Can information campaigns eradicate AIDS? The effect of HIV knowledge and risk behavior on HIV status: the case of three Sub-Saharan countries

Markus Frölich, Rosalia Vazquez-Alvarez[1]

Department of Economics, University of St.Gallen

First version (very preliminary): February, 2007. Comments welcome. Do not circulate

Abstract: AIDS continues to have a devastating effect on many developing economies, particularly in Sub-Saharan Africa. Given the lack of a vaccine to stop HIV transmission and the very expensive medical treatment, most public policy emphasis has been placed on education and particular information campaigns. In this paper, we examine the impact of AIDS education from two sides. First, we examine to what extent information campaigns have been successful in reducing HIV prevalence and incidence. Second, we examine the impact of actual AIDS knowledge on HIV rates. The basic policy issue can be expressed as follows: Suppose that everyone knew and understood the basic facts about AIDS, would this reduce HIV rates to (almost) zero? If so, public policy should target groups with incomplete knowledge. If not, information campaigns alone are bound to fail and much stronger interventions are required to eradicate AIDS. Using rich data sets from three Sub-Saharan economies (Kenya, Tanzania and Ethiopia) we investigate the effect of observed HIV related knowledge on the probability of catching the virus using data on individuals. Our analysis controls for detailed individual specific characteristics including variables reflecting innate risk behaviour that may drive the risk of becoming HIV positive irrespective of HIV related health knowledge. We examine further how these effects differ between different groups, thereby identifying target groups that public information campaigns should direct more attention to. Results so far are preliminary.

Keywords: AIDS/HIV, education, information campaigns, Africa

JEL classification: I10, O15, C21, C14

[1]This paper is part of the project Community Based Development Programme (CoBaDeP). Both authors are affiliated with the Swiss Institute for International Economics and Applied Economic Research (SIAW). The first author is further affiliated with the Institute for Labour Market Policy Evaluation (IFAU), Uppsala, and the Institute for the Study of Labor (IZA), Bonn. We are grateful to the Richard Büchner Stiftung for financial support. markus.froelich@unisg.ch, rosalia.vazquez-alvarez@unisg.ch, University of St. Gallen, Bodanstrasse 8, 9000 St. Gallen, Switzerland, www.siaw.unisg.ch

1 Introduction

According to the latest published figures, at the end of 2006 there were approximately 39 million people world wide living with the Human Immunodeficiency Virus (HIV), the agent that causes Acquired Immunodeficiency Syndrome (AIDS). Although the global HIV-positive incidence rate peaked in the late 1990s, many countries in Africa keep on experiencing significant increases in incidence rates and also prevalence rates. For example, while antenatal clinics surveillance systems and national surveys show a decline in HIV prevalence since the year 2000 in Kenya, Uganda and Zimbabwe, other countries including Mozambique, South Africa and Swaziland are still in the steep section of the prevalence curve. Even when prevalence rates level off, they remain at rather high levels, e.g. at about 6 to 7% in Kenya or Uganda in 2005.[2] Thus, new infections keep the prevalence high.

Without an effective vaccine to stop HIV transmission and so far very expensive medical treatment, information campaigns are considered to be the most (cost) effective public response to reduce incidence, i.e. new infections. Information campaigns aim to deliver knowledge and awareness about AIDS/HIV (the consequence of HIV, transmission routes, preventive methods and available medical treatments) to induce individuals to preventive and prudent behaviour. Yet, information and knowledge alone may be insufficient to eradicate new infections if individuals are not infinetely risk averse.

For public policy purposes it is important to know whether HIV could be eradicated via information campaigns alone or whether additional interventions should be pursued now.

In this paper we assess the further potential of continued information campaings in Ethiopia, Kenya and Uganda by estimating the effect of HIV knowledge on HIV prevalence for those who have *not* yet acquired comprehensive knowledge and awareness about HIV/AIDS. We estimate by how much HIV prevalence could be *reduced* through universal information and how much HIV should we expect to *remain* even then. If the remaining prevalence rate is low, public policy may focus on information alone. If not, further efforts would be required to stop the pandemic.

We further examine how these effects vary between different sociodemographic groups to identify those groups that further information campaigns should target and those groups that

[2] Implying a steady state figure of approximately 7 out of each 100 people in the population between the ages of 15 and 49 living with a HIV-positive status.

are least responsive to information. Specific information campaigns might need to be developed to reach these particular groups.

The empirical analysis is based on individual data from the most recent Demographic and Health Surveys (DHS) for Ethiopia, Kenya and Uganda[3], which included a module for blood testing among their survey instruments.[4] We use possession of and exposure to television and/or radio as instruments for the acquisition of comprehensive HIV knowledge. Conditional on detailed individual data comprising education, general knowledge, general health knowledge, general risk attitudes, investments into health, awareness and responsibility of health care for relatives and children, religion, migration, wealth, household possessions and family structure as well as indicators of pleasure preferences such as smoking and tobacco use and abuse and village characteristics, we believe that these instruments are neither related to the unobservables determining health status nor knowledge status.

To avoid a potential reverse causality, we focus on individuals who are unaware of their own HIV status but nevertheless declare a subjective probability of being HIV positive. Since the actual HIV status is unknown to them, reverse causality between HIV knowledge and HIV status is of less concern. Their subjective probability of being HIV positive may be an indicator of risk but may also drive their stock of knowledge, such that we include this as an important confounding variable.

Although unobserved mechanisms may drive both the actual outcome (either true or subjective HIV status) and the acquired HIV health related knowledge (e.g., motivation, ability), we believe that our data set is sufficiently rich to account for this second source of endogeneity that may also bias our final results. Finally, it is clear that the effect of information campaigns is conditional on the risk attitudes of individuals since HIV infection is to a large extent the consequence of making risky choices. Our estimates allow for effect heterogeneity with regards to innate risk related behaviour that aim at comparing the effect of HIV knowledge on both true and subjective HIV status between sub-populations with clearly differentiated risk attitudes.

This paper is organized as follows. Section 2 reviews the background of the HIV/AIDS

[3]The results are still preliminary in this version of the paper offerening results regarding Kenya in with respect to the estimaiton of the parametric methods as defined in Section 3

[4]The DHS are large scale and representative surveys organized by MEASURE DHS Macro International Inc., which has more than 20 years experience in conducting such surveys worldwide.

pandemic and the state of potential interventions that aim at eradicating the rate of infecting. Section 3 reviews the identification issues that mark the econometric methods employed in this paper. Section 4 reviews the data sets used in the empirical section. Section 5 summarizes findings and results and Section 6 concludes.

2 Background on HIV/AIDS

Recent estimates from UNAIDS, the United Nations Agency for the control and prevention of HIV/AIDS, show that economies in Sub-Saharan Africa account for 63% of all HIV infected cases world wide (25 out of 40 million HIV cases) while 65% of all new infections during 2006 were also within the Sub-Saharan region (2.4 million new cases of HIV infection). For the region itself the figures translate into a 6% HIV prevalence (compared to 1% prevalence worldwide) and a 12% incidence rate (i.e., 2.4 new cases in 2006 relative to the existing HIV population of 24 million over the same period).[5]

Some countries in Sub-Saharan Africa, notable Kenya, Tanzania, Rwanda and Uganda, have experienced a levelling off of HIV prevalence since the end of the 1990s or beginning of 2000. For example, in Kenya and Uganda there has been a steady prevalence rate of 6 to 7% since the late 1990s. It is believed that the leveling off resulted from the initial information campaigns in the middle to late 1980s with major prevention methods mounted in Kenya since the beginning of the year 2000. Surveillance data suggest that these efforts have had positive outcomes (e.g., observed a significant delay in sexual debut, increased use of condoms and decline in the number of partners), although the levelling off is also partly the consequence of mortality from old AIDS cases and saturation of infection among people more at risk (Cheluget et al., 2006). The problem is not identical among sub-groups in the population and neither is the response behaviour. For example, based on surveillance data UNAIDS figures for 2006,[6] prevalence rates in Kenya differ by gender (the HIV prevalence in males is estimated at 5% and 8% for females), between rural and urban areas (10% versus 6%, respectively) and between sub-groups according to infection route: while the nationwide prevalence has levelled off, prevalence among intravenous drug users has risen significantly in the past decade, driving the rates observed in urban areas (e.g., prevalence rates of 50% in Mombasa and 53% in Nairobi). Uganda has

[5] See UNAIDS, 2006 epidemiological updates, www.unaids.org

[6] See UNAIDS country reports, 2005, 2006.

experienced similar trends than Kenya, with rates stabilizing at around 7% since 2004, higher rates for females (8%) than males (5%) and an increase in prevalence in urban areas where intravenous drug users mark the difference with respect to the observed nationwide rates.

Relative to other Sub-Saharan countries (including other Eastern African countries) Uganda is a country where information campaigns were placed in the populations at a relatively early stage of the epidemic in the middle of the 1980s. Kenya has implemented similar campaigns only since the late 1990s. In both countries these campaigns may account for much of the decline in HIV prevalence because of the effect these have had in terms of modifying the risk-related behaviour of individuals. Using surveillance data on a selected number of villages in the south of Uganda, DeWaqe, 2002 finds that information campaigns are successful at reducing the risk of infection although educated individuals, and specifically those with education levels beyond secondary education, are significantly more responsive to HIV/AIDS information campaigns than those with educational achievements below secondary education. Likewise, data from Kenya suggests that the HIV/AIDS information campaigns implemented nationwide since the beginning of the year 2000 have contriubted to modifying the behaviour in the population thus reducing prevalence rates in high risk groups (Baltazar, 2005 and Cheluget et al., 2006). For example, using pre-natal surveillance data, the rate of infection among pregnant women in Kenya has declined from 25% in 1998 to 8% in 2005.

Information campaigns are seen as the most effective tool so far in the aim to eradicate current HIV/AIDS trends. Clearly, the aim of such campaigns is to modify individual's behaviour in situations where contracting the virus becomes more likely, for example, modify individual's perceptions on the use of condoms, number of partners, time of sexual debut and the effect of drug use with regards to transmission rates.Information campaigns also aim at breaking down myths that may increase the risk of discrimination while creating misunderstanding with regards to HIV/AIDS transmission. But information campaigns are costly and divert scarce resources from other policy actions. For example, UNAIDS reports from 2006 suggest that the lowest income Sub-Saharan economies have allocated approximately US$2.5 billion in a decade. These expenditures have aimed at both increasing medical coverage (i.e., aim at providing medical treatment for at least 50% of those who need it) and promoting knowledge transfer that hopes to modify the risk-related behaviour in the population. It is therefore natural to question the effectiveness of such expenditure

in terms of modified health (HIV/AIDS) related behaviour. If information campaigns are working, the result would eventually lead to a decline in HIV prevalence as result of an effective reduction in new cases (incidence) in the population, and not just as the result of mortality among those who develop AIDS. However, the steady state levels observed in Eastern Africa seem to suggest that new incidences are still occurring at inacceptable rates. Employing more resources on information campaigns that target only those who have access to such resources is not an effective policy tool if those with access are not a representative sample in the population. For example, estimates from Kenya's 2003 Demographic and Health Survey (DHS) show that 30% from the lowest wealth deciles have not heard about sexually transmitted infections (STI) while 68% of these have poor or no knowledge in terms of avoiding HIV/AIDS. In contrast only 2% from the top wealth deciles have not heard about STI while at most 13% of these have poor or no knowledge in terms of avoiding HIV/AIDS. Thus, there is a clear and significant difference in terms of health and HIV/AIDS knowledge that may be driven by wealth or other related socioeconomic status. HIV/AIDS comparative analysis between low and high education groups reveals a similar outcome in terms of knowledge distribution. The message, therefore, is that information campaigns (and preventive methods relying on knowledge spread in general) may not be equally distributed among all sociodemographic cells in the population.

Likewise, even in the presence of correctly acquired information by means of information campaigns, individual's innate risk behavior may be the driving force behind the HIV/AIDS problem, a concept that applies to all countries irrespective of their developing status Individual's differ with regards to risk attitudes with some individuals discounting the future such that the potential benefits associated with certain events (e.g., not using condoms, sharing needles) may justify the risks of seroconversion. Cultural and religious attitudes are also factors that may undermine the effect of holding the correct stock of knowledge with regards to HIV/AIDS.

To understand the effect of information campaigns in the population, sound quantitative data on both the distribution of health knowledge and the behavioural attitudes of individuals in the population is needed. We emphasize the word 'distribution', that is, for any quantitative analysis to have relevance in terms of policy impact there is a need to study the effect of information and knowledge campaigns at the microeconomic level where estimation of hetero-

geneous effects are part of the analysis. For example, allowing for aggregates may lead us to conclude that 50% of those between ages 15 to 18 in Kenya have poor knowledge in terms of HIV/AIDS prevention (DHS, Kenya 2003). These numbers hide facts such as the social economic composition of those with poor HIV/AIDS knowledge. At the same time, macroeconomic aggregates do not allow targeting of specific sub-groups who may actually be worst off than the average in terms of preventive information.

Analysis at the microeconomic level that allows for heterogeneity to drive policy evaluation further requires the use of random *representative* data from the population. It is often the case that the data used to analyze the spread of HIV in low income countries (e.g., Sub-Saharan economies) are data from pre-natal clinics where the target group are mainly women. Other data sources are experimental data or quasi-experimental based on surveillance data that may only target a few villages or communities but which do not represent the population as a whole (e.g., deWaqe, 2002), thus inference from such selected groups to the population require caution.

Alternative data sets are available that can help solve the problem and provide unbiased population estimates from the sample analogue. These are national level surveys that cover a random sample of villages thus surveying individuals, households and communities on issues directly relevant to socioeconomic outcomes (e.g., employment, income or consumption) as well as extensive information on HIV knowledge, HIV related risk behaviour and clinical data on HIV testing. Examples of these are the Demographic Health Surveys (DHS) and companion data sets on HIV-testing (AIS data). The DHS and AIS data do not provide panel structure for the dynamic analysis of policy implementation but nevertheless allow for cross-sectional analysis allowing for effects heterogeneity.

The main topic of our empirical analysis is to understand the causal relation between HIV-status (positive or negative) and the stock of knowledge individuals hold with regards to HIV/AIDS (knowledge of transmission routes and treatment). Clearly, in the absence of time varying data it is difficult to determine with certainty the causal relation between HIV/AIDS specific knowledge and the HIV status of the individual, but the data is very rich to control for the potential of reserve causality. Our analysis hopes to provide unbiased and informative estimates on the probability of being HIV positive stratified to different sub-groups in the population by socioeconomic background. Furthermore, we use all the available information

in the data to understand how health knowledge (or lack of it) can explain the HIV status of individuals with estimates that take into account possible endogeneity between HIV status and HIV related knowledge. Our aim is to use these estimates to come up with policy recommendations with regards to the distribution of information campaigns among mutually exclusive sub-groups in the population. Comparing different countries (Kenya, Uganda, Ethiopia and Tanzania) provides comparative analysis that helps to capture country specific effect on the evolution of HIV prevalence and incidence.

3 Econometric methodology

In the first part of the paper, we examine the impact of AIDS knowledge on HIV status. In other words, to which extent does awareness about the dangers and sources of HIV reduce the incidence of HIV. We start by defining notations and assumptions that lead to the econometric framework for our analysis. Let $Y_i \in \{0, 1\}$ denote the binary HIV-status with $Y_i = 1$ representing a HIV-positive status and $Y_i = 0$ a non-HIV-positive status, at some particular point in time. Let D_i denote the knowledge of person i about HIV/AIDS. In some of the analysis, we will treat D as binary in that a person either does have comprehensive knowledge about HIV and transmission channels ($D_i = 1$) or does not ($D_i = 0$).[7] In later analysis, we will also consider a finer measurement of knowledge. Our central interest is to understand the effect of HIV/AIDS related knowledge (D) on HIV status (Y).

The basic conceptual model in mind is as follows: An individual's HIV status essentially depends on *two* sources: prudent *behaviour* and factors not under the direct control of the individual. The latter contains e.g. the disease *environment*. Clearly, if the surrounding HIV prevalence is zero, the risk of infection is zero. It also controls other external influences such as violence or transmission due to blood transfer during surgery, blood donation, infection trough mother-to-child transmission etc. Hence, even the most prudent person faces some risk.

We further suppose that behaviour is driven by two factors: *information* and individual

[7] We pursue different versions in how we treat persons who knows all transmission channels and perhaps additionally believes that there are further transmission channels, e.g. by mosquitoes. The argument could go in both directions. A person knowing the correct channels plus assuming additional ones, may be even more cautious. On the other hand, also a more fatalistic attitude might result weakening caution since mosquitoes are so frequent that their complete avoidance is impossible.

risk preferences. An individual with incomplete knowledge about HIV/AIDS may be unaware that prudent behaviour reduces risk. On the other hand, individuals clearly differ in their risk aversion. We therefore aim to exploit exogenous variation in information while controlling for individual risk preferences.

Since our estimations are based on survey data collected at one point in time, we may consider these two variables to be determined as such:

$$\begin{aligned} Y_i &= \varphi(D_i, U_i) \\ D_i &= \zeta(Y_i, Z_i, V_i) \end{aligned}$$

where φ and ζ are unknown nonparametric functions, U_i and V_i are characteristics of person i, including preferences for risk, pleasure, time preferences and the like. U affects the risk of HIV infection, whereas V represents factors that make individual i more or less likely to have acquired comprehensive knowledge about HIV, i.e. not only having received the information, but also having it fully understood and believing in it. Z_i will be a set of instrumental variables that represent *exogenous* exposure to information. These contain the availability and exposure to radio and television as discussed in more detail below.

For the further discussions it will be helpful to separate the characteristics of person i into observed characteristics X_i and unobserved characteriscs U_i and V_i and denoting the model as:

$$\begin{aligned} Y_i &= \varphi(D_i, X_i, U_i) \\ D_i &= \zeta(Y_i, Z_i, X_i, V_i) \end{aligned} \tag{1}$$

For an individual i with characteristics U_i and V_i, we define the *potential* HIV outcomes as

$$Y_i^d = \varphi(d, X_i, U_i),$$

which is the outcome that would have materialized if an exogenous external intervention had fixed D_i to take the level d without affecting any of the other characteristics of this individual. I.e. Y_1 is the potential outcome (HIV-status) if the individual is fully informed ($D = 1$) and Y_0 is the potential outcome if the individual lacks the correct information on HIV transmission. The effect we are mostly interested in is

$$E\left[Y^1 - Y^0 | D = 0\right] \tag{2}$$

which is the effect on HIV prevalence if those with insufficient HIV knowledge were to be made fully informed. Similarly,

$$E\left[Y^1 - Y^0\right] \tag{3}$$

is the effect of knowledge in the population at large.

In addition to this, we will be interested in how these effects differ by gender and among certain subpopulations, e.g. young adults etc. Let A define a subgroup of the population, e.g. women younger than 20 years, and define the expectation operator:

$$E_A\left[\cdot\right] = E\left[\cdot|X \in A\right].$$

We are thus interested in identifying

$$E_A\left[Y^1 - Y^0|D = 0\right] \tag{4}$$

$$E_A\left[Y^1 - Y^0\right] \tag{5}$$

for several different groups A, to identify those groups that benefit most from information and those that benefit little from it. If certain groups display a particularly strong treatment effect, information campaigns should be developed to target particularly these groups.

Two aspects render the estimation of the (2) difficult. First, U_i and V_i are most likely to be correlated. A person with a different attitude towards education and knowledge may also have different risk preferences. Similarly, unobserved ability may determine both the individual's HIV status and his ability to pick up information.[8]

In addition to that D_i may actually also be a function of Y_i, which may be called *reverse causality*. If a person knows to be HIV positive, he will be more inclined to become informed about the disease, either to prevent transmission to others or in order to reduce the chances of increased morbidity over time. However, even if this individual is not aware of being positive, he may after a while become sick more often and may therefore be more exposed to medical information. In addition, this higher morbidity may even lead to participating in a HIV test with subsequent comprehensive health counselling. Hence, while for some individuals there

[8] As another example, let V be the underlying unobserved HIV prevalence surrounding the individual (e.g., others who know their HIV status constantly provide information to the individual, even if he does not know those providing information are HIV positive) so that V is unobserved random variability determinant of Y.

may be an effect from D to Y, it could be the other way around for others.[9]

To identify the average treatment effect, we pursue different routes.

1) We will estimate (1) by instrumental variable methods. Nonparametric identification is very difficult since Y is binary. If Y was continuous, the approach of Chernohukov et al. (2004, 2005) could be used. For Y discrete, alternative approaches rely on an identification at infinity argument (Tamer, 2003). In other words, they require a large support of the instruments, which we do not have. Therefore we have to rely on parametric or semiparametric identification.

We will use 2SLS estimation as well as *simultaneous equations probit* models, see e.g. Blundell and Smith, 1986, 1989, 1994, and Rivers & Voug, 1998.

2) We ignore reverse causality, i.e. assume that Y does not feature in the function ζ in (1). This opens a large array of (nonparametric) control function type estimators, e.g. Chesher ,2003, 2005, 2006, 2006b, Froelich 2002 and Imbers and Newey, 2003 among others. We expect these results to be *upward biased* if reverse causality really existed, since the reverse relationship would lead from HIV positive to more knowledge.

3) We analyze the relationship in the subpopulation of individuals who *do not know their HIV status*. We define this subpopulation as those individuals who have never taken an HIV test and presume that they do not know their HIV status. For this subpopulation, reverse causality shall be of lesser concern since individuals do not acquire knowledge actively as a result of being HIV positive. This may be most credible for *younger persons* where higher morbidity due to HIV is unlikely to have developed yet. We therefore ignore reverse causality and pursue control function type estimators.

For older individuals, AIDS may have been leading to higher morbidity such that they may have become more exposed to medical information in general, leading to reverse causality and thus to positive bias. On the other hand, by conditioning on the subpopulation of individuals who never have taken a test, we may even obtain negatively biased results, if HIV knowledge increases the probability of test taking. In this case, either comprehensive knowledge or high morbidity increases the probability of test taking, such that, in the absence of any causal effect, these two variables would be negatively correlated among those who did never take a

[9] This implies a relation with feedback effects between the two latent processes that complicates the identification process: only at equilibrium point where the both outcomes are fixed at steady state values would it be possible to identify the weights associated with the regressors.

test. Hence, for the older individuals the overall direction may be unclear.

4) As an alternative to point-identification, we will also examine interval-identification (or set identification) that can be obtained under weaker assumptions. This line of research was pursued most vehemently by Manski, but also received recent attention e.g. in Chesher 2007. These issues will be discussed in more detail in the next section. We start with elaborating the first instrumental variable approach. Since, Y is binary as well as D, we may write the above model as

We may think that two vectors, X and X, Z, contain observed variable information that determinants Y and D, respectively. For simplicity, let both outcomes be determined by latent models such that the following framework applies:

$$\begin{aligned} Y &= I\left(Y^* \geq 0\right) & Y^* &= \varphi(D, X, \beta; U) \\ D &= I\left(D^* \geq 0\right) & D^* &= \zeta(Y, X, Z, \gamma; V), \end{aligned}$$

where we additionally imposed that the functions φ and ζ are parametrically specified by vectors β and γ. We assume further that the instruments Z are independent of the unobservables conditional on X[10]

$$(U, V) \perp\!\!\!\perp Z | X.$$

These will be the crucial instrumental variable conditions, which will be elaborated in more detail in the following Section on data. To foreshadow this discussion, we will consider the availability and exposure to radio and television as Z variables conditional on a huge number of individual and household X variables. These X variables include information about education, general knowledge, health knowledge other than HIV/AIDS knowledge, general risk attitudes, investments into health, awareness and responsibility of health care for relatives and children, religion, migration, wealth, household possessions and family structure as well as indicators of smoking and tobacco use and abuse. These variables capture the inclination towards pleasure and risk, long term versus short term preferences, altruism and general education, health investments and wealth. Since X includes general health knowledge, D only captures very specific AIDS/HIV knowledge. The basic assumption thus is that listening to the radio/television does not affect Y directly and that conditional on these many X variables including wealth

[10] Depending on the precise estimation strategy we may also need that $(U, V) \perp\!\!\!\perp X$.

and education, the posession and exposure to media is not related to other unobserved characteristics. We focus in the first instance on *simultaneous equations probit models.*[11]

3.1 Interval identification

The use of parametric or semiparametric specifications may induce specification bias. Alternatively, we may aim for interval identification using only weak nonparametric assumptions. We will assume that the function φ is weakly *decreasing* in its first argument and weakly *increasing* in its third argument. The first assumption is often called *monotonous treatment assumption* in Manski (1989, 1990, 1997, 2000, 2002). The latter assumption is explored in Chesher 2007.

$$Y_i^d = \varphi(d, X_i, U_i)$$

We pursue to estimate the treatment effect for a subpopulation defined by A.

1) Exclusion restrictions:

The assumption is that, conditional on X:

$$E\left[Y^d|X,Z\right] = E\left[Y^d|X\right],$$

which implies that

$$\max_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=d\right]\cdot P(D=d|X,Z)\right\} \leq E\left[Y^d|X\right]$$
$$\leq \min_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=d\right]\cdot P(D=d|X,Z) + P(D\neq d|X,Z)\right\}$$

such that

$$\max_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=1\right]\cdot p\left(X,Z\right)\right\} - \min_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=0\right]\cdot\left(1-p\left(X,Z\right)\right)+p\left(X,Z\right)\right\}$$
$$\leq E\left[Y^1-Y^0|X\right]$$
$$\leq \min_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=1\right]\cdot p\left(X,Z\right)+\left(1-p\left(X,Z\right)\right)\right)\} - \max_{Z\in supp(Z)} \left\{E\left[Y|X,Z,D=0\right]\cdot\left(1-p\left(X,Z\right)\right)\right\}$$

[11] The coefficient vectors β and γ proceeds by assuming functional forms for both $\varphi()$ and $\zeta()$ together with assumptions on the joint distribution of the error terms (U,V) thus leading to estimates of the parameter vectors (β,γ). Given the binary nature of both outcome and treatment, assuming additive errors such that $(U,V)\sim N\left[0,\Sigma\right]$, where $\Sigma=\left(1,1,\rho_{U,V}\right)$ and allowing for linear specifications to define the functions $\varphi()$ and $\zeta()$ with respect to the covariates (i.e., $\varphi = X\beta+\alpha D+U$; $\zeta = X\gamma_1+Z\gamma_2+V$) a binary probit estimate would lead to the conditional identification of $E\left[Y_1-Y_0|X\right]$ such that $E\left[Y_1-Y_0|X\right]=\Phi(X\beta+\alpha D)-\Phi(X\beta)$, since $E\left[Y_1|X\right]=\Phi(X\beta+\alpha D)$, $E\left[Y_0|X\right]=\Phi(X\beta)$ under the given assumptions.

where $p(x,z) = P(D=1|X=x, Z=z)$.

Therefore we obtain an interval for $E_A\left[Y^1 - Y^0\right]$ as[12]

$$\int_A \left\{ \max_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=1\right] \cdot p\left(X,Z\right) \right\} - \min_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=0\right] \cdot \left(1-p\left(X,Z\right)\right) + p\left(X,Z\right) \right\} \right\} dF$$

$$\leq E_A\left[Y^1 - Y^0\right]$$

$$\leq \int_A \left\{ \min_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=1\right] \cdot p\left(X,Z\right) + \left(1-p\left(X,Z\right)\right) \right\} \right\} - \max_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=0\right] \cdot \left(1-p\left(X,Z\right)\right) \right\} dF$$

where $\int_A dF_X$ refers to the integral with respect to X in the subpopulation A.

2) Exclusion restrictions with monotoniticy

The treatment effect monotonicity implies that

$$E\left[Y^1|X,Z\right] \leq E\left[Y^0|X,Z\right],$$

which together with the exclusion restriction conditional on X

$$E\left[Y^d|X,Z\right] = E\left[Y^d|X\right],$$

gives

$$\max_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=1\right] \cdot p\left(X,Z\right) \right\} \leq E\left[Y^1|X\right] \leq \min_{Z \in supp(Z)} E\left[Y|X,Z\right]$$

and

$$\max_{Z \in supp(Z)} E\left[Y|X,Z\right] \leq E\left[Y^0|X\right] \leq \min_{Z \in supp(Z)} \left\{ p\left(X,Z\right) + E\left[Y|X,Z,D=0\right] \cdot \left(1-p\left(X,Z\right)\right) \right\}$$

such that

$$\int_A \left\{ \max_{Z \in supp(Z)} \left\{ E\left[Y|X,Z,D=1\right] \cdot p\left(X,Z\right) \right\} - \min_{Z \in supp(Z)} \left\{ p\left(X,Z\right) + E\left[Y|X,Z,D=0\right] \cdot \left(1-p\left(X,Z\right)\right) \right\} \right\} dF$$

$$\leq E_A\left[Y^1 - Y^0\right]$$

$$\leq \int_A \left\{ \min_{Z \in supp(Z)} E\left[Y|X,Z\right] - \max_{Z \in supp(Z)} E\left[Y|X,Z\right] \right\} dF_X.$$

[12] This clearly is an extension of the Manski stuff. So we could claim this in the introduction also.

3.1.1 Bounds and HIV-testing Nonresponse

So far we have assumed that individuals outcome variable (HIV-status) is fully observed so that either with parametric assumptions or using nonparametric identification the problem of nonresponse does not play a role. Unfortunately the data is such that a large percentage of individuals refuses to provide blood for the anonymous test (see Section 4 on Data issues). It is well known in the survey literature that refusal to respond on missing information are survey response processes that do not happen at random. In the case of providing blood for anonymous testing, issues such as confidentiality, fear of infection or other relevant behaviour explanations may determine the refusal to participate outcome. Thus, if the sample of size N is representative of the population and HIV-test response is such that only $n_1 = N - n_0$ respond positively to the blood request, we cannot assume that these n_1 individuals are a random sample from N since there must be behavioural reasons that leads the other n_0 not to provide blood, while such behavioural reasons are highly likely to be associated with the outcome variable of interest $Y = HIV - test\ result$. Ignoring the n_0 non-tested may lead to biased estimates of the policy parameter. For example, it may be that those who refuse to be tested are more risk averse and this explains their fears to have a needle injected on them for the sake of a survey. But their risk averse attitude may also determine their risk behavioural attitude in general, including more care and less likeliness to engage in risk related behaviour that may lead towards a HIV-positive status. Under this assumption HIV prevalence in the n_1 group is likely to be much higher than in the n_0 sub-population. Estimates of either (1), (**??**) or (6) will be biased estimates with a bias in the direction of the HIV-prevalence in the sub-population of n_1. Solving the problem implies three options. One option is to assume that refusals do so at random so that using only those individuals for whom we have information provides unbiased estimates of the policy parameters. The second option is to think of on a set of variables Gthat may determine both the response behaviour of individuals as well as determinant of HIV status, and use these variables to impute the missing observations (e.g., as in Rubin, 1987). The approach is widely known and widely used because it provides a complete data set with which to estimate using standard software packages. The problem with such an approach is that it needs underlying underlaying untestable assumptions that may also lead to biased estimates. The third and final option is to bound the actual estimates using the fact that nonresponse itself is a particular treatment. Thus, if we let $\delta = 1$ indicate response and

$\delta = 0$ indicate nonresponse, Bayes' theorem implies the following conditional representation of the outcome variable:

$$\begin{aligned} P(Y_j &= 1|X) = P(Y_j = 1|X, \delta = 1)P(\delta = 1|X) + P(Y_j = 1|X, \delta = 0)P(\delta = 0|X) \quad (6) \\ where\ j &= 0, 1. \end{aligned}$$

Clearly, for either $P(Y_1 = 1|X)$ or $P(Y_0 = 1|X)$ the sampling process is uninformative with regards to $P(Y_j = 1|X, \delta = 0)$, therefore bounds as defined in (**??**), (**??**) or (6) are not identified either since these rely on full response (non-refusal) on the outcome variable Y (HIV-status). As was the case in the previous sub-section, the unknown measure $P(Y_j = 1|X, \delta = 0)$ has natural bounds $(0, 1)$ that may lead to worst case bounds as those defined in (**??**), thus identifying each of the potential outcomes $P(Y_j = 1|X)$ up to a bounding interval. Likewise, we can think of variables Z that are informative on the response (refusal) behaviour of individuals but do not affect individual's HIV-status. Examples of Z are, for example, interviewer's characteristics. Using these exclusion type of restrictions would lead to minimizing and maximizing bounds with respect to discrete values of Z the result of which would be identification regions by means of exclusion restrictions, as in (6). Finally, the use of monotonic conditions is also possible in this set up. We could, for example, follow the example above and assume that those who refuse to be tested are less likely to be infected by the HI-virus, therefore we could allow for bounds on (**??**) derived using this monotonic assumption interpreted as $P(Y_j = 1|X, \delta = 0) \leq P(Y_j = 1|X, \delta = 1)$. Theoretically, allowing for bounding intervals that account for refusals to the HIV-test widens the identification regions implied by (**??**)-(6), but des not change the nature of the bounds. Empirically, the use of bounds on (**??**)-(6) to account for refusals to HIV-testing implies wider identification regions for the policy parameter $E[Y_1 - Y_0|X]$. In the following section we will see that the empirical issue is relevant and this may justify some preliminary results that are based on estimates from (1).

this set up. We could, for example, follow the example above and assume that those who refuse to be tested are less likely to be infected by the HI-virus, therefore we could allow for bounds on (**??**) derived using this monotonic assumption interpreted as $P(Y_j = 1|X, \delta = 0) \leq P(Y_j = 1|X, \delta = 1)$. Theoretically, allowing for bounding intervals that account for refusals to the HIV-test widens the identification regions implied by (4)-(7), but des not change the nature of the bounds. Empirically, the use of bounds on (4)-(7) to account for refusals to HIV-testing implies wider identification regions for the policy parameter $E[Y_1 - Y_0|X]$. In the following section we will see that the empirical issue is relevant and this may justify some preliminary results that are based on estimates from (1).

4 Data (preliminary)

We use data from three Eastern African countries (Ethiopia, Kenya and Tanzania) to estimate and contrast the conditional effect of HIV/AIDS knowledge on HIV-status. Our data sets draw from the original Demographic and Health surveys, DHS-surveys, www.measuredhs.org. The DHS data collection and data distribution is an initiative from USAID (www.usaid.org) that aims at collecting, analysing and diseminating accurate and representative data on socio-demographic aspects (economic, social, population and health outcomes, including extensive informaiton on aspects such as family planning, nutrition and violence indicators) for developing economies. In particular we focuse our attention to DHS data and HIV-test data.

For any given country and any given year the following information applies: DHS-data sets are cross-sections of data collected at particular points in time to cover a representative sample from the population distributed among a finite number of geographical clusters. The DHS data selects a finite number of households to represent the underlying population, and from these, each household is represented by a household member that will answer questions at household level (e.g., household composition, wealth items and migratory movement of household members). From each household the interviewer will select females and males according to some random process so that surveyed individuals are representative of their respective gender groups. Females answer question on their own socio-economic position, schooling, work and health (both subjective, objective health and health knowledge), as well as family planning issues (contraception, fertility, etc), sexual activity and specifically HIV-AIDS knowledge related questions. Females are also asked to provide informaiton on

their children, if they ever had any. Males answer similar socio-economic and health quesitons as females, but males are not asked to provide any information on their children; they are only asked to provide information on thier own family planning preferences. From the randomly selcted female sample, there is a further random selection of females that are asked specific quesitons regarding gender violence towards them or their environment. Furthermore, from the original household members there is a random selection of individuals (both males and females) who are asked to provide blood for the purpose of HIV-testing. The blood test are anonimous and individuals are given full guarantee of hygine, anonymity and respect. Refusal to be tested is accepted by the interviewer. Finally, it is not necessarely the case that the partner of an interviewed female will also be interviewed, thus, the number of couples such that both male and females are interview are a sub-sample from the total possible number of couples in the surveys.

We focuse our attention on the country of Kenya and the year 2003.[14]. The DHS data has information for Kenya for the years 1989, 1993, 1998 and 2003. We choose 2003 because this is the only year that provides HIV-test data from a random sample of surveyed individuals. The Kenya 2003 surveyed sample is reprenterntative of the Kenyan population collecting informaiton from 400 geographical clusters (country sections) that cover a total of 8,651 distinct households. The family structure in Kenya is such that a household may contain more than one family unit, and while distinct family units within a housheold may be blood related this may not necessarily be the case with all surveyed houshold. The 2003 survey registeres household with number of household members ranging from 1 to 24, while. Including household children, the survey collectes information (directly or indirectly) from 37,612 individuals. Each household is assigned a representative household member that answers questions at household level (e.g., wealth, household make up, materials, household composition, etc). Once households are selected (within geographical clusters) the interviewers select randomly from each household to determine the sample of males and females that eventually provide information at individual level. The female sample includes 8,195 sampled individuals whereas the male sample includes

[14] This preliminary version of the paper contains information, estimates and policy conclusions with regards to the country of Kenya. Further results will complete the paper with estimates based on Tanzania and Ethiopia. We select these three economies because the DHSMeasure surveys provides both DHS-data and HIV-test data in all three cases. At the same time, the three economies are sufficiently close geographically to imply some meaningful comparison between them.

3,578 individuals. Both males and females answer questions on their own background (e.g., migration, schooling, employment), reproduction (i.e., descendents information), use and knowledge of contraception and family planning, marriage and sexual activity, fertily preferences, work situation of individual and spouse – if any – and a section on AIDS and other sexually transmitted disease (i.e., on own status knowledge and knwoledge of specific disease by the respondent). Besides this, females also answer a section on pregancy-postnatal care, a section related to their children's health and nutritional backgroud and a section on domestic violence although this is only asked from a random sample selected from the original female random sample. Only 1,430 individual couples within the 8,561 are such that both the male and the female form part of the independently selected male and female samples. Besides the DHS-core data, the same year collected information on HIV-status. This was done by selecting randomly from the 8,561 household to define a sample with a total of 8,800 selected individuals (4,377 are males and 4,423 are females). All those selected for the HIV-test data are ages 15 to 49. If the selected person is a child under age (15-18) the person inmediately responsible for the child is asked for permision to perform the test on the selected child. In any case, all selected members are read a note which specifies and guarantees hygine, anonymity and respect in the event the person refuses to participate. The interviewer emphasise that the results of the test are not given to the person providing the blood. A total of 2,337 individuals from the 8,800 either refuse (1,185), are not present when selected for the test (916) or other problems arise (236) so that testing is only possible for 6,360 of the original 8,800 random sample. We do not necessarely consider the 27% of missing test a random selection from the 73% for whom we have HIV-test data.

Our selection criteria is based on the random sample of individuals that have been asked to provide blood for the purpose of HIV testing. These are the originally randomly selected 8,800. From these we select those that have also answered to the independent females and males DHS surveys so that we have informaiton on them with regards to household characteristics, own characteristics and information on their individual health knowledge referent to HIV/AIDS. The 8,800 individuals are distributed such that 4,377 are male and 4,423 are females. From the 4,377 only 3,578 males asnwer to the DHS core question on socio-economic and demograhic information, whereas 3,421 females selected for HIV testing provide DHS core information. We are interested on understanding the effect that knowledge related to HIV/AIDS has on current

HIV-status. Clearly, if an individual knows his or her status, such knowledge can affect their knowledge-seeking behaviour. Thus, elimination of all those who declare 'ever to have had a formal test for AIDS' implies creating samples of individuals that ignore their own HIV status. The DHS survey asks individuals such question and therefore we can eliminate anyone in the sample that may have actual information on their HIV-positive or negative status.[15]. Thus, from our initial sample sizes (3,578 and 3,421 males and females, respectively), the final sample size upon which to based our estimates (see Section 5, results) consists of 2,993 males and 3,421 females. Table 1 shows the distribution of these with respect to the result of the anonimous HIV-testing :

Table 1: Distribution of selected sample between HIV-Status

	Males (2,993)	**Females (3,421)**
HIV-Negative	2,351	2,569
HIV-Positive	*105*	*217*
Refuse testing	537	635

The outcome of interst is that of 'HIV-Positive'. Table 1 indicates that ignorning those who refuse to be tested, the HIV-prevalence in Kenya (2003) was 4.3% for males and 7.8% for females. These numbers agree with UNAIDS latest estimates (5% and 8%, respectively) and show that the effect of HIV on females is significantly larger (again, ignoring the refusal units, a test of significant difference between the two sample probabilities (3.5%) results in a t-statistic of 5.4, so that such difference is highly significant).

In our aim to understand HIV status, we condition on variables that may explain both the knowledge that individuals have with regards to HIV/AIDS (or health in general) and variables that may explain the outcome HIV-positive. These variables are groupped as household variables, individual specific characteristics, variables that may control for the risk taking behaviur of individuals, and variables that explain the individual's understanding on HIV/AIDS. Table 2 summarizes all these variables according to groups.

[15]Notice that selection of th random sample of individuals who are asked to provide blood is done before they are (later in the survey) asked if they were ever tested for HIV in the past. Therefore, random selection of the 8,800 who are eventually asked to provide blood is independent from the individual's past HIV-testing experience.

Table 2: Covariates sub-groups

Household Characteristics	Individual specific characteristics	Risk behaviour identification	HIV/AIDS Knowledge variables	Instrumental variables
Household size	Age	Smoke	Believes about what causes AIDS	Listen to the radio often
Number of kids age 5 and below	Age of sexual debut	Drink alcohol	Condom knowledge	Watch TV often
Rural/Urban	Education	Household violence	IDU knowledge	Read newspapers and magazines often
Time to nearest water source	Own number of children	Use of nets on children	Look of AIDS person knowledge	Visited by Family Planner in last 12 months
Capital, town, small city, countryside	Use of mosquito net	Vaccination of children in household	AIDS and mosquitoes	
Five Wealth quantiles	Religion		AIDS and sharing food	
Durable items (e.g., car, bike, telephone)	Family planning information		AIDS and traditional healer	
Household ownership	Partner present		AIDS and number of sexual partners	
Land ownership	Employment status		AIDS and blood transfusion	
Building stage of the house	Travelled away last 12 months (males only)		AIDS and pregnacy	
Share toilet facilities	Number of wives and regular partners (males only)			

The variables in the category HIV/AIDS knowledge help us to build up the 'TREATMENT' variable of interest. Recall that our interest is to understand the effect of present (adquired) knowledge on past (adquired but unknown) HIV status. There are multiple variables that would serve as single treatment variables. At this preliminary stage we let the treatment variable DKNOWS be 1 if the person is able to identify all known routes of transmission that may cause a HIV-positive status, if the person answers 'no' when asked if sharing cooking and eating utensils my transfer AIDS and if the person answers 'yes' when asked if a healthy looking person is compatible with a HIV-positive status. Furthermore, individuals are also asked to declare if they know anyone that has ever died of AIDS. These is taken as an independent indicator of 'knowledge' of AIDS because even if it provides information on individual's knowlege of the illness, it does not directly inform us on the effect this information may have on individual's risk taking behaviur. The variable DKNOWS is, therefore, a dummy variable that equals 'one' if the person has sound knowledge that may help him or her avoid becoming HIV positive in the event of confronting a risk episode, and 'zero' otherwise. Table 3 summarises this variable

for each of the gender's subgroups and with regards to their positive and negative HIV status:

Table 4: Relation between HIV status and Health knowledge

	MALES			FEMALES	
	Has sound health knowledge DKNOWS=1	Does not have sound health knowledge DKNOWS=0		Has sound health knowledge DKNOWS=1	Does not have sound health knowledge DKNOWS=0
HIV-NEGATIVE	988	1363		1,419	1,150
HIV-POSITIVE	33	72		83	134
Total	**1,021**	**1,435**		**1,502**	**1,284**

Table 4 already provides information with regards to the totals without accounting for the refusals (these will be dealt with later in the results section). We could think of the summary statistics in Table 4 as an unconditional anticipation to the effect that HIV-knowledge (DKNOWS) has on the outcome HIV-Positive. Clearly, the percentage of individuals who unknown to them are HIV-positive and at the same time have no knowledge or insufficient knowledge about HIV/AIDS is singnificantly larger than those who are HIV-positive but are sufficiently aware about HIV/AIDS. In fact, 8% of those with poor HIV/AIDS knowledge are HIV-positive, whereas 5% of those with sound HIV/AIDS knowledge are HIV-positive. Thus, from this very basic unconditional correlations there seems to be a positive relation between low or poor knowledge and becoming HIV-positive. We use the two samples as given in Table 4 (2,456 males and 2,786 females) to estimate conditional outcomes allowing for the treatment 'HIV/AIDS knowledge'.

5 Results (very preliminary!)

In this section our aim is to present conditional results on the effect of HIV/AIDS knowledge (DKNOWS) on HIV-positive status. The outcome variable (HIV-positive, HIV-negative) is binary in nature and indicates the result of anonimous testing so that the outcome is in fact unknown to the respondent (to all other conditional variables). Assuming complete exogeneity of all variables in Table 3, we proceed to estimate a probit model where the main treatment variable of interest is the bynary outcome DKNOWS=1 if the person has solid knowledge on avoiding HIV, and DKNOWS=0 otherwise. We estimate separately for males and females since there are clear behavioural difference regarding genders with respect to both outcome and treatment. Table 5 summarizes the estimates for the conditional outcome for the sub-group of

females

Table 5: Females estimate, Conditional probability of HIV-positive

Probit regression — Number of obs = 2786; LR chi2(20) = 122.55; Prob > chi2 = 0.0000; Log likelihood = -700.92813; Pseudo R2 = 0.0804

hivp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0097033	.0067341	1.44	0.150	-.0034952	.0229018
edu	-.1487783	.0602557	-2.47	0.014	-.2668773	-.0306793
numkids	.019141	.0236445	0.81	0.418	-.0272013	.0654834
hsize	-.0887317	.0188247	-4.71	0.000	-.1256275	-.051836
nkids5	.0507527	.0470461	1.08	0.281	-.0414561	.1429614
urban	.1200246	.1051315	1.14	0.254	-.0860293	.3260785
twater	.2396629	.0844589	2.84	0.005	.0741265	.4051993
wqtile1	-.3872261	.1498796	-2.58	0.010	-.6809847	-.0934675
wqtile2	-.1199298	.1230944	-0.97	0.330	-.3611903	.1213306
wqtile3	-.1481841	.1234968	-1.20	0.230	-.3902333	.0938651
ownh	-.1333402	.1118422	-1.19	0.233	-.3525468	.0858665
nomsq	-.1655555	.0855805	-1.93	0.053	-.3332903	.0021793
catolic	.0301739	.2862044	0.11	0.916	-.5307764	.5911243
protest	.0756138	.2800369	0.27	0.787	-.4732484	.624476
muslim	-.3924145	.3153661	-1.24	0.213	-1.010521	.2256917
partner	-.1782787	.0831218	-2.14	0.032	-.3411944	-.015363
employed	.0484098	.0801115	0.60	0.546	-.1086058	.2054254
sxage	.0022691	.0014915	1.52	0.128	-.0006542	.0051924
dknows	.2475981	.0785316	3.15	0.002	.0936789	.4015172
knowperaids	.0131941	.091013	0.14	0.885	-.1651882	.1915764
_cons	-1.046599	.3543488	-2.95	0.003	-1.741109	-.3520877

Note1 :Edu: education in ascending categories, numkids is number of own children, hsize is household size, nkids5 is number of children 5 or below living in the household, urban equals 1 if household in urban region and 0 if rural region, twater equals 1 if it takes the household more than 30 minutes to collect running water for household use, wqtile1 to wqtile3 are the three lowest wealth quantiles in the population, ownh equals 1 if household own by household members, nomsq is one if there are mosquito nets in the house, partner equals 1 if partner present, employment equals one if the individual is employed, sxage is number of years since sexual debut, dknows is 1 if person has solid knowledge of HIV/AIDS, KNOPERAIDS equals one if person knowns anyone died of AIDS

Table 5 shows that controlling for all variables that may cause both the outcome HIV-positive and the treatment DKNOWS, the latter is significant and positive. Thus, it would suggest the possibility of endogeneity between outcome and treatment. This is possible if we think that despite individuals not knowning their own HIV status, it is nevertheless true that they know better their true unobserved risk behaviour and can therefore seek information to re–affirm a HIV netagive status. The possibility of endogeneity between DKNOWS and HIVP leads us to the possible use of credible instruments that relate closely to the variable DKNOWS but does not necessarely relate to the HIV-positive status of individuals. The questionnaire asks individuals to declare if they listen to the radion, watch televion often and if they read newspapers and magazines. In the case of females they also ask them if they were visited by a family planning officer in the last 12 months. We may think that listening to radio, watchin tv and reading newspapers as well as visits from family planning officers, as variables that provide information on the HIV/AIDS situation in the country but may not necessarely determine the HIV status of individuals (other than allowing for information may lead to less risky behaviour which is what we are trying to capture). Therefore, we use these four variables as instruments

to estimate the effect that instrumented HIV/AIDS knwledge may have on the HIV-status of individuals. Table 6 below shows the results.

Table 6: Two Stage estimation, Instrumenting DKNOWS, Females

```
Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    278
-------------+------------------------------           F( 19,  2766) =    5.7
       Model | -20.7288654    19 -1.09099291           Prob > F      =  0.000
    Residual |  220.826855  2766  .079836173           R-squared     =      .
-------------+------------------------------           Adj R-squared =      .
       Total |   200.09799  2785   .07184847           Root MSE      =  .2825

------------------------------------------------------------------------------
        hivp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dknows |  -.1923331    .119203    -1.61   0.107     -.426069    .041402
         age |   .0028198   .0007439     3.79   0.000      .0013611    .004278
         edu |   .0113986   .0181455     0.63   0.530     -.0241815    .046978
       hsize |  -.0130308   .0026289    -4.96   0.000     -.0181857   -.007875
      nkids5 |   .0170491   .0077489     2.20   0.028      .0018549    .032243
       urban |   .0433051   .0181669     2.38   0.017      .0076831    .078927
      twater |   .0216528   .0127055     1.70   0.088     -.0032604    .046566
     wqtile1 |  -.0799551   .0273335    -2.93   0.003     -.1335512   -.026359
     wqtile2 |  -.0432168   .0214674    -2.01   0.044     -.0853104   -.001123
     wqtile3 |  -.0361638   .0185787    -1.95   0.052     -.0725933    .000266
       ownh |  -.0294247   .0178861    -1.65   0.100     -.0644962    .005646
      nomsq |  -.0431093   .0144002    -2.99   0.003     -.0713455   -.014875
     catolic |   .0047351   .0392389     0.12   0.904     -.0722055    .081675
     protest |   .0093845   .0383056     0.24   0.806      -.065726    .084495
      muslim |  -.0601035   .0411177    -1.46   0.144      -.140728    .020520
     partner |  -.0254273   .0144898    -1.75   0.079     -.0538392    .002984
    employed |    .021805   .0146194     1.49   0.136     -.0068611     .05047
       sxage |   .0006044    .000265     2.28   0.023      .0000847     .00112
    electric |  -.0605652   .0192118    -3.15   0.002     -.098236   -.022894
       _cons |    .212137   .0557715     3.80   0.000      .102779    .321494
------------------------------------------------------------------------------
Instrumented:  dknows
Instruments:   electric ivradio ivtele ivread vistiFP radio
               F( 23,  2762) =   22.41
------------------------------------------------------------------------------
```

Note :See Note 1, Table 5. The variables IVREAD, IVTELE, IVRADIO and VISTIFP are all instruments for the variable DKNOWS. The variable IVREAD is one if person reads newspapers often, the variable IVTELE is 1 if person watches TV often, the variable IVRADIO is one if person listens to radio often and the variable VISTIFP is one if person has been visited by family planning officer in past 12 months

An intial first stage estimate shows the four instruments used in the analysis to be significant and positively associated with the variable DKNOWS. The F-test cannot reject these variables as valid instruments and therefore we use them to instrument the treatment variable DKNOWS. From Table 6 we see that having HIV/AIDS knowledge is negatively associated with the outcome variable HIV-positive, therefore it seems that sound knowledge of HIV/AIDS leads to a reduced risk of becoming HIV-positive (controlling for ascending order of education). The same results suggest that living with a partner and in large household sizes is associated with lower chances of HIV-positive status, thus single female are more likely to be HIV positive even in the presence of HIV/AIDS knowledge.

We now turn out attention to the sample of males to understand if we get similar results as those obtained with females. In the case of males we add information with regards to number of wives (numwives), if the person works away from home in past 12 months (away12) and if the person was circumsised (cirmums). In the case of males, an initial probit estimating the effect of DKNOWS on the outcome HIV-positive (controlling for similar covariares as in the

subsample of females) shows that the variable DKNOWS results in a positive coefficient that is not significantly different than zero. Thus, in the case of male the direction of the treatment is similar to that of females but insignificant. Table 7 shows these results:

Table 7: Conditional estimate on the probability of HIV-positive, given HIV/AIDS knowledg

Probit regression — Number of obs = 2456; LR chi2(23) = 141.32; Prob > chi2 = 0.0000; Log likelihood = -363.05757; Pseudo R2 = 0.1629

hivp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0180888	.0100667	1.80	0.072	-.0016416	.0378192
edu	.0621138	.0759404	0.82	0.413	-.0867267	.2109542
numkids	-.0290899	.0279659	-1.04	0.298	-.0839021	.0257223
hsize	-.0761938	.0264038	-2.89	0.004	-.1279444	-.0244432
nkids5	.1647939	.069045	2.39	0.017	.0294682	.3001196
urban	.2460474	.1383856	1.78	0.075	-.0251834	.5172782
twater	.1813965	.1210144	1.50	0.134	-.0557874	.4185804
wqtile1	-.1544965	.2004769	-0.77	0.441	-.5474241	.238431
wqtile2	-.0393109	.1787655	-0.22	0.826	-.3896849	.3110631
wqtile3	-.2981633	.2010459	-1.48	0.138	-.692206	.0958794
ownh	-.0615504	.1543911	-0.40	0.690	-.3641514	.2410506
nomsq	-.1495934	.1167912	-1.28	0.200	-.3784999	.0793131
catolic	-.0830222	.2205248	-0.38	0.707	-.5152428	.3491984
protest	-.0374403	.2096496	-0.18	0.858	-.448346	.3734654
muslim	-.1651305	.2865392	-0.58	0.564	-.726737	.396476
partner	-.2024158	.215893	-0.94	0.348	-.6255584	.2207268
employed	.4117261	.1691823	2.43	0.015	.0801349	.7433172
numwives	.1920702	.151003	1.27	0.203	-.1038903	.4880308
away12	.0064901	.0041807	1.55	0.121	-.0017039	.0146841
smoke	.1226853	.124778	0.98	0.325	-.1218751	.3672458
circums	-.7781325	.1203659	-6.46	0.000	-1.014045	-.5422197
sxage	.0059276	.0082352	0.72	0.472	-.010213	.0220682
dknows	.0905149	.1133308	0.80	0.424	-.1316093	.3126392
_cons	-1.947951	.3746336	-5.20	0.000	-2.682219	-1.213683

Note :See Notes 1, Table 5

The outstanding result in Table 7 is the effect of circumcision on HIV-positive status: there is a strong causal effect such that circumcision drastically reduces the probability of becoming HIV postive, a result that is nowadays well known in the medical field. Notice that we can accentuate 'causal effect' and not just mainly a correlation, because circumcision is most likely to occurre at a very early age so that reverse causality between circumcision and HIV-positive status may is unlikely (although it could be discussed if there is a possibility for third unobserved variables to be determinant of these effects). In order to deal with the possible endogeneity of DKNOWS, we perform a similar analysis as for the sample of females: using three instruments (IVREAD, IVTELE and IVRADIO) we instrument the treatment variable DKNOWS. The estimated second stage of a two stage probit outcome is preseneted in Table 8

Table 8: Two Stage estimation, Instrumenting DKNOWS, Males

```
Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    2456
-------------+------------------------------           F( 23,  2432) =    6.75
       Model | -2.94429522    23 -.128012836           Prob > F      =  0.0000
    Residual |  103.455289  2432  .042539181           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |  100.510993  2455  .040941342           Root MSE      =  .20625

------------------------------------------------------------------------------
        hivp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dknows |   .1446922   .0700734     2.06   0.039     .0072824     .282102
         age |   .0006795   .0009198     0.74   0.460    -.0011241    .0024832
         edu |  -.0153924   .0131226    -1.17   0.241    -.0411251    .0103402
    numkids |  -.0018498   .0025276    -0.73   0.464    -.0068062    .0031067
       hsize |  -.0054023   .0018453    -2.93   0.003    -.0090209   -.0017838
      nkids5 |   .0092501   .0056341     1.64   0.101    -.0017981    .0202982
       urban |   .0148947    .012951     1.15   0.250    -.0105015    .0402909
      twater |   .0242623   .0103172     2.35   0.019     .0040309    .0444937
     wqtile1 |   .0018727    .016262     0.12   0.908     -.030016    .0337614
     wqtile2 |   .0024468   .0144178     0.17   0.865    -.0258258    .0307193
     wqtile3 |   -.013726   .0136698    -1.00   0.315    -.0405316    .0130797
       ownh  |  -.0140697    .013861    -1.02   0.310    -.0412503    .0131108
       nomsq |  -.0132652   .0104758    -1.27   0.206    -.0338076    .0072773
     catolic |  -.0095061   .0196003    -0.48   0.628    -.0479412     .028929
     protest |  -.0030856   .0183213    -0.17   0.866    -.0390126    .0328413
      muslim |   .0198215   .0255532     0.78   0.438    -.0302868    .0699297
     partner |  -.0327985   .0219239    -1.50   0.135      -.07579     .010193
    employed |   .0083311    .01431     0.58   0.560     -.01973    .0363921
    numwives |   .0316731   .0170935     1.85   0.064    -.0018463    .0651925
     away12  |   .0013862   .0004999     2.78   0.006     .0004077    .0023646
       smoke |   .0146492   .0105276     1.39   0.164    -.0059948    .0352931
    circums  |  -.0934734   .0122838    -7.61   0.000    -.1175611   -.0693857
       sxage |    .001354   .0007319     1.85   0.064    -.0000811    .0027892
       _cons |   .0337292   .0338444     1.00   0.319    -.0326376     .100096
------------------------------------------------------------------------------
Instrumented:  dknows
Instruments:   sxage ivread ivradio ivtele
               F( 25,  2430) =   20.80
------------------------------------------------------------------------------
```

Note :See Notes 1, Table 5

In the case of males, the result seem to contradict the findings that determined the relation between HIV/AIDS knowledge and HIV-status for the sub-sample of females: for males, if individuals have sound knoweledge on HIV/AIDS the effect of this is to increase the chances of becoming HIV-positive. Notice also that the variable SXAGE is positive and significant. This variable explains the number of years an individual has spent as sexually active. The positive relation suggest that longer time exposed to the virus leads to higher probability of becoming HIV positive; this could be explained by the fact that information campaigns were late to arrive in Kenya (late 1999, early 2000) when the prevalence had already picked to high levels relative to other countries in the area. A high prevalence rate may induce individuals to pick up knowledge even specially if these are in contact to environments where morbidity and mortality as result of AIDS is the norm. Thus, high level of knowlege is aquired by those who are more likely to find themselves more exposed to the virus, so that environmental conditions (and not knowledge itself) that determines individual's knowledge, would explain the high positive correlation between HIV/AIDS knowledge and HIV-positive status. Notice that this may not be the case for females since they are not as exposed to the environment that alerts of the presense of HIV/AIDS (e.g., females do not visit sex workers, females are not equally likely to work outside their own village, etc).

Within this preliminary results we have found that knowledge does seem to explain a HIV-positive status although in the case of females this seems to suggest that knowledge empowers

them to avoid infection, whereas in the case of males having solid knowledge on HIV/AIDS seems to suggest knowledge of an environment that has high contact with HIV/AIDS related risks, but contact with such an environment, at the same time, increases the chances that males acquire the virus.

5.1 Subjective beleives and Objective HIV-status by health knowledge

Clearly, knowledge seems to help females to avoid becoming HIV-positive, while males seem to react adversely. In either case, the transmission of HIV to those who are HIV-negative in the population (incidence rate) is mostly determined by the risk-behaviour of individuals who are HIV-positive. In our data set individuals do not know their own HIV status and at the same time they are asked to declare 'their subjective beleives in terms of becoming HIV-positive'. The answer to this question is a question that classifies individual's beleives into 5 different categories. Table 9 shows the distribution of these between males and females (only individuals that undergo the anonimous HIV-test):

Table 9: Relation between Subjective HIV by Gender and Health Knowledge (actual HIV-positive in brackets by sub-groups)

	FEMALES			MALES	
	Has sound health knowledge DKNOWS=1	Does not have sound health knowledge DKNOWS=0		Has sound health knowledge DKNOWS=1	Does not have sound health knowledge DKNOWS=0
No risk at all	453 (**25**)	797 (**23**)		479 (**13**)	596 (**15**)
Small risk	663 (**64**)	615 (**34**)		990 (**16**)	508 (**47**)
Moderate	280 (**24**)	219 (**17**)		176 (**3**)	76 (**7**)
Great	142 (**21**)	143 (**9**)		84 (**1**)	43 (**3**)
TOTAL	**1,538 (134)**	**1,774 (83)**		**1,729 (33)**	**1,223 (72)**

Table 9 shows that informaiton does not necessarely affect individual's perceived subjective chances of becoming HIV-positive. The less inform individuals are regarding HIV/AIDS, the more likely they are to suggest 'not to be at any risk, or very small riks', while at the same time the percentage of actual HIV-positive indivuals is found in the groups who beleive 'inmunity' towars the virus and whose HIV/AIDS health knowledge is at the lowest. For example, Table 9 shows that in the case of females, the larger pecentage declaring poor HIV/AIDS knowledge are also the group in declaring 'not to be in any risk of becoming HIV positive' (797 out of 1,774, or 45% of those who ignore most of what is important to know to avoid HIV infection). At the

same time, the group of females that hold poor HIV/AIDS knowledge is such that it holds 39% of those that are HIV positive. Moreover, 67% of those who are HIV positive and declare poor knowledge in the female group do in fact beleive to have small or no chance to become HIV positive thus leading to fatal outcomes in terms of transmission to others that may not be HIV positive. For the groups of males the picture is equally pesimistic. 6% of those who declare poor HIV/AIDS knowledge are in fact HIV-positive, and among these HIV-positive individuals 86% beleive to have no chance or very small change of ever becoming HIV positive. Therefore, these individuals are at a very high risk of tranmission to others. Table 9 complements our previous results: these showed that only females are likely to react to health knowledge and use it to stop becoming HIV positive, whereas the opposite is true for males. At the same time, risk behaviour on behalf of males who do not react ot HIV/AIDS knowledge suggest that the tranmission rate will continue to determine the HIV prevalence curve, at least in the case of Kenya.

6 Conclusions (preliminary)

We have looked at a random sample of individuals from Kenya who provided information on socio-economic conditions as well as allowing for blood samples to understand their HIV-status even fi this was not disclosed to them. We aimed at understanding the effect that past acquired HIV/AIDS related knowledge could have on individual's HIV-status. The problem is that of endogeneity because even if individuals are not aware of their onw HIV-status potential subjective knowledge may may lead them to seek information, even if this is to re-inforce their own beleive of a negative HIV-status. Thus, to solve the endogeneity problem this paper makes use of instruments that we beleive are capable of explaining HIV/AIDS related knowledge and at the same time these instruments do not necessarely determine HIV status. Instrumenting HIV/AIDS knowledge we find that females effectively use informaton to prevent becoming HIV-positive whereas for males the picture is different: knowledge of HIV/AIDS suggest a higher probability of becoming HIV-positive. Since the endogeneity problem (reverse causality) cannot explain such an outcome, we believe that the result shows that males may be more likely to become in contact with environments where HIV/AIDS is more pronounced (e.g., sex workers, migration) and that such environments are also determining their own HIV-positive status. Furthermore, our results suggest that even in the presence of solid HIV/AIDS knowledge

individual's subjective beleives on their own HIV-status are in clear conflict with their actual HIV-status. Those who unknown to themselves are HIV-positive are also more likely to have relatively poor HIV/AIDS knowledge and, at the same time, are more likely to beleive that becoming HIV-positive is a very unlikely event. This is more the case for males than for females. This results seem to support the conclusion that information campaigns are perhaps reaching only some sub-sectors in the population (e.g., females as result, posibly, of antinatal care) but do not necessarely reach those who are more at risk of becoming HIV-positive and at the same time, are more likely to transmite the illness to others in the population.

7 Reference

Beerenwinkel, N., T. Sing, T. Lengauer, J. Rahnengfuhrer, et al., "Computational methods for the design of effective therapies against drug resistant HIV strains", Bioinformatics, Review, 2005, 21, 3943-3950.

De Luca, A, A. Cozzi-Lepri, CF Perno, C. Balotta, S. Di Giambenedetto, A. Poggio, G Pagano, et al., "Variability in the interpretation of transmitted genotypic HIV-1 drug resistance and prediction of VL outcomes of the initial HAART by distinct systems", 2004, Antiviral Therapy, 9, 743-752

Fang, C.T, H.M. Hsu, S.J.Twu, M.Y. Chen, J.S. Hwang, J.D.Wang, C.Y. Chuang, "Decreasing HIV transmission after the policy of providing free access to HAART in Taiwan", The Journal of Infections Diseases, 2004, 190, 879-85

Grandvalet, Y., "Least Absolute Shrinkage in Equivalent to quadratic penalization", 1998, Université de Technologie de Compiègne Working paper, France.

Hosseinipour, M., S. Cohen, P. Vernazza, A. Kashuba, "Can Antiretroviral therapy be used to prevent sexual transmision of human immunodeficiency?", Clinical Infections Diseases, 2002:34, 1391-5

Law, M., G. Prestage, A. Grulich, P.de Ven and S. Kippax, "Modelling the effect of combination antiretroviral treatments on HIV incidence", AIDS, 2001, 15: 1287-1294

Montaner, J., R. Hogg, E Wood, T Kerr, M Tyndall, AR Levy, R. Harrigan, "The case for exapanding access to highly active antiretroviral therapy to curv the growth of the HIV epidemic", 2006, The Lancet, 368: 531-536.

Kantor, R., R. Shafer, S. Follansbee, J. Taylor, D. Shilane, L.Hurley, D.P. Nguyen, D. Katzenstein and J. Fessel, " Evolution of resistance to drugs in HIV-1-infected patients failing therapy", AIDS, 2004, 18, 1503-1511.

Quinn, T and S. Reynolds, "A randomized controlled trial of short cycle intermittent versus continuous HAART for the treatment of chronic HIV infections in Uganda", WP from Johns Hopkins Bloomberg School of Public Health, 2007

Rabinowitz, M., L.Myers, M. Banjevic, A. Chan, J. Singer, J. Haberer, K. McCann, R. Wolkowicz, "Accurate prediction of HIV-1 drug response from the RT and Protease AA squences using spare models created by convex optimization", Bioinformatics, 2005 (advanced edition)

Schackman, B., K. Gebo, R. Walesky, E. Losina, T. Muccio, PE Sax, MC Weinstein, GR Seage, RD Moore, KA Freedberg, " The lifetime cost of current HIV care in the USA", 2006, Med Care 44, 990-997

Srinivasa, R., M. Kakehashi, "Incubation-time distribution in back-calculation applied to HIV/AIDS in India", Mathematical Biosciences and Engineering, 2000, 2:2, 263-277

Tibshirani, R., "Regresson shrinkage and selection via the LASSO", 1996, Journal of the Royal Statistical Society, 1996, 267-288.

Manski, C., (1989), Anatomy of the selection problem, Journal of Human Resources, 24, 343-360

Manski, C., (1990), Non-parametric bounds on treatment effects, American Economic Review, Papers & Proceedings, 80, 319-323

Manski, C., (1994), The selection problem in 'Advances in Econometrics', C. Sims (ed.), Cambridge University Press, 143-170

Manski, C., (1995), Identification problems in the social sciences, Harvard University Press

Manski, C., (1997), Monotone treatment response, Econometrica, 68, 1311-1334

Manski, C., and J. Pepper, (2000), Monotone instrumental variables with an application to the returns to schooling, Econometrica, 68, 997-1010