t} \\ s_{m;t+1} &= s_{m;t} - (\mathbb{1}\mu_m + \alpha_m\tau_m)\delta_m i_{f;t} s_{m;t} \\ i_{m;t+1} &= i_{m;t} + (\mathbb{1}\mu_m + \alpha_m\tau_m)\delta_m i_{f;t} s_{m;t} \end{cases}$$

[78] Superscripts M and P denote Matrilineality and Patrilineality, respectively.

with $s_{f;t}$ and $s_{m;t}$ denoting the proportions of "susceptible" (i.e. not infected) individuals among female and male populations respectively; and $i_{f;t}$ and $i_{m;t}$ denoting the proportions of "infected" individuals among female and male populations respectively.

According to my conceptual framework as well as my empirical results, the highest rates of HIV among matrilineal females are mainly driven by matrilineal females adopting more promiscuous extramarital sexual behaviours. Further, I found that matrilineal females are mainly contaminated through extramarital routes of infection. To illustrate how this leads to higher gender differential in HIV rates among matrilineal populations, I assume a hypothetical closed population (i.e. no births, no deaths, no migration) of 100 females and 100 males who are in a long-term committed sexual relationship (e.g. formal marriage), with 1 infected female (1%) and 1 infected male (1%) at the initial period, and I simulate HIV rates dynamics on a monthly basis, over a 10 years period.

Parameters values. $\mathbb{1}$ is an indicator of a long-term committed sexual relationship (e.g. formal marriage). I assign to the parameters α_f and α_m the values I have estimated in column 4 of Table 7. More precisely, I assume no difference between matrilineal and patrilineal males and assign $\alpha_m^M = \alpha_m^P = \alpha_m = 0.204$ for both matrilineal and patrilineal males, based on the mean of patrilineal males in my regression sample; and I assign $\alpha_f^M = 0.076$ for matrilineal females; and $\alpha_f^P = 0.008$ for patrilineal females.[79]

The probability that a male is infected per each sexual intercourse with a HIV positive female is $\delta_m = \gamma_m \times \rho_m$, namely the product of the probability that he does not use condom during sexual intercourse with the transmission risk for one-time male unprotected sex (assuming for simplicity that condom is totally efficient against HIV transmission). Therefore, the value of γ_m is simply 1 minus the probability that a male uses condom. Consequently, based on the mean probability that a patrilineal male reports using condom as a contraceptive method in my regression sample of column 1 of Table 8, $\gamma_m = 1 - 0.138 = 0.862$. Again, I assign the same value for both matrilineal and patrilineal males. ρ_m is taken from Greenwood et al. (2019) and is set equal to 0.0045.[80] As a result, $\delta_m = 0.862 \times 0.0045 = 0.003879$.

The probability that a female is infected per each sexual intercourse with a HIV positive male is $\delta_f = \gamma_f \times (\rho_m \times \omega_f)$, with ω_f capturing gender difference in biological susceptibility of contracting the virus when exposed to it. Following my estimates in column 1 of Table 8, $\gamma_f^M = 1 - (0.138 - 0.106 - 0.031) = 0.999$, and $\gamma_f^P = 1 - (0.138 - 0.106) = 0.968$, for matrilineal and patrilineal females respectively. Further, building on several medical studies, Greenwood et al. (2019) assume that women are 75% more likely to get infected than males for physiological and anatomical reasons. I follow them and assign $\omega_f = 1.75$. Therefore, $\delta_f^M = 0.999 \times 0.0045 \times 1.75 \approx 0.007867$ for matrilineal females; and $\delta_f^P = 0.968 \times 0.0045 \times 1.75 = 0.007623$ for patrilineal females.

Unfortunately, the DHS does not directly provide data on the frequency of sexual intercourses within couples. Therefore, I perform the simulation exercise for different values of μ_m, τ_m, μ_f, and τ_f. For consistency purpose, I assume that $\mu_m = \mu_f$ in both matrilineal and patrilineal context. In other words, wife and husband have the same number of intra-couple sexual intercourses per time period. Further, for simplicity I assume that in matrilineal

[79] As for the other females parameter values based on my estimations, I assign values as follows: mean of patrilineal males + "Female" estimate for patrilineal females; mean of patrilineal males + "Female" estimate + "Female x Matrilineal" estimate for matrilineal females. Here, it gives 0.204-0.196=0.008 for patrilineal females; and 0.204-0.196+0.068=0.076 for matrilineal females, based on my estimates in column 4 in Table 7.

[80] This number falls in the range of estimates reported by a variety of studies, according to Greenwood et al. (2019).

context $\tau_m = \tau_f$ (i.e. males and females have the same number of sexual intercourses per time period with *each* of their extramarital partner[81]). Finally, given that according to my empirical findings patrilineal females report a lower sexual activity than their matrilineal counterparts, to determine τ_f^P I lower the value of τ_f^M in the same proportion than patrilineal females reported sexual activity. According to my estimate in column 2 of Table 7, patrilineal females are 4.1 percentage points less likely to report any sexual activity in the last month than matrilineal females. Accordingly, I set $\tau_f^P = 0.959 \times \tau_f^M$. As detailed below, I perform the exercise for different values of μ_m and τ_m. The values and definitions of parameters are summarized in Table A.13 in appendix.

Simulation results. The main results of this simulation exercise[82], computed under a scenario where $\mu_m = \tau_m = 4$, are presented numerically in Table 9, and graphically in Figure 4. Several lessons can be drawn. First, at any time period matrilineal female HIV rate is higher than patrilineal female HIV rate, and this gap is further increasing through time (comparing column 1 and 4). This is consistent with the first prediction of my model. Second, the gap in HIV rates between matrilineal and patrilineal males is of weaker magnitude. This is consistent with the second prediction of my model. Finally, in line with these two previous results and as indicated in the last column, at any time period the gender gap in HIV rates is relatively higher in the matrilineal context, and is further increasing through time. This result is consistent with the third prediction of my model.

Table 9: **Simulated HIV Rates**

	Scenario 1 ($\mu_m = \tau_m = 4$)						
	Matrilineality			*Patrilineality*			**Raw**
Months	Females	Males	Difference	Females	Males	Difference	**Diff-in-Diff**
1	0.010	0.010	0.000	0.010	0.010	0.000	**0.000**
25	0.021	0.017	0.004	0.020	0.017	0.003	**0.001**
50	0.039	0.030	0.009	0.036	0.029	0.007	**0.002**
75	0.071	0.054	0.017	0.064	0.050	0.014	**0.003**
100	0.126	0.095	0.031	0.111	0.087	0.027	**0.004**
120	0.193	0.146	0.047	0.167	0.132	0.035	**0.012**

Notes: This table reports the simulated HIV rates, based on the compartmental SI epidemic model detailed in section 6, and assuming that $\mu_m = \tau_m = 4$.

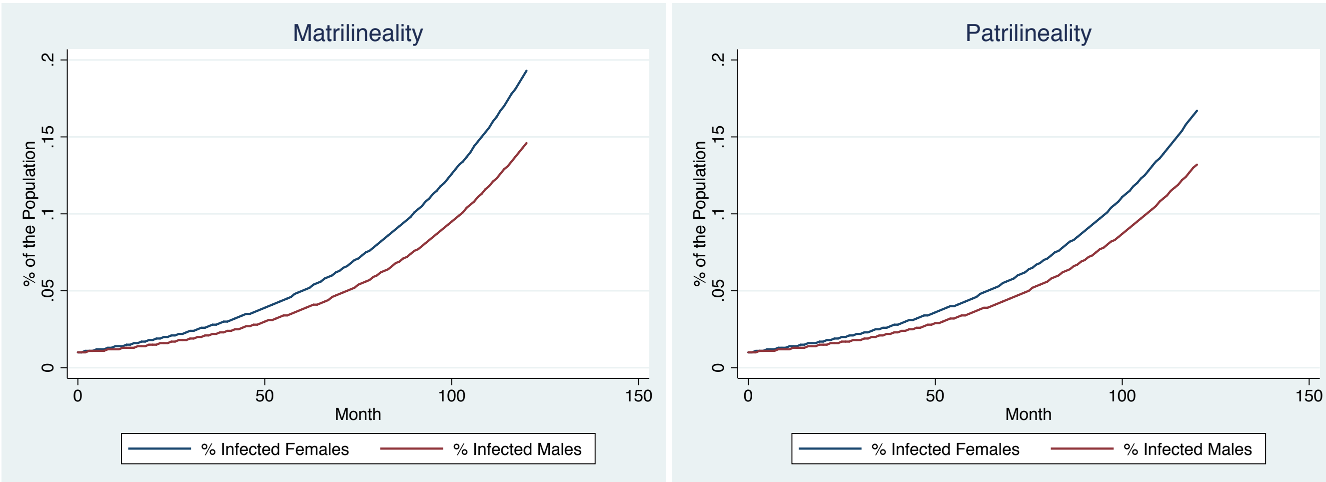
As a robustness check, in Table A.14 as well as Figure A.6 and Figure A.7 in appendix, I perform the same simulation exercise under alternative scenarios, namely $\mu_m = 4$ and $\tau_m = 2$ to allow for the possibility that the frequency of sexual intercourse per extramarital partner is lower than the frequency with long-term committed partner; and $\mu_m = 2$ and $\tau_m = 2$ to emphasize an overall lower frequency of sexual intercourses within both long-term committed and extramarital couples. The conclusions drawn from these alternative exercises are qualitatively similar to my main simulation exercise.

To conclude, this simulation exercise illustrates the different HIV propagation dynamics at play and confirms the main finding of my paper: ancestral matrilineality is, through its impact on females' sexual and contraceptive

[81] This does not mean that females and males have the same number of extramarital partners. These latter are captured by α_f and α_m respectively.

[82] I only report proportions of infected individuals. According to my SI model, the proportion of susceptible individuals is simply the proportion of non-infected individuals (= 1 - proportion of infected individuals).

Figure 4: Simulation - Scenario 1 ($\mu_m = \tau_m = 4$)



behaviour, a driver of contemporary female HIV in Sub-Saharan Africa.

7 Conclusions

In this paper I build on the latest evolutionary psychological theories as well as on the anthropological literature to test the hypothesis that females originating from ancestrally matrilineal ethnic groups are more likely to be infected by HIV today than their patrilineal counterparts. Indeed, ancestral matrilineal kinship organizations constituted environments in which substituting long-term committed sexual relationships with sexual variety may have been more beneficial for females' reproductive success. This strategy would have allowed them to substitute matrilineal males' lower propensity to commit in the long-term with an increased access to better genes for their offsprings. Furthermore, benefiting from better marriage outside option as well as being inherently valued more, matrilineal females are expected to benefit from a greater sexual autonomy and ability to implement their sexual preferences. Consequently, I test the hypothesis that ancestral matrilineality has shaped more promiscuous contemporary females' sexual and contraceptive behaviour, and therefore regretfully drived higher HIV prevalence among these populations.

Using data from 18 countries and exploiting within-country variation in ethnic groups' ancestral kinship organizations, I find that women originating from ancestrally matrilineal ethnic groups are significantly more likely to be HIV positive today. This result is robust to a large set of cultural, historical, geographical and environmental factors, as well as the inclusion of (within-country) region-survey (year) fixed effects. In addition, I show that this long-lasting "matrilineal effect" on female HIV is not driven by any differential selection into DHS HIV testing, nor by differences in overall women's health status, but is specific to sexually transmitted diseases.

I go one step further by formally testing for omitted bias, computing Altonji et al. (2005) ratios and estimating Oster (2017) bias-adjusted lower bound coefficients. These latter provide very little support for the presence of an omitted bias in my OLS estimates.

Going beyond in identifying causal relationships and correcting for potential reverse causality, I exploit GPS location of DHS villages as well as digitized Murdock's map of ancestral ethnic groups in Africa to compute the distance between DHS villages and nearest ancestral matrilineal ethnic boundary. I use this measure to conduct a geographic regression discontinuity design estimation strategy at ethnic boundaries, as well as an instrumental variable strategy. I show that my main estimates are qualitatively robust across all these specifications. Finally, I estimate the average treatment over treated effect of being located on an ancestrally matrilineal geographic area on DHS villages' proportion of HIV positive females, matching villages with their nearest neighbor in an ancestrally non-matrilineal geographic area based on a large array of geographic observables. I find that villages located in ancestrally matrilineal areas have significantly higher females HIV rates, supporting my main OLS finding.

Consistent with my conceptual framework, I find that the higher HIV rates found for matrilineal females can be explained by these latter benefiting from a higher sexual autonomy and ability to implement their sexual preferences, and therefore adopting more promiscuous sexual behaviours. Along these lines, I find that matrilineal females are mainly infected through extramarital routes of infection.

Further, I highlight differences in contraceptive behaviour as a second main mechanism. More precisely, I find evidence that matrilineal females higher contraception-related decision-making power translates into them substituting condom with long-term contraceptive methods. By doing so, matrilineal women incidentally substitute protective contraceptive methods with less protective ones. Nevertheless, I find evidence that when they have internalized the risk of HIV transmission, matrilineal individuals are then more likely to adopt condom as a contraceptive method. Finally, I discard differences in access to condom as well as in sexual debuts as additional mechanisms.

In the end, I build on the epidemiological literature to perform a numerical simulation exercise, aiming at illustrating how the differences between matrilineal and patrilineal populations in gender-specific sexual and contraceptive behaviours translate into different gender-specific HIV rates dynamics. Under credible parameter values, these simulations show dynamics that are consistent with my conceptual framework and my empirical findings.

While recent research literature and policy recommendations have put the emphasis on the need to empower women to reduce the spread of HIV among female populations in Sub-Saharan Africa, this paper highlights mechanisms which call for complementary policies. As a matter of fact, my results have shown that matrilineal women, despite their higher sexual autonomy and ability to impose safe sexual practices to their husbands, are more likely to adopt risky sexual and contraceptive behaviours. Nevertheless, I have also highlighted that, once beliefs about actual risks of HIV transmission are correctly internalized, matrilineal women take advantage of their higher decision-making power to increase their condom use. This promising result calls for complementary policies targeting empowered women, matrilineal women specifically, and raising their awareness about the actual riskiness of adopting promiscuous sexual behaviours. It is hoped that these policies will induce behavioural changes that will restrain the spread of HIV in Sub-Saharan Africa.

References

Abadie, Alberto and Guido W Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 2006, *74* (1), 235–267.

_ and _ , "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics*, 2011, *29* (1), 1–11.

Aberle, David F, "Matrilineal Descent in Cross-Cultural Perspective," *Matrilineal Kinship*, 1961, pp. 655–727.

Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara, "Violence Against Women: a Cross-Cultural Analysis for Africa," *NBER Working Paper*, 2016.

_ , Paula Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Anderson, Siwan, "Legal Origins and Female HIV," *The American Economic Review*, 2018, *108* (6), 1407–39.

Andrews, Isaiah, "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 2018, *100* (2), 337–348.

Ashraf, Nava, Erica Field, Alessandra Voena, and Roberta Ziparo, "Maternal Mortality Risk and Spousal Differences in Desired Fertilitye," *Working Paper*, 2020.

_ , Natalie Bau, Nathan Nunn, and Alessandra Voena, "Bride Price and Female Education," *Journal of Political Economy*, 2020.

Ashraf, Quamrul and Oded Galor, "Genetic Diversity and the Origins of Cultural Fragmentation," *The American Economic Review: Papers and Proceedings*, 2013, *103* (3), 528–33.

_ and _ , "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic development," *The American Economic Review*, 2013, *103* (1), 1–46.

Baland, Jean-Marie and Roberta Ziparo, "Intra-Household Bargaining in Poor Countries," *Towards Gender Equity in Development*, 2018.

Bargain, Olivier, Jordan Loper, and Roberta Ziparo, "Traditional Norm, Access to Divorce and Women's Empowerment: Evidence from Indonesia," *Working Paper*, 2020.

Becker, Anke, "On the Economic Origins of Constraints on Women's Sexuality," *Working Paper*, 2018.

BenYishay, Ariel, Pauline Grosjean, and Joe Vecci, "The Fish is the Friend of Matriliny: Reef Density and Matrilineal Inheritance," *Journal of Development Economics*, 2017, *127*, 234–249.

Berman, Nicolas, Mathieu Couttenier, and Victoire Girard, "Natural Resources and Ethnic Identity," *Working Paper*, 2019.

Bertocchi, Graziella and Arcangelo Dimico, "The Long-Term Determinants of Female HIV Infection in Africa: The Slave Trade, Polygyny, and Sexual Behavior," *Journal of Development Economics*, 2019, *140*, 90–105.

Besley, Timothy and Marta Reynal-Querol, "The Legacy of Historical Conflict: Evidence from Africa," *American Political Science Review*, 2014, *108* (2), 319–336.

Buss, David M, *The Evolution of Desire - Strategies of human mating*, Basic Books, 2016.

Cagé, Julia and Valeria Rueda, "Sex and the Mission: The Conflicting Effects of Early Christian Investments on the HIV Epidemic in Sub-Saharan Africa," *Journal of Demographic Economics*, Forthcoming.

Case, Anne and Christina Paxson, "HIV Risk and Adolescent Behaviors in Africa," *American Economic Review: Papers & Proceedings*, 2013, *103* (3), 433–38.

Cassidy, Rachel, M Groot Bruinderink, Wendy Janssens, and Karlijn Morsink, "The Power to Protect: Household Bargaining and Female Condom Use," *Working Paper*, 2018.

Chimbiri, Agnes M, "The Condom is an "Intruder" in Marriage: Evidence from Rural Malawi," *Social Science & Medicine*, 2007, *64* (5), 1102–1115.

Cohen, Craig R, Michele Montandon, Adam W Carrico, Stephen Shiboski, Alan Bostrom, Alfredo Obure, Zachary Kwena, Robert C Bailey, Rosemary Nguti, and Elizabeth A Bukusi, "Association of Attitudes and Beliefs Towards Antiretroviral Therapy with HIV-Seroprevalence in the General Population of Kisumu, Kenya," *PloS one*, 2009, *4* (3), e4573.

Cordero-Coma, Julia and Richard Breen, "HIV Prevention and Social Desirability: Husband-wife Discrepancies in Reports of Condom Use," *Journal of Marriage and Family*, 2012, *74* (3), 601–613.

Corno, Lucia and Damien De Walque, "Mines, Migration and HIV/AIDS in Southern Africa," *Journal of African Economies*, 2012, *21* (3), 465–498.

Delavande, Adeline and Hans-Peter Kohler, "HIV/AIDS-Related Expectations and Risky Sexual Behaviour in Malawi," *The Review of Economic Studies*, 2015, *83* (1), 118–164.

Diamond, Jared M, *Why is sex fun?: the evolution of human sexuality*, Basic Books, 1998.

Fernández, Raquel, "Does Culture Matter?," in "Handbook of social economics," Vol. 1, Elsevier, 2011, pp. 481–510.

Fortunato, Laura, "The Evolution of Matrilineal Kinship Organization," *Proceedings of the Royal Society London*, 2012.

Fox, Robin, *Kinship and Marriage: An Anthropological Perspective*, Cambridge: Cambridge University Press, 1934.

Friedman, Willa Helterline, "Antiretroviral Drug Access and Behavior Change," *Journal of Development Economics*, 2018, *135*, 392–411.

Giuliano, Paola, "Gender: An Historical Perspective," *The Oxford Handbook of Women and the Economy*, 2017.

_ **and Nathan Nunn**, "Ancestral Characteristics of Modern Populations," *Economic History of Developing Regions*, 2018, *33* (1), 1–17.

_ **and** _ , "Understanding Cultural Persistence and Change," *Working Paper*, 2019.

González, Felipe and Edward Miguel, "War and Local Collective Action in Sierra Leone: A Comment on the Use of Coefficient Stability Approaches," *Journal of Public Economics*, 2015, *128*, 30–33.

Greenwood, Jeremy, Philipp Kircher, Cezar Santos, and Michèle Tertilt, "An Equilibrium Model of the African HIV/AIDS Epidemic," *Econometrica*, 2019, *87* (4), 1081–1113.

Greiling, Heidi and David M Buss, "Women's Sexual Strategies: The Hidden Dimension of Extra-Pair Mating," *Personality and Individual Differences*, 2000, *28* (5), 929–963.

Holden, Clare Janaki and Ruth Mace, "Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis," *Proceedings of the Royal Society of London B: Biological Sciences*, 2003, *270* (1532), 2425–2433.

_ **, Rebecca Sear, and Ruth Mace**, "Matriliny as Daughter-Biased Investment," *Evolution and Human Behavior*, 2003, *24* (2), 99–112.

Islam, Mohammad Amirul, M Rakibul Islam, and Banya Banowary, "Sex Preference as a Determinant of Contraceptive Use in Matrilineal Societies: a Study on the Garo of Bangladesh," *The European Journal of Contraception & Reproductive Health Care*, 2009, *14* (4), 301–306.

Kibira, Simon PS, Patricia Ndugga, Elizabeth Nansubuga, Andrew Sewannonda, and Betty Kwagala, "Contraceptive Uptake Among Married Women in Uganda: Does Empowerment Matter?," *African Population Studies*, 2014, *28*, 968–975.

Lowes, Sara, "Kinship Systems, Gender Norms, and Household Bargaining: Evidence from the Matrilineal Belt," *Working Paper*, 2018.

_ , "Matrilineal Kinship and Gender Differences in Competition," *Working Paper*, 2018.

_ **and Eduardo Montero**, "The Legacy of Colonial Medicine in Central Africa," *Working Paper*, 2018.

_ **and Nathan Nunn**, "Bride Price and the Wellbeing of Women," *WIDER Working Paper*, 2017.

Michalopoulos, Stelios, Louis Putterman, and David N Weil, "The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa," *Journal of the European Economic Association*, 2019.

Mishra, Vinod, Martin Vaessen, J Boerma, Fred Arnold, Ann Way, Bernard Barrere, Anne Cross, Rathavuth Hong, and Jasbir Sangha, "HIV Testing in National Population-Based Surveys: Experience from the Demographic and Health Surveys," *Bulletin of the World Health Organization*, 2006, *84*, 537–545.

Montalvo, José G and Marta Reynal-Querol, "Ethnic Polarization, Potential Conflict, and Civil Wars," *The American Economic Review*, 2005, *95* (3), 796–816.

Moscona, Jacob, Nathan Nunn, and James A Robinson, "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa," *Econometrica*, 2020.

Nunn, Nathan, "Religious Conversion in Colonial Africa," *The American Economic Review: Papers & Proceedings*, 2010, *100* (2), 147–52.

_ **and Leonard Wantchekon**, "The Slave Trade and the Origins of Mistrust in Africa," *The American Economic Review*, 2011, *101* (7), 3221–52.

Nyqvist, Martina Björkman, Lucia Corno, Damien De Walque, and Jakob Svensson, "Incentivizing Safer Sexual Behavior: Evidence from a Lottery Experiment on HIV Prevention," *American Economic Journal: Applied Economics*, 2018, *10* (3), 287–314.

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

_ , "HIV and Sexual Behavior Change: Why not Africa?," *Journal of Health Economics*, 2012, *31* (1), 35–49.

_ , "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa," *Journal of the European Economic Association*, 2012, *10* (5), 1025–1058.

_ , "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2017, pp. 1–18.

Palamuleni, Martin Enock and Ayo Stephen Adebowale, "Women Empowerment and the Current Use of Long Acting and Permanent Contraceptive: Evidence from 2010 Malawi Demographic and Health Survey," *Malawi Medical Journal*, 2014, *26* (3), 63–70.

Paula, Áureo De, Gil Shapira, and Petra E Todd, "How Beliefs About HIV Status Affect Risky Behaviors: Evidence from Malawi," *Journal of Applied Econometrics*, 2014, *29* (6), 944–964.

Porta, Rafael La, Florencio Lopez de Silanes, and Andrei Shleifer, "The Economic Consequences of Legal Origins," *Journal of Economic Literature*, 2008, *46* (2).

Samari, Goleen, "WomenâĂŹs Empowerment and Short-and Long-Acting Contraceptive Method Use in Egypt," *Culture, health & sexuality*, 2018, *20* (4), 458–473.

Sun, Liyang, "Implementing Valid Two-Step Identification-Robust Confidence Sets for Linear Instrumental-Variables Models," *The Stata Journal*, 2018, *18* (4), 803–825.

Tequame, Miron, "HIV, Risky Behavior and Ethno-Linguistic Heterogeneity," *Journal of Economics and Statistics*, 2012, *232* (6), 606–632.

Teso, Edoardo, "The Long-Term Effect of Demographic Shocks on the Evolution of Gender Roles: Evidence from the Transatlantic Slave Trade," *Journal of the European Economic Association*, 2019.

Thornton, Rebecca L, "The Demand for, and Impact of, Learning HIV Status," *The American Economic Review*, 2008, *98* (5), 1829–63.

UNAIDS, "Miles to Go - Closing Gaps. Breaking Barriers. Righting Injustices.," *Joint United Nations Programme on HIV/AIDS*, 2018.

_ , "Women and Girls and HIV," *Joint United Nations Programme on HIV/AIDS*, 2018.

Walque, Damien De, Harounan Kazianga, and Mead Over, "Antiretroviral Therapy Perceived Efficacy and Risky Sexual Behaviors: Evidence from Mozambique," *Economic Development and Cultural Change*, 2012, *61* (1), 97–126.

Worden, Lee and Travis C Porco, "Products of Compartmental Models in Epidemiology," *Computational and Mathematical Methods in Medicine*, 2017, *2017.*

Yi, Tae Joon, Brett Shannon, Jessica Prodger, Lyle McKinnon, and Rupert Kaul, "Genital Immunology and HIV Susceptibility in Young Women," *American Journal of Reproductive Immunology*, 2013, *69*, 74–79.

Appendices

Appendix A Data Description

A.1 List of the countries included in the analysis

All Demographic Health Surveys (DHS) from Sub-Saharan African countries whith data on both HIV, individual's ethnicity as well as village's GPS information are included in the analysis. Based on this selection criterion, the country surveys used include 32 surveys from 18 countries as follow:

- Burkina-Faso (2003, 2010)
- Cameroon (2004, 2011)
- Chad (2014)
- Congo Democratic Republic (2007, 2013)
- Ethiopia (2005, 2011, 2016)
- Gabon (2012)
- Ghana (2003, 2014)
- Guinea (2005, 2012)
- Ivory Coast (2011)
- Kenya (2003, 2008)
- Liberia (2013)
- Malawi (2004, 2010, 2014)
- Mali (2006, 2012)
- Senegal (2005, 2010)
- Sierra Leone (2008, 2013)
- Togo (2013)
- Uganda (2011)
- Zambia (2007, 2013)

A.2 Description of the main controls

A.2.1 Individual-level Data

- **Marital status:** Categorical variable of the current marital status of the respondent (0 *"Never married"*; 1 *"Married"*; 2 *"Living together"*; 3 *"Widowed"*; 4 *"Divorced"*; and 5 *"Not living together"*). **Source:** DHS (variable v501)

- **Actual polygyny:** Dummy variable indicating whether an individual is in a polygynous union. **Source:** DHS (computed from variable v505)

- **Age:** Individual's Age in completed years. **Source:** DHS (variable v012)

- **Age squared:** Square of the individual's age in completed years. **Source:** DHS (computed from variable v012)

- **Number of children:** Individual's total number of children ever born. **Source:** DHS (variable v201)

- **Urban:** Dummy variable indicating whether an individual resides in an urban location. **Source:** DHS (variable v025)

- **Education:** Individual's education in single years. **Source:** DHS (variable v133)

- **Working:** Dummy variable indicating whether an individual is currently working. **Source:** DHS (variable v714)

- **Wealth:** Categorical variable of wealth index (a composite measure of a household's cumulative living standard) (1 *"Poorest"*; 2 *"Poorer"*; 3 *"Middle"*; 4 *"Richer"*; and 5 *"Richest"*). **Source:** DHS (variable v190)

- **Religion:** Categorical variable of religion (1 *"Christian"*; 2 *"Muslim"*; 3 *"Other religion"*; and 4 *"No religion"*). **Source:** DHS (computed and harmonized between countries by the author, from variable v130)

A.2.2 Ethnicity-level Data

These variables are computed from the *Ethnographic Atlas (E.A.)*, a worldwide anthropological database containing ethnographic information on cultural aspects and ways of life of ethnic groups prior to industrilization and colonial contact.

- **Women's historical participation in agriculture:** Variable increasing in women's historical participation in agriculture, and ranging from 1 *"Males only or almost alone"* to 6 *"Females only or almost alone"*. **Source:** Ethnographic Atlas (variable v54)

- **Polygyny:** Dummy variable indicating whether an ethnic group practiced polygyny (v8 = 2; 4 or 5 in the E.A.). **Source:** Ethnographic Atlas (variable v8)

- **Bride Price:** Dummy variable indicating whether an ethnic group practiced bride price (v6 not equal to 6 or 7 in the E.A.). **Source:** Ethnographic Atlas (variable v6)

- **Plough:** Categorical variable of plough use (1 *"No plough"*; 2 *"Not aboriginal but well established plough use at period of observation"*; and 3 *"Aboriginal plough use prior to contact"*). **Source:** Ethnographic Atlas (variable v39)

- **Pastoralism:** Index of ethnic group's dependence on pastoralism, computed following Becker (2018) by interacting the index of ethnic group dependence on animal husbandry (v4 in the E.A.) with a dummy indicating whether herd animals were the predominant ethnic group's type of animal husbandry (i.e v40 = 3, 4, 5, 6 or 7 in the E.A.). **Source:** Ethnographic Atlas (computed from variables v4 and v40)

- **Clans:** Dummy variable indicating whether an ethnic group was organized in clan communities (v15 = 6 in the E.A.). **Source:** Ethnographic Atlas (computed from variable v15)

- **Settlement patterns:** Variable increasing in the complexity of ethnic group's settlement patterns, ranging from 1 *"Nomadic or fully migratory"* to 8 *"Complex settlements"*. **Source:** Ethnographic Atlas (variable v30)

- **Jurisdictional Hierarchies:** Categorical variable of the number of levels in the ethnic group's jurisdictional hierarchy beyond local community (= v33 - 1 in the E.A.). **Source:** Ethnographic Atlas (variable v33)

- **Reliance on hunting:** Categorical variable of ethnic group's reliance on hunting, ranging from 0 *"0 - 5% dependence"* to 9 *"86 - 100% dependence"*. **Source:** Ethnographic Atlas (variable v2)

- **Reliance on fishing:** Categorical variable of ethnic group's reliance on fishing, ranging from 0 *"0 - 5% dependence"* to 9 *"86 - 100% dependence"*. **Source:** Ethnographic Atlas (variable v3)

- **Reliance on gathering:** Categorical variable of ethnic group's reliance on gathering, ranging from 0 *"0 - 5% dependence"* to 9 *"86 - 100% dependence"*. **Source:** Ethnographic Atlas (variable v1)

- **Reliance on animal husbandry:** Categorical variable of ethnic group's reliance on animal husbandry, ranging from 0 *"0 - 5% dependence"* to 9 *"86 - 100% dependence"*. **Source:** Ethnographic Atlas (variable v4)

- **Reliance on agriculture:** Categorical variable of ethnic group's reliance on agriculture, ranging from 0 *"0 - 5% dependence"* to 9 *"86 - 100% dependence"*. **Source:** Ethnographic Atlas (variable v5)

- **Large domesticated animals:** Dummy variable indicating whether an ethnic group practiced animal husbandry (v40 > 1 in the E.A.). **Source:** Ethnographic Atlas (computed from variable v40)

- **Intensity of agriculture:** Categorical variable increasing in ethnic group's intensity of agriculture, and ranging from 1 *"No agriculture"* to 6 *"Intensive irrigated agriculture"*. **Source:** Ethnographic Atlas (variable v28)

- **Year of observation in the E.A.:** Year of observation of the ethnic group in the Ethnographic Atlas. **Source:** Ethnographic Atlas (variavle v102)

A.2.3 Village-level Data

Except when otherwise indicated, most of village-level data are computed from DHS geospatial covariates. Documentation on DHS geospatial covariates can be found here: https://spatialdata.dhsprogram.com/covariates/.

- **Latitude:** Latitude of DHS village (in decimal degrees). **Source:** DHS Geospatial Covariates
- **Longitude:** Longitude of DHS village (in decimal degrees). **Source:** DHS Geospatial Covariates
- **Altitude:** Altitude of DHS village (in meters). **Source:** DHS Geospatial Covariates
- **Nightlight composite:** The average nighttime luminosity, of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village location (2015). **Source:** DHS Geospatial Covariates
- **Population density (2010):** The average UN-adjusted population density of the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village location (2010). **Source:** DHS Geospatial Covariates
- **Distance to lake or coastline:** The geodesic distance to either a lake or the coastline. **Source:** DHS Geospatial Covariates
- **Distance to international border:** The geodesic distance to the nearest international borders. **Source:** DHS Geospatial Covariates
- **Average time urban center: (2015):** The average time (minutes) required to reach a high-density urban center, from the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village location, based on year 2015 infrastructure data. **Source:** DHS Geospatial Covariates
- **Malaria incidence (2010):** The average number of people per year who show clinical symptoms of plasmodium falciparum malaria within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village location (2010). **Source:** DHS Geospatial Covariates
- **Vegetation index (2010):** The average vegetation index value within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village (2010). **Source:** DHS Geospatial Covariates
- **Length of the growing season:** Categorical variable of the length of the growing season in days for the area within the 2 km (urban) or 10 km (rural) buffer surrounding the DHS village location; ranging from 1 *"0 days"* to 16 *" $>$ 365 days"*. **Source:** DHS Geospatial Covariates
- **Distance to nearest active mine:** Distance (in km) to the nearest active mine (active the same year than the village is surveyed in DHS). **Source:** Calculation from the author, exploiting data on the location of large-scale active mines in Africa, provided by S&P Global Market Intelligence (https://www.spglobal.com/marketintelligence/en/campaigns/metals-mining).
- **Active mine within 1000 km:** Dummy indicating whether the distance to the nearest active mine is lower or equal to 1000 km. **Source:** Calcultation from the author, exploiting data on the location of large-scale active mines in Africa, provided by S&P Global Market Intelligence (https://www.spglobal.com/marketintelligence/en/campaigns/metals-mining).

- **Ethnic fractionalization:** Index of ethnic fractionalization computed at the DHS village level using information on individual's (who reside in the village) ethnicity, following the formula of Montalvo and Reynal-Querol (2005) as follows:

$$Ethnic\_Fractionalization = 1 - \sum_{i=1}^{N} \pi_i^2 = \sum_{i=1}^{N} \pi_i(1 - \pi_i) \quad (9)$$

where π_i is the proportion of people who belong to the ethnic group i, and N is the number of thnic groups. The index of ethnic fractionalization, which is a Herfindhahl index, has a simple interpretation as the probability that two randomly selected individuals from a given geographic area do not belong to the same ethnic group. **Source:** Calculation from the author

- **Ethnic polarization:** Index of ethnic polarization computed at the DHS village level using information on individual's (who reside in the village) ethnicity, following the formula of Montalvo and Reynal-Querol (2005) as follows:

$$Ethnic\_Polarization = 1 - \sum_{i=1}^{N} \left(\frac{1/2 - \pi_i}{1/2} \right)^2 \pi^2 = 4 \sum_{i=1}^{N} \pi_i^2 (1 - \pi_i) \quad (10)$$

The original purpose of this Ethnic Polarization (EP) index is to capture how far the distribution of the ethnic groups is from the (1/2, 0, 0, ... 0, 1/2) distribution (bipolar), which represents the highest level of polarization. Ranging between 0 and 1, a higher value of the EP index indicates a higher ethnical polarization, with EP equal to 0 indicating an ethnical homogeneity, and an EP equal to 1 for two ethnic groups of the same size. **Source:** Calculation from the author

Appendix B Additional Figures

Figure A.1: Ancestral Ethnic Group Boundaries and Contemporary Gender Differences in HIV Rates

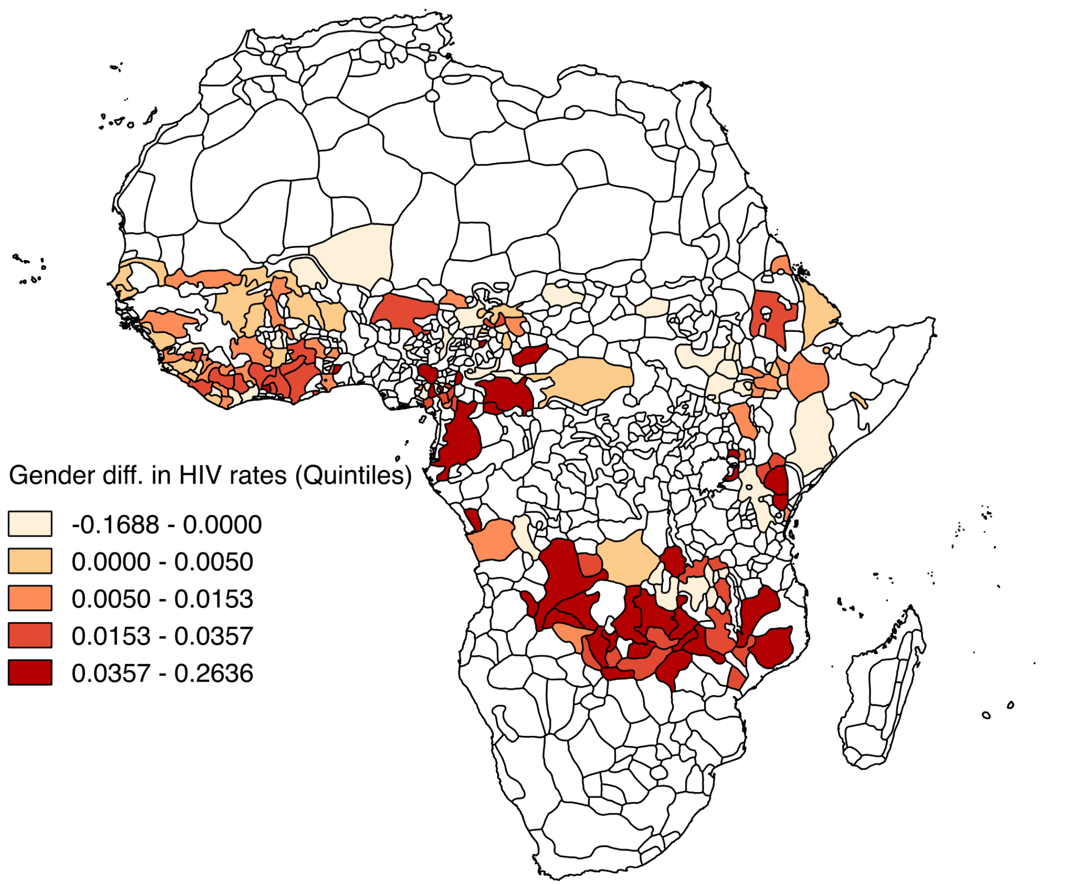


Figure A.2: Ancestral Ethnic Group Boundaries and Contemporary Male HIV Rates

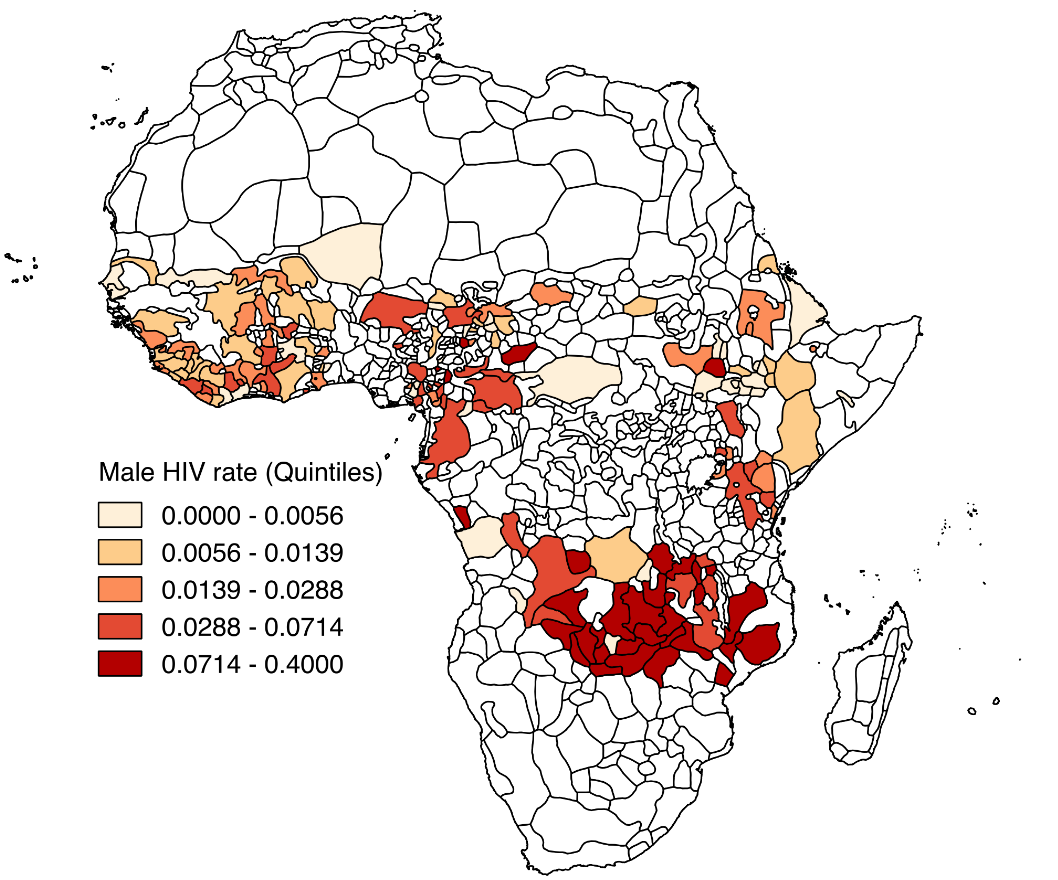
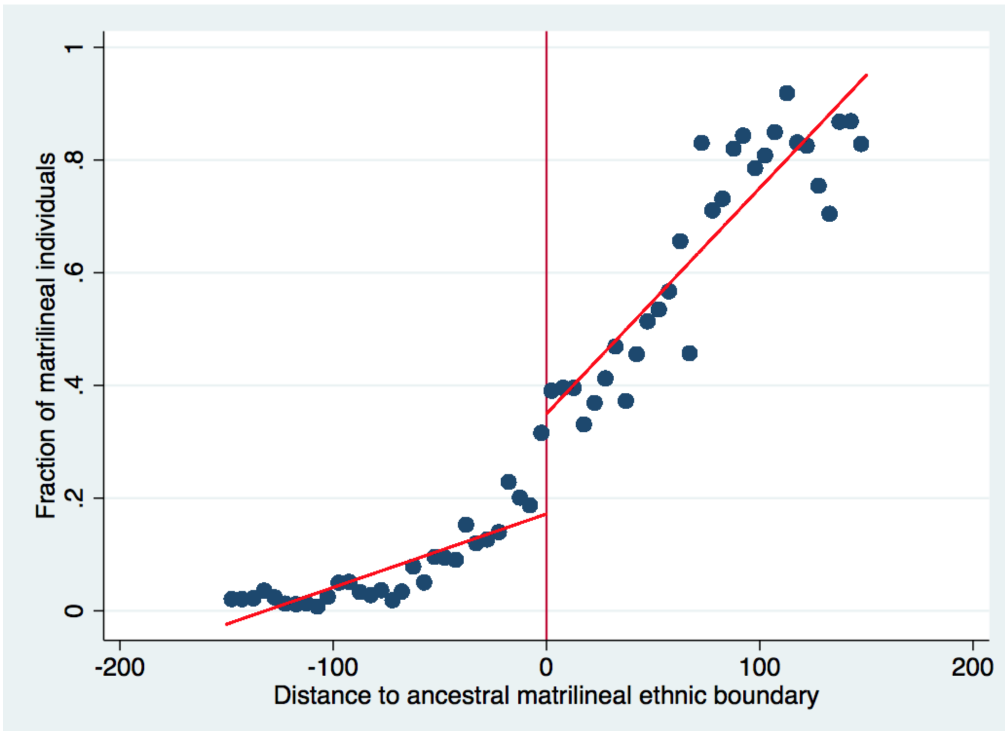


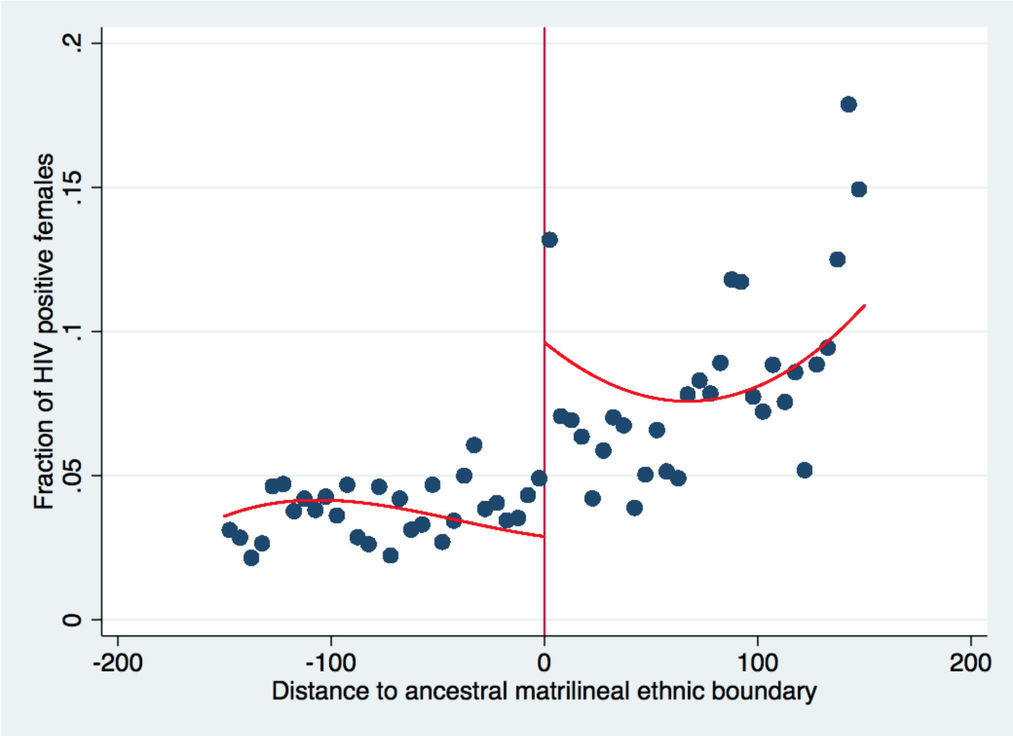
Figure A.3: Location of DHS Villages

Figure A.4: Individual's Matrilineality and Distance to Nearest Ancestral Matrilineal Ethnic Boundary (RDD)



This graph presents the unconditional relationship between individual's ethnic group's ancestral matrilineality and individual's DHS village geographic location, for which a linear polynomial is estimated separately at each side of the boundary. The sample is limited to individuals living in villages located within 150 km of an ancestral matrilineal ethnic boundary. The x-axis reports geographic distance. Positive values are kilometers into the territory of an ancestrally matrilineal ethnic group and negative values are kilometers into the territory of an ancestrally non-matrilineal (i.e. patrilineal or other) ethnic group. The y-axis measures the fraction of the population at each distance that originates from an ancestrally matrilineal ethnic group.

Figure A.5: Female HIV Rate and Distance to Nearest Ancestral Matrilineal Ethnic Boundary (RDD)



This graph presents the relationship between female HIV rate and individual's DHS village geographic location, for a specification that conditions on region (within-country) × survey (time) FE, and for which a cubic polynomial is estimated separately at each side of the boundary. The sample is limited to females living in villages located within 150 km of an ancestral matrilineal ethnic boundary. The x-axis reports geographic distance. Positive values are kilometers into the territory of an ancestrally matrilineal ethnic group and negative values are kilometers into the territory of an ancestrally non-matrilineal (i.e. patrilineal or other) ethnic group. The y-axis measures the fraction of the female population at each distance that is HIV positive.

Figure A.6: Simulation - Scenario 2 ($\mu_m = 4$ and $\tau_m = 2$)

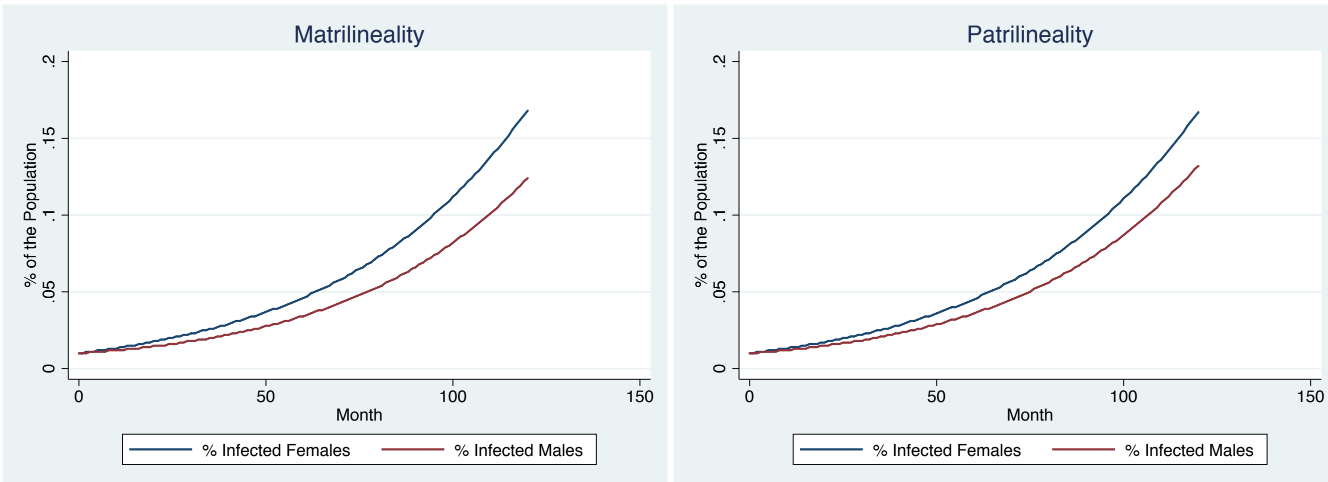
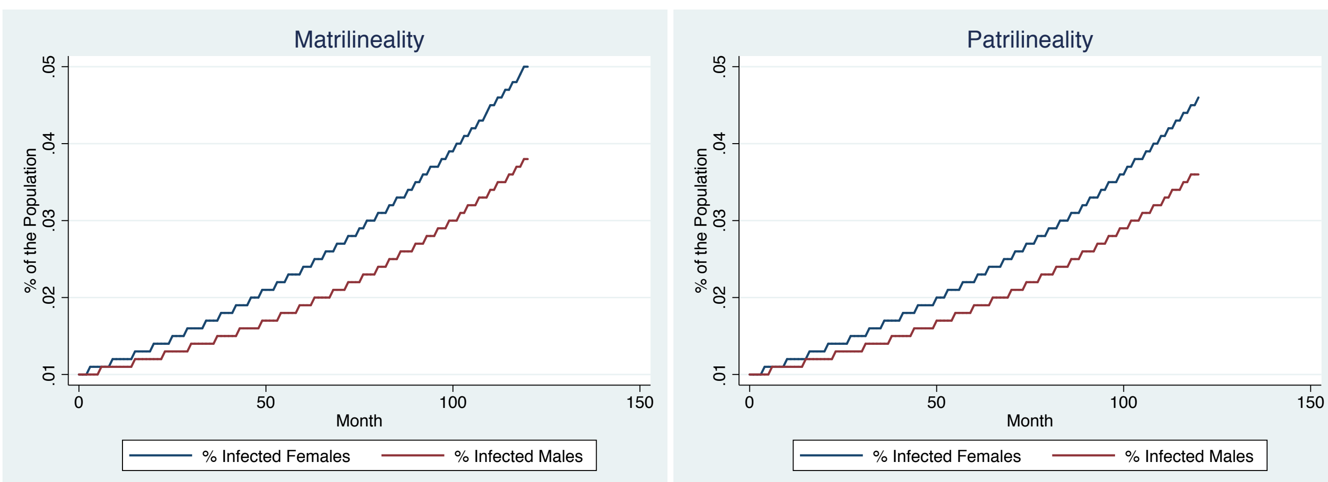


Figure A.7: Simulation - Scenario 3 ($\mu_m = \tau_m = 2$)



Appendix C Additional Tables

Table A.1: **Summary Statistics (Final Regression Sample)**

	Females			Males			All Sample		
	Matrilineal	Patrilineal	Difference	Matrilineal	Patrilineal	Difference	Matrilineal	Patrilineal	Difference
HIV	0.098	0.039	0.059***	0.066	0.023	0.043***	0.083	0.033	0.051***
	(0.194)	(0.298)	(0.002)	(0.150)	(0.249)	(0.002)	(0.277)	(0.177)	(0.001)
Individual-level variables									
Marital status	1.204	1.082	0.122***	0.810	0.786	0.025**	1.021	0.960	0.061***
	(1.228)	(1.037)	(0.009)	(0.934)	(0.888)	(0.008)	(1.119)	(0.989)	(0.006)
Actual polygyny	0.106	0.272	-0.166***	0.055	0.121	-0.066***	0.082	0.210	-0.128***
	(0.308)	(0.445)	(0.004)	(0.227)	(0.326)	(0.003)	(0.275)	(0.407)	(0.002)
Age	28.236	28.937	-0.701***	30.205	31.277	-1.073***	29.151	29.896	-0.745***
	(9.410)	(9.970)	(0.085)	(11.498)	(12.096)	(0.113)	(10.478)	(10.952)	(0.069)
Age squared	885.820	936.766	-50.946***	1044.514	1124.572	-80.058***	959.575	10103.752	-54.177***
	(578.551)	(635.385)	(5.416)	(776.859)	(835.740)	(7.790)	(682.563)	(730.113)	(4.599)
Nb. children	3.062	3.283	-0.221***	2.960	3.155	-0.195***	3.014	3.230	-0.216***
	(2.790)	(3.126)	(0.027)	(3.560)	(4.114)	(0.038)	(3.172)	(3.565)	(0.022)
Urban	0.274	0.310	-0.035***	0.262	0.341	-0.078	0.269	0.322	-0.054***
	(0.446)	(0.462)	(0.004)	(0.474)	(0.440)	(0.004)	(0.003)	(0.001)	(0.003)
Education	5.338	4.185	1.153***	6.419	5.516	0.903***	5.841	4.731	1.110***
	(3.956)	(4.349)	(0.037)	(3.907)	(4.757)	(0.044)	(3.970)	(4.567)	(0.029)
Working	0.575	0.636	-0.061***	0.782	0.788	-0.006	0.671	0.698	-0.027***
	(0.494)	(0.481)	(0.004)	(0.412)	(0.409)	(0.004)	(0.470)	(0.459)	(0.003)
Wealth	3.013	2.997	0.016	3.038	2.993	0.045***	3.025	2.995	0.029**
	(1.416)	(1.430)	(0.012)	(1.385)	(1.424)	(0.013)	(1.402)	(1.428)	(0.009)
Religion	1.237	1.520	-0.283***	1.314	1.615	-0.300***	1.273	1.559	-0.286***
	(0.565)	(0.651)	(0.006)	(0.715)	(0.689)	(0.007)	(0.640)	(0.668)	(0.004)
Ethnicity-level variables									
Women's participation in agri.	4.624	3.758	0.866***	4.647	3.600	1.047***	4.635	3.693	0.942***
	(1.023)	(1.236)	(0.010)	(1.000)	(1.224)	(0.011)	(1.012)	(1.233)	(0.008)
Polygyny	0.770	0.524	0.247***	0.788	0.484	0.304***	0.779	0.508	0.271***
	(0.421)	(0.499)	(0.004)	(0.409)	(0.500)	(0.005)	(0.415)	(0.500)	(0.003)
Bride Price	0.867	1.000	-0.132***	0.880	1.000	-0.119***	0.873	1.000	-0.126***
	(0.339)	(0.022)	(0.001)	(0.325)	(0.022)	(0.001)	(0.332)	(0.022)	(0.001)
Plough	1.000	1.019	-0.019***	1.000	1.000	0.000	1.000	1.011	-0.011***
	(0.000)	(0.137)	(0.001)	(0.000)	(0.022)	(0.000)	(0.000)	(0.106)	(0.001)
Pastoralism	0.087	0.208	-0.121***	0.089	0.195	-0.107***	0.088	0.203	-0.115***
	(0.045)	(0.131)	(0.122)	(0.043)	(0.131)	(0.001)	(0.044)	(0.131)	(0.001)
Clans	0.234	0.262	-0.028***	0.223	0.259	-0.035***	0.229	0.261	-0.032***
	(0.423)	(0.440)	(0.004)	(0.417)	(0.438)	(0.004)	(0.002)	(0.001)	(0.003)
Settlement patterns	6.679	6.628	0.051***	6.699	6.828	-0.129***	6.688	6.710	-0.022***
	(0.610)	(1.129)	(0.009)	(0.599)	(1.044)	(0.009)	(0.605)	(1.100)	(0.007)
Jurisdictional hierarchies	1.470	1.574	-0.103***	1.509	1.525	-0.016*	1.488	1.554	-0.066***
	(0.637)	(0.973)	(0.008)	(0.631)	(0.935)	(0.008)	(0.635)	(0.958)	(0.006)
Reliance on hunting	15.451	8.723	6.729***	15.809	8.877	6.932***	15.618	8.786	6.831***
	(6.505)	(5.432)	(0.048)	(6.585)	(5.604)	(0.055)	(6.545)	(5.504)	(0.036)
Reliance on fishing	14.762	9.229	5.533***	14.629	9.312	5.317***	14.700	9.263	5.437***
	(7.758)	(7.429)	(0.065)	(7.732)	(7.517)	(0.071)	(7.746)	(7.466)	(0.048)
Reliance on gathering	7.198	5.863	1.336***	7.237	6.416	0.821***	7.216	6.089	1.127***
	(4.414)	(4.779)	(0.041)	(4.357)	(4.979)	(0.046)	(4.388)	(4.870)	(0.031)
Reliance on animal husbandry	9.666	20.651	-10.985***	9.720	19.472	-9.752***	9.691	20.168	-10.477***
	(2.655)	(12.684)	(0.101)	(2.524)	(12.571)	(0.108)	(2.595)	(12.651)	(0.074)
Reliance on agriculture	54.339	58.996	-4.657***	53.969	59.235	-5.266***	54.167	59.094	-4.927***
	(5.659)	(9.489)	(0.078)	(5.552)	(10.006)	(0.088)	(5.613)	(9.705)	(0.058)
Large domesticated animals	0.839	0.960	-0.122***	0.856	0.946	-0.090***	0.847	0.954	-0.108***
	(0.368)	(0.195)	(0.002)	(0.351)	(0.227)	(0.002)	(0.360)	(0.209)	(0.002)
Intensity of agriculture	3.016	3.449	-0.432***	3.015	3.348	-0.332***	3.016	3.407	-0.391***
	(0.180)	(0.889)	(0.007)	(0.175)	(0.830)	(0.007)	(0.178)	(0.867)	(0.005)
Year of observation in E.A.	1918.87	1916.973	1.896***	1919.177	1917.302	1.875***	1919.013	1917.108	1.905***
	(11.451)	(18.575)	(0.153)	(11.297)	(15.702)	(0.142)	(11.380)	(17.455)	(0.106)
Village-level variables									
Latitude	-10.973	5.025	-15.998***	-11.046	5.667	-16.714***	-11.007	5.288	-16.295***
	(6.955)	(7.643)	(0.065)	(6.805)	(8.235)	(0.075)	(6.885)	(7.898)	(0.049)
Longitude	27.569	10.705	16.864***	27.709	5.996	21.713***	27.634	8.775	18.860***
	(11.371)	(19.016)	(0.156)	(11.196)	(17.501)	(0.067)	(11.290)	(18.555)	(0.112)
Altitude	856.539	656.037	200.502***	879.954	534.783	345.171***	867.421	606.332	261.089***
	(389.289)	(535.854)	(4.461)	(386.619)	(516.611)	(4.679)	(388.220)	(531.406)	(3.252)
Nightlight composite	2.563	1.601	0.962***	2.448	1.720	0.728***	2.510	1.650	0.860***
	(6.897)	(4.658)	(0.044)	(6.808)	(4.857)	(0.050)	(6.856)	(4.741)	(0.033)
Pop. density (2010)	1297.833	1056.708	241.125***	1050.502	958.689	91.813***	1182.883	1016.528	166.355***
	(5170.735)	(3149.892)	(30.431)	(4312.437)	(3103.504)	(31.634)	(4792.504)	(3131.321)	(21.984)
Distance (geodesic) lake/coastline	103,597	117,080	-13,483***	105,490	131,855	-26,365***	104,477	123,137	-18,660***
	(113,291)	(111,648)	(966)	(109,640)	(112,878)	(1,060)	(111,611)	(112,388)	(715)
Distance (geodesic) international border	46,026	81,903	-35,877***	46,250	81,936	-35,687***	46,130	81,817	-35,787***
	(41,586)	(74,721)	(611)	(40,845)	(79,587)	(700)	(41,243)	(76,752)	(460)
Average time urban center (2015)	77.479	78.707	-1.228***	78.590	86.703	-8.113***	77.995	81.985	-3.989***
	(76.205)	(96.880)	(0.812)	(76.041)	(108.687)	(0.978)	(76.130)	(101.961)	(0.625)
Malaria incidence (2010)	0.379	0.368	0.011***	0.380	0.349	0.031***	0.379	0.360	0.019***
	(0.171)	(0.213)	(0.002)	(0.174)	(0.215)	(0.002)	(0.172)	(0.214)	(0.001)
Vegetation index (2010)	3066.037	3233.712	-167.676***	3077.646	3143.876	-66.230***	3071.432	3196.886	-125.454***
	(582.642)	(975.880)	(8.012)	(590.357)	(987.358)	(8.764)	(586.259)	(981.593)	(5.914)
Length growing season	8.008	9.248	-1.240***	7.967	8.769	-0.802***	7.989	9.052	-1.062***
	(1.511)	(2.733)	(0.022)	(1.481)	(2.601)	(0.023)	(1.497)	(2.690)	(0.016)
Distance mine	3480.909	2399.391	1081.518***	3494.092	2107.588	1386.504***	3487.036	2279.775	1207.261
	(1018.664)	(1094.061)	(9.352)	(993.290)	(1027.745)	(9.641)	(5.869)	(2.755)	(6.785)
Mine within 1000 km	0.034	0.041	-0.007***	0.035	0.054	-0.019***	0.035	0.047	-0.012***
	(0.182)	(0.199)	(0.002)	(0.185)	(0.227)	(0.002)	(0.183)	(0.211)	(0.001)
Ethnic fractionalization	0.370	0.270	-0.100***	0.363	0.270	0.093***	0.367	0.270	0.097***
	(0.280)	(0.265)	(0.002)	(0.281)	(0.265)	(0.003)	(0.280)	(0.265)	(0.002)
Ethnic polarization	0.622	0.486	0.136***	0.611	0.487	0.125***	0.617	0.486	0.131***
	(0.370)	(0.393)	(0.003)	(0.372)	(0.396)	(0.004)	(0.371)	(0.394)	(0.002)
Observations	15,758	90,206		13,683	62,665		29,441	152,871	

Notes: This table reports summary statistics (means; and standard deviations in parenthesis) computed on my main regression sample (column 5 in Table 1). Standard errors of t-test of equality of means are reported in parenthesis in columns reporting mean differences. A description of variables is provided in subsection A.2 of this appendix.

Table A.2: **Ancestral Matrilineality, Genetic Diversity and Father Diversification (OLS)**

	Country-Level		**Individual-Level**	
	Mobility index-predicted genetic diversity	Mobility index-predicted genetic diversity (ancestry andjusted)	Women's children's number of different fathers	Women's children's father diversification
	(1)	(2)	(3)	(4)
Matrilineality	0.005**	0.007*	0.007**	0.014**
	(0.002)	(0.004)	(0.003)	(0.007)
Controls	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	No	No
Region-survey FE	No	No	Yes	Yes
Observations	124	124	66,895	66,895
R-squared	0.994	0.897	0.023	0.333
Clusters	N.A.	N.A.	100	100
Mean Dep. Var.	0.711	0.722	1.019	0.537

Notes: **In columns 1 and 2**, OLS estimates are reported with robust standard errors in brackets. The unit of observation is a country. **"Matrilineality"** is the estimated proportion of countries' citizens with ancestors that had matrilineal inheritence rule (source: Giuliano and Nunn, 2018). This variable ranges from 0 to 1. Outcomes are from Ashraf and Galor (2013b) and are computed at the country level. The **"Mobility index-predicted genetic diversity"** is the expected heterozygosity (genetic diversity) of a given country as predicted by migratory distance (human mobility index), calculated for the journey from Addis Ababa (Ethiopia) to the country's modern capital city. **"Mobility index-predicted genetic diversity (ancestry adjusted)"** is the expected heterozygosity (genetic diversity) of a given country as predicted by migratory distance (human mobility index), calculated for the journey from Addis Ababa (Ethiopia) to each of the year 1500 CE locations of the ancestral populations of the country's component ethnic groups in 2000 CE, as well as for the journey between each pair of these ancestral populations. *The Controls* are from Ashraf and Galor (2013b) and Ashraf and Galor (2013a), computed at the country level and include: the *"aerial"* great circle distance "as the crow flies" from Addis Ababa (Ethiopia) to the country's modern capital city; the square of this previous measure of distance; the *"migratory"* great circle distance from Addis Ababa (Ethiopia) to the country's modern capital city along a land-restricted path forced through one or more of five intercontinental waypoints; the square of this previous measure of distance; the absolute value of the latitude of a country's geodesic centroid; a geospatial index of the suitability of land for agriculture (based on ecological indicators of climate suitability for cultivation); the percentage of a country's arable land; the country's mean distance to nearest waterway; the country's total land area; the average monthly temperature of a country; the average monthly precipitation of a country; the fraction of a country's land area that is located in tropical and subtropical climate zones; the total number of different types of infectious diseases in a country; a dummy indicating whether a nation is an island; and a dummy indicating whether a nation is landlocked. **In column 3 and 4**, OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is a woman surveyed in DHS and being part of my final sample, originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. **"Matrilineality"** indicates (dummy) whether a woman belongs to a traditionally matrilineal ethnic group. **"Women's children's number of different fathers"** is the number of different males with whom a female had her biological children (aged less than 17 and member of the household at the time of the survey). **"Women's children's father diversification"** is the the number of different males with whom a female had her biological children divided by the number of a female's biological children (aged less than 17 and member of the household at the time of the survey). *The controls* consist of the *Individual controls*, the *(Ancestral) Ethnic Group Controls* as well as the *Village-Geographic Controls* defined in subsection 3.3 and Table 1. The number of female's biological children is not controlled for anymore in column 4. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. In columns 3 and 4, "R-squared" reported is adjusted; and "Mean Dep. Var." reported is the mean of patrilineal women in the regression samples. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A.3: **Selection And Falsification Tests**

	Selection		Falsification Tests		
	Consent HIV Test	Take HIV Test	Anemia	BMI	Rohrer Index
	(1)	(2)	(3)	(4)	(5)
Matrilineality	-0.028***	-0.030***	-0.008	-0.105	-0.042
	(0.006)	(0.007)	(0.018)	(0.113)	(0.075)
Female	0.018***	0.020***			
	(0.003)	(0.003)			
Female × Matrilineality	-0.000	0.001			
	(0.005)	(0.006)			
Ind. Controls	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes
Gender	Both	Both	Female	Female	Female
Observations	145,619	145,877	69,999	83,908	83,908
Adj. R-squared	0.053	0.053	0.062	0.197	0.167
Clusters	74	74	63	81	81
Mean Dep. Var. (Patri. Males)	0.941	0.934	0.473	22.195	13.980

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Consent HIV Test" is a dummy indicating whether an individual consented to take the DHS HIV test. "Take HIV Test" is a dummy indicating whether an individual actually took the DHS HIV test. "Anemia" is a dummy indicating whether an individual has any level of anemia (available only for women). "BMI" is the Body Mass Index (available only for women). "Rohrer Index" is available only for women. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. * *p<0.10,** p<0.05,*** p<0.01*

Table A.4: **Heterogeneous Effects by Subsamples: Common Law vs. Civil Law Countries / Polygynous vs. Non Polygynous Individuals**

	HIV			
Sample	Common Law Countries	Civil Law Countries	Ind. not in Polygynous Union	Ind. in Polygynous Union
	(1)	(2)	(3)	(4)
Matrilineality	-0.001	-0.002	-0.000	0.002
	(0.009)	(0.006)	(0.005)	(0.019)
Female	0.009***	0.008***	0.011***	-0.001
	(0.003)	(0.002)	(0.002)	(0.003)
Female × Matrilineality	0.015**	0.002	0.014**	0.007
	(0.008)	(0.007)	(0.006)	(0.015)
Ind. Controls	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes
Observations	81,422	100,890	147,799	34,513
Adj. R-squared	0.089	0.027	0.083	0.060
Clusters	57	57	100	99
Mean Dep. Var. (Patri. Males)	0.043	0.014	0.023	0.024
Mean Female x Matri.	0.158	0.029	0.095	0.049
Std. Dev. Female x Matri.	0.365	0.167	0.294	0.215

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. *Common Law Countries* and *Civil Law Countrie* classification is from La Porta et al. (2008) dataset. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. **"HIV"** is a dummy indicating whether an individual is HIV positive (from DHS HIV Tests). Controls are defined in Table 1. Actual polygyny is not included in the controls in columns 3 and 4. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10,** p<0.05,*** p<0.01*

Table A.5: **Considering Ancestral Matrilocality**

	HIV		
Sample	Full (Controlling for Matrilocality)	Not Matrilocal	Matrilocal
	(1)	(2)	(3)
Matrilineality	-0.003	0.001	
	(0.007)	(0.007)	
Female	0.008***	0.009***	-0.027*
	(0.002)	(0.002)	(0.013)
Female × Matrilineality	0.017***	0.014*	0.048***
	(0.006)	(0.007)	(0.013)
Ind. Controls	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes
Ancestral Matrilocality	Yes		
Observations	182,312	160,079	22,233
Adj. R-squared	0.078	0.067	0.107
Clusters	100	90	10
Mean Dep. Var. (Patrilineal Males)	0.023	0.023	0.021
Mean Female x Matrilineality	0.086	0.031	0.488
Std. Dev. Female x Matrilineality	0.281	0.172	0.500

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. The sample in column 2 is composed of individuals originating from an ethnic group which was not ancestrally matrilocal; while the sample in column 3 is composed of individuals originating from an ethnic group which was ancestrally matrilocal. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group (this is dropped in column 3 for collinearity reason). "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. **"HIV"** is a dummy indicating whether an individual is HIV positive (from DHS HIV Tests). Controls are defined in Table 1. Ancestral matrilocality is additionally controlled for in column 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10,** p<0.05,*** p<0.01*

Table A.6: **Robustness of OLS Estimates to Gender-Specific Alternative Channels**

	HIV			
	(1)	(2)	(3)	(4)
Matrilineality	-0.000	-0.003	0.001	-0.004
	(0.006)	(0.005)	(0.005)	(0.005)
Female	-0.066***	-0.187	0.248	-0.276
	(0.015)	(0.299)	(0.276)	(0.319)
Female × Matrilineality	0.015**	0.016***	0.010*	0.016**
	(0.007)	(0.005)	(0.006)	(0.007)
Ind. Controls	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes
Ind. Controls × Female	Yes	Yes	Yes	Yes
Ethnic Group Controls × Female		Yes	Yes	Yes
Village-Geographic Controls × Female			Yes	Yes
Region-survey × Female FE				Yes
Observations	182,312	182,312	182,312	182,312
Adj. R-squared	0.081	0.081	0.081	0.082
Clusters	100	100	100	100
Mean Dep. Var. (Patri. Males)	0.023	0.023	0.023	0.023

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group (it therefore excludes individuals originating from ethnic groups with alternate inheritance rules (ambilineality, bilinearity, duolinearity, etc.)). **"HIV"**, "Matrilineality", "Female", **"Female × Matrilineality"**, "Ind. Controls", "Ethnic Group Controls", "Village-Geographic Controls", and "Region-survey FE" are defined in Table 1. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A.7: **Assessing the importance of bias from unobservables by controlling for observable characteristics**

		(1)	(2)	(3)	(4)	(5)	(6)
Robustness Tests:		**Coeff. Ratio Test (Altonji et al., 2005)**	**Minimum Coeff. Lower Bound (Oster, 2017)**				$\hat{\beta_F} -/+$ **2.8 S.E.**
Controls in Restricted (R) set	Controls in Full (F) set		$R^2_{Max} = min(\mathbf{1.3}\hat{R^2_F};1)$	$R^2_{Max} = min(\mathbf{1.5}\hat{R^2_F};1)$	$R^2_{Max} = min(\mathbf{2}\hat{R^2_F};1)$	$R^2_{Max} = \mathbf{1}$	
Matrilineality; Female	Full set of controls from Equation 1	15.79	0.017	0.017	0.018	0.031	[0.0004 ; 0.0328]
Matrilineality; Female; Region-survey FE	Full set of controls from Equation 1	-4.89	0.014	0.012	0.008	-0.088	[0.0004 ; 0.0328]
Matrilineality; Female; Region-survey FE; Ind. Controls	Full set of controls from Equation 1	-509.27	0.016	0.016	0.015	-0.001	[0.0004 ; 0.0328]
Matrilineality; Female; Region-survey FE; Ind. Controls; Ethnic Group Controls	Full set of controls from Equation 1	167.42	0.019	0.020	0.024	0.099	[0.0004 ; 0.0328]

Notes: Each cell in column 1 report ratios based on the coefficient of $Matrilineality \times Female$ in two regressions; in one regression a "restricted" set of controls is included and in the other, a "full" set of controls is included. In both regressions, the sample sizes are the same. The controls included in each set are listed on the left side of the table (see Table 1 for a full description of the full set of controls from Equation 1). If $\hat{\beta_R}$ is the coefficient in the restricted set and $\hat{\beta_F}$ is the coefficient in the full set, then the ratio is $\hat{\beta_F}/(\hat{\beta_R} - \hat{\beta_F})$ (see Altonji et al., 2005). Each cell in columns 2-5 report bias-adjusted coefficient lower bounds of $Matrilineality \times Female$ based on Oster (2017). If $\hat{R^2_R}$ is the R^2 of the regression with the restricted set of controls, and $\hat{R^2_F}$ is the R^2 of the regression with the full set of controls, then the minimum coefficient lower bound is: $\hat{\beta_F} - (\hat{\beta_R} - \hat{\beta_F}) \times ((R^2_{Max} - \hat{R^2_F})/(\hat{R^2_F} - \hat{R^2_R}))$. Column 6 reports the bounds of the 99.5% confidence interval of the fully controlled estimate of $Matrilineality \times Female$.

Table A.8: **Ancestral Matrilineality and Couples HIV Discordance (OLS)**

	Dep. Var.: Couple's Serostatus							
	Wife - / Husband -		Wife + / Husband -		Wife - / Husband +		Wife + / Husband +	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Matrilineality	0.0103	0.0107	0.0133*	0.0131*	-0.0018	0.0025	-0.0218***	-0.0263***
	(0.0086)	(0.0096)	(0.0070)	(0.0076)	(0.0066)	(0.0068)	(0.0062)	(0.0058)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Husband's Controls		Yes		Yes		Yes		Yes
Observations	25,272	23,917	25,272	23,917	25,272	23,917	25,272	23,917
Adj. R-squared	0.100	0.106	0.021	0.024	0.027	0.030	0.056	0.061
Clusters	82	82	82	82	82	82	82	82
Mean Dep. Var. (Patri.)	0.955	0.957	0.014	0.013	0.016	0.015	0.015	0.015
Prop. Mixed Ethnicity	0.193	0.156	0.193	0.156	0.193	0.156	0.193	0.156
Prop. Mixed Matrilineality	0.048	0.044	0.048	0.044	0.048	0.044	0.048	0.044

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is a non-polygynous formally married couple with both wife and husband originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" is a dummy indicating whether the wife belongs to a traditionally matrilineal ethnic group. "Wife - / Husband -" is a dummy indicating whether both the wife and the husband are HIV negative. "Wife + / Husband -" is a dummy indicating whether the wife is HIV positive while the husband is HIV negative. "Wife - / Husband +" is a dummy indicating whether the wife is HIV negative while the husband is HIV positive. "Wife + / Husband +" is a dummy indicating whether both the wife and the husband are HIV positive. *Individual controls*, *(Ancestral) Ethnic Group Controls* and *Village-Geographic Controls* are computed for the wife and are defined in Table 1. *Husband's Controls* consist of the *Individual controls*, *(Ancestral) Ethnic Group Controls* and *Village-Geographic Controls* defined in Table 1 and computed for the husband, in addition of a dummy indicating whether the wife and the husband originate from a different Ethnographic Atlas ancestral ethnic group (*"Mixed ethnicity"*), as well as a dummy indicating whether the wife and the husband originate from ethnic groups with different ancestral kinship organizations (matrilineal vs. patrilineal) (*"Mixed matrilineality"*). *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. *Note: * p<0.10,** p<0.05,*** p<0.01*

Table A.9: **Ancestral Matrilineality and Fertility (OLS)**

	Want another child (1)	Nb. children (2)
Matrilineality	-0.070***	-0.129***
	(0.020)	(0.057)
Female	-0.247***	0.188***
	(0.012)	(0.041)
Female × Matrilineality	0.127***	0.153***
	(0.030)	(0.051)
Ind. Controls	Yes	Yes
Ethnic Group Controls	Yes	Yes
Village-Geographic Controls	Yes	Yes
Region-survey FE	Yes	Yes
Observations	120,501	182,312
Adj. R-squared	0.310	0.643
Clusters	84	100
Mean Dep. Var. (Patri. Males)	0.787	3.155

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Want another child" is a dummy indicating whether the individual wants another child. "Nb. children" is the number of biological children ever born. Controls are defined in Table 1 (However, number of children is not controlled for anymore). *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A.10: **Heterogeneity in Condom Use by Perception of Risk of HIV Transmission (OLS)**

	Condom							
	Females				Males			
	Full sample		Ever tested	Never tested	Full sample		Ever tested	Never tested
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Matrilineality × Condom reduces HIV	-0.005				0.009			
	(0.005)				(0.009)			
Matrilineality × Condom not reduces HIV	-0.011**				-0.006			
	(0.005)				(0.009)			
Matrilineality × HIV		0.021***	0.034***	-0.003		0.048***	0.084***	-0.008
		(0.006)	(0.009)	(0.007)		(0.016)	(0.023)	(0.021)
Matrilineality × No HIV		-0.007*	-0.003	-0.005		0.003	-0.008	0.006
		(0.004)	(0.007)	(0.006)		(0.008)	(0.016)	(0.011)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	72,855	89,081	30,795	54,814	52,796	58,094	17,683	39,954
Adj. R-squared	0.073	0.073	0.079	0.067	0.237	0.239	0.240	0.231
Clusters	84	84	82	84	74	74	74	73
Mean Dep. Var. (Patri.)	0.038	0.031	0.055	0.021	0.151	0.138	0.236	0.106
F-Test Equality of coeff. (p-value)	0.310	0.000	0.000	0.241	0.000	0.001	0.000	0.414

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Condom (not) reduces HIV" is a dummy indicating whether an individual thinks that always using condoms during sex (does not) reduces chance of getting HIV. "(No) HIV" is a dummy indicating whether an individual is HIV (negative) positive according to DHS test. "Condom" is a dummy indicating whether an individual reports "condom (male)" as her current contraceptive method. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. Samples "Ever tested" in column 3 and 7 consist of individuals who have ever been tested for AIDS before DHS survey. Samples "Never tested" in column 4 and 8 consist of individuals who have never been tested for AIDS before DHS survey. ** p<0.10,** p<0.05,*** p<0.01*

Table A.11: **Ancestral Matrilineality, Acknowledgment of HIV Risks and Access to Condom (OLS)**

	Acknowledgment of Risks				Access to condom	
	Heard of AIDS	Heard of STI	Think condom reduces HIV	Think having one partner reduces HIV	Know a source to get condom	Can get condom
	(1)	(2)	(3)	(4)	(5)	(6)
Matrilineality	-0.007	-0.006	0.005	-0.002	-0.026	0.007
	(0.005)	(0.004)	(0.009)	(0.007)	(0.017)	(0.017)
Female	-0.027***	-0.025***	-0.048***	-0.031***	-0.191***	-0.286***
	(0.005)	(0.004)	(0.007)	(0.006)	(0.014)	(0.025)
Female × Matrilineality	0.022***	0.021***	0.007	-0.007	0.092***	0.037
	(0.006)	(0.005)	(0.013)	(0.009)	(0.023)	(0.030)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village-Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-survey FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	182,284	182,218	156,133	164,673	164,368	103,316
Adj. R-squared	0.121	0.104	0.048	0.051	0.337	0.161
Clusters	100	100	100	100	100	99
Mean Dep. Var. (Patri. males)	0.966	0.974	0.852	0.913	0.761	0.878

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is a woman originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Heard of AIDS" is a dummy indicating whether an individual has ever heard of AIDS. "Heard of STI" is a dummy indicating whether an individual has ever heard of any STI. "Think condom reduces HIV" is a dummy indicating whether an individual thinks that always using condoms during sex reduces chance of getting HIV. "Think having one partner reduces HIV" is a dummy indicating whether an individual thinks that having only one sexual partner reduces chance of getting HIV. "Know wource to get condom" is a dummy indicating whether an individual knows a source to get male condoms. "Can get condom" is a dummy indicating whether an individual can get herself a male condom. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10,** p<0.05,*** p<0.01*

Table A.12: **Ancestral Matrilineality and Sexual Debuts (OLS)**

	Age first sex (1)	Age first marriage (2)
Matrilineality	-0.783***	-0.739***
	(0.197)	(0.216)
Female	-1.807***	-4.947***
	(0.257)	(0.251)
Female × Matrilineality	1.257***	1.405***
	(0.298)	(0.336)
Ind. Controls	Yes	Yes
Ethnic Group Controls	Yes	Yes
Village-Geographic Controls	Yes	Yes
Region-survey FE	Yes	Yes
Observations	150,197	128,860
Adj. R-squared	0.238	0.444
Clusters	100	100
Mean Dep. Var. (Patri. Males)	18.602	23.996

Notes: OLS estimates are reported with standard errors clustered at the ethnicity level in brackets. The unit of observation is an individual originating from either a traditionally matrilineal or a traditionally patrilineal ethnic group. "Matrilineality" indicates (dummy) whether an individual belongs to a traditionally matrilineal ethnic group. "Female" indicates (dummy) whether an individual is a female. **"Female × Matrilineality"** indicates (dummy) whether an individual is a female belonging to a traditionally matrilineal ethnic group. "Age first sex" is the age at first sexual intercourse. "Age first marriage" is the age at start of first marriage or union. Controls are defined in Table 1. *Region-survey* is a subnational region defined in DHS, interacted with its survey-year. ** p<0.10,** p<0.05,*** p<0.01*

Table A.13: **Summary of Model's Parameters**

Parameter	Value	Interpretation	Source
α_f^M	0.076	*Matrilineal* female's number of extramarital partners	The author (Table 7)
α_f^P	0.008	*Patrilineal* female's number of extramarital partners	The author (Table 7)
α_m	0.204	Male's number of extramarital partners	The author (Table 7)
γ_f^M	0.999	*Matrilineal* female's probability of not using condom	The author (Table 8)
γ_f^P	0.968	*Patrilineal* female's probability of not using condom	The author (Table 8)
γ_m	0.862	Male's probability of not using condom	The author (Table 8)
ρ_m	0.0045	Male's contamination risk for one-time unprotected sex	Greenwood et al. (2019)
ω_f	1.75	Gender difference in biological susceptibility to contract the virus	Greenwood et al. (2019)
δ_f^M	0.007867	*Matri.* female's proba. of infection per each contact with a HIV+ male ($= \gamma_f^M \times \rho_m \times \omega_f$)	The author (Calculation)
δ_f^P	0.007623	*Patri.* female's proba. of infection per each contact with a HIV+ male ($= \gamma_f^P \times \rho_m \times \omega_f$)	The author (Calculation)
δ_m	0.003879	Male's probability of infection per each contact with a HIV+ female ($= \gamma_m \times \rho_m$)	The author (Calculation)
μ_m	4	Male's nb. of sexual intercourses with LT committed partner	The author (Assumption)
μ_f	4	Female's nb. of sexual intercourses with LT committed partner	The author (Assumption)
τ_m	4	Male's nb. of sexual intercourses with each extramarital partner	The author (Assumption)
τ_f^M	4	*Matrilineal* female's nb. of sexual intercourses with each extramarital partner	The author (Assumption)
τ_f^P	3.836	*Patrilineal* female's nb. of sexual intercourses with each extramarital partner ($= 0.959 \times \tau_f^M$)	The author (Assumption)
$i_{f,0}$	0.01	Initial female's HIV rate	The author (Assumption)
$i_{m,0}$	0.01	Initial male's HIV rate	The author (Assumption)

Table A.14: **Simulated HIV Rates (Alternative Scenarios)**

	Scenario 2 ($\mu_m = 4$ and $\tau_m = 2$)						
	Matrilineality			*Patrilineality*			**Raw**
Months	Females	Males	Difference	Females	Males	Difference	**Diff-in-Diff**
1	0.010	0.010	0.000	0.010	0.010	0.000	**0.000**
25	0.020	0.016	0.004	0.019	0.016	0.003	**0.001**
50	0.037	0.028	0.009	0.035	0.027	0.008	**0.001**
75	0.065	0.048	0.017	0.061	0.046	0.015	**0.002**
100	0.112	0.082	0.030	0.103	0.078	0.025	**0.005**
120	0.168	0.124	0.044	0.153	0.116	0.037	**0.007**
	Scenario 3 ($\mu_m = \tau_m = 2$)						
	Matrilineality			*Patrilineality*			**Raw**
Months	Females	Males	Difference	Females	Males	Difference	**Diff-in-Diff**
1	0.010	0.010	0.000	0.010	0.010	0.000	**0.000**
25	0.015	0.013	0.002	0.014	0.013	0.001	**0.001**
50	0.021	0.017	0.004	0.020	0.017	0.003	**0.001**
75	0.029	0.022	0.007	0.027	0.022	0.005	**0.002**
100	0.039	0.030	0.009	0.036	0.029	0.007	**0.002**
120	0.050	0.038	0.012	0.046	0.036	0.010	**0.002**

Notes: This table reports the simulated HIV rates, based on the compartmental SI epidemic model detailed in section 6, assuming $\mu_m = 4$ and $\tau_m = 2$ in the upper part of the table; and $\mu_m = \tau_m = 2$ in the lower part of the table.