Universität St.Gallen

The Virgin HIV Puzzle: Can Misreporting Account for the High Proportion of HIV Cases in Self-Reported Virgins?

Eva Deuchert

July 2010 Discussion Paper no. 2010-24

Department of Economics

University of St. Gallen

Editor: Martina Flockerzi
University of St. Gallen
Department of Economics
Varnbüelstrasse 19
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email vwaabtass@unisg.ch

Publisher: Department of Economics
University of St. Gallen
Varnbüelstrasse 19
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35

Electronic Publication: http://www.vwa.unisg.ch

The Virgin HIV Puzzle: Can Misreporting Account for the High Proportion of HIV Cases in Self-Reported Virgins?[1]

Eva Deuchert

Author's address:

Dr. Eva Deuchert
Center for Disability and Integration
Rosenbergstrasse 51
9000 St. Gallen
Phone +41 71 2242318
Fax +41 71 2203290
Email eva.deuchert@unisg.ch
Website http://www.cdi.unisg.ch

[1] This paper has been presented at seminars at the University of St. Gallen, Cost Conference on the Evaluation of European Labour Market Programmes (2008, Paris), Annual Conference of the Verein für Socialpolitik: Research Committee Development Economics (2008, Zurich), Conference on Sustainable Development (2008, Strasbourg), Health Econometrics Workshop (2008, Milan), and the World Congress on Public Health (2009, Istanbul). I thank participants for helpful comments and discussions. The usual disclaimer applies.

Abstract

It is widely believed that HIV is predominantly sexually transmitted in Sub Saharan Africa. This claim which is inconsistent with national representative data from Lesotho, Zimbabwe, and Swaziland, which reveals that a significant proportion of HIV infections occurred in adolescents who claim to be virgins. Two explanations for this observation have been proposed: adolescents misreport sexual status or non-sexual risks are more prevalent than previously asserted. This paper empirically uncovers the implicit assumptions underlying this discussion, by estimating the proportion of sexually transmitted HIV infections assuming that misreporting is irrelevant, and the proportion of misreporting necessary to conclude that HIV is predominantly sexually transmitted. It shows that under the no-misreporting assumption, 70% of HIV cases in the respective sample of unmarried adolescent women is not due to sexual transmission. The assumption that HIV is predominantly sexually transmitted is only valid, if more than 55% of unmarried adolescent women who are sexually active have misreported sexual activity status. This research is designed to gain better understanding on the importance of different transmission modes. This is important to design combination prevention to achieve maximum impact on HIV prevention.

Keywords:

Population attributable fraction; non-classical measurement error; HIV transmission mode

JEL Classification

C13, C14, I12

1. Introduction

It is commonly believed that sexual HIV transmission is the dominant transmission mode in Sub Saharan Africa, accounting for more than 95% of all HIV infections (Ezzati, et al., 2006; Schmid, et al., 2004). For this reason, HIV prevention strategies usually focus on the prevention of sexual transmission through HIV counseling, testing and disclosure, delay of sexual debut, partner limitation, couples counseling and testing, safer sex and normalizing condom use (UNAIDS, 2007).

The belief that HIV is predominantly sexually transmitted however, is inconsistent with national representative data, where various surveys from Sub Saharan Africa document high infection rates among self-reported virgins (Brewer, Potterat, Muth, & Brody, 2007; Gavin, et al., 2006). A study on HIV infections in adolescent women in Zimbabwe for example, shows that **41% of HIV positive women** report themselves to be virgins (Gavin, et al., 2006). Two alternative explanations for the high infection rates in self-reported virgins were proposed: (1) adolescent women misreport sexual status (Gavin, et al., 2006), or (2) a significant number of adolescents have acquired HIV by non-sexual HIV transmission routes (Brewer, Potterat, Muth, & Brody, 2007). The issue whether "virgin" HIV cases are explained by misreporting or non-sexual HIV transmission has yet not been solved. This question however, is of primary importance for research and policy. A high proportion of women misreporting sexual behavior raises the question whether self-reported sexual data should be used to monitor the HIV/AIDS epidemic. Alternatively, a high proportion of non-sexual HIV transmission may bring up the issue whether the prevention of non-sexual HIV risks is sufficiently covered by the current prevention paradigm. This paper contributes to the "virgin" HIV debate by empirically uncovering the implicit assumptions underlying the discussion. It proposes an empirical model that allows to estimate the proportion of sexually transmitted HIV infections assuming that misreporting is irrelevant, and the proportion of misreporting necessary to conclude that HIV is predominantly sexually transmitted. From such an analysis one can gain a better understanding of the nature of the problem, and of the validity of the arguments.

The spread and the control of HIV have traditionally been discussed in the medical literature. With the availability of large nationally representative surveys however, the African HIV/AIDS epidemic became focus of economic research aimed at understanding how behavioral and socio-economic risk-factors facilitate the spread of HIV in Sub Saharan Africa. Recent examples are studies from Bongaarts, et al. (2008),

Bene and Merten (2008), or Awusabo-Asare and Annim (2008). This research usually focuses on sexual behavior as the major or the only mode of transmission. Oster (2005) for example, models the HIV epidemic in the United States and Sub-Saharan Africa assuming that HIV is only heterosexually transmitted. De Walque (2007) analyzes data from discordant partnerships and argues that those women who are HIV infected but do not report extramarital sexual relationships have misreported sexual behavior. He concludes that unreported extramarital sex is substantial source for HIV acquisition among cohabiting women.

The focus on sexual transmission alone in research and policy however, may fall too short if alternative transmission modes, particularly transmission through unsafe health care (i.e. re-using injection equipment, administering contaminated blood transfusions, and using unsterile surgery equipment), plays an important role in spreading HIV/AIDS in Sub Saharan Africa (Gisselquist, Rothenberg, Potterat, & Drucker, 2002; Schneider & Drucker, 2006). It is therefore important to revisit and challenge previous assumptions regarding the importance of different transmission modes in Sub Saharan Africa.

When empirically analyzing the importance of different transmission modes, one is hampered by corrupt data as people are likely to misreport sexual behaviors. Various studies demonstrate that misreporting is relevant in many sex surveys because one can observe inter-couple disagreements concerning certain sexual practices and condom use, or gender discrepancies in the number of sex partners (de Boer, Celentano, Tovanabutra, Rugpao, Nelson, & Suriyanon, 1998; Catania, et al., 1995; Smith T. , 1992). It is also argued that the initiation of sexual intercourse is subjected to misreporting, because data from subsequent cross-sectional surveys document inconsistencies in within-cohort responses (Gersovitz, 2005; Glick & Sahn, 2008), and because studies using biomarkers (tests for sexually transmitted infections -STI) reveal high STI prevalence in self-declared virgins (Buvé, et al., 2001; Mensch, Hewett, Gregory, & Helleringer, 2008). This literature however, can only document inconsistencies but cannot quantify the extent of misreporting. It remains therefore unclear, whether misreporting is important enough to account for the majority of "virgin" HIV cases observed in the data.

To empirically uncover the assumptions underlying the debate on "virgin" HIV this paper estimates population attributable fractions (= HIV prevalence that would be observed if the population were sexually inactive, compared with its current sexual

activity pattern). In its standard form, this estimator assumes accurate data. To empirically analyze how much misreporting would be necessary for a predominantly sexually transmitted epidemic, this paper applies a framework developed in the empirical evaluation literature that seeks to measure a causal effect of a “treatment” variable, when the treatment is misreported and the measurement error is not independent of its true value (non-classical measurement error).

Several empirical models have been developed to address the problem of non-classical measurement errors. Mahajan (2006), Lewbel (2007), and Hu (2008) for example, show that a point identification of a causal effect can be achieved if an instrument (i.e. an independent measure of the treatment) was available. For the research question of this paper however, this would require identifying a variable which is related to "true" sexual activity but not to misreporting sexual behavior. Such a variable is not available. Battistin and Sianesi (2006) show how to construct bounds for the true effect, when an instrument is not available by making a priori assumptions on the extent of misclassification. Here, this method will be modified to estimate population attributable fractions under various assumptions on misreporting. From this analysis one can identify the assumption on misreporting that leads to a predominantly sexually transmitted epidemic.

To empirically estimate population attributable fraction under various assumptions on the extend of misreporting, the paper uses data from unmarried adolescent women living in Lesotho, Swaziland, and Zimbabwe. Assuming that reported behavior is accurate, only 30% of HIV infections in the population can be attributed to sexual HIV transmission. For a predominantly sexually transmitted epidemic within the considered population, where more than 95% of HIV infections is caused by sexual transmission, one need to assume that more than 55% of unmarried women aged 15-19 who are sexually active have misreported sexual activity.

The rest of the paper is organized as follows: The next section outlines the model that is used to estimate the proportion of sexual HIV transmission. Section 3 provides an empirical application using data from Lesotho, Swaziland, and Zimbabwe. The final section concludes.

2. An empirical model of sexual HIV transmission under misreporting

2.1. Identification

The public health and health economic literature typically models sexual HIV transmission in the context of a Bernoulli-risk simulation model (see for instance in Oster 2005). In these models each sexual episode or sexual partnership is treated as an independent trial and the probability of a "success" (HIV infection) is assumed to equal the per-contact (or per-partner) probability of HIV transmission. This approach requires having knowledge about the number of partnerships, the number of unprotected sexual episodes per partner, the partner matching process, the HIV prevalence in partners, and the sexual HIV transmission efficiency. These variables are often not available and must be based on assumptions, are measured with a great uncertainty, or are constrained by misreporting. Simulation outcomes are therefore constrained by great uncertainty.

An alternative to simulation models is to empirically estimate population attributable fractions -*PAFs*-, which are defined as the *"proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal"* (Rockhill, Newman and Weinberg 1998, 15). PAFs rely on an empirical counterfactual analysis, where the effect of a risk factor is estimated by comparing the current disease burden with the levels that would be expected under an alternative (hypothetical) scenario.

PAFs are widely used in the public health literature, because they allow to link information about a hazardous effects to data on exposure to this risk factor. In the most general form, the PAF can be expressed as (Eide and Heuch 2001):

$$PAF = \frac{E(Y) - E(Y|A^* = 0)}{E(Y)} \quad (1)$$

where $E(Y)$ is the proportion of HIV (prevalence) in the population, and $E(Y|A^* = 0)$ represents the proportion of disease if the exposure to the risk-factor “sexual activity” A^* would be reduced to zero (or in other words if no one would be sexually active). The PAF is therefore equivalent to the proportion of sexually acquired HIV infections in the population.

To estimate equation (1), one needs to consistently estimate the prevalence of HIV from non-sexual transmission. By definition, all HIV infections in true virgins have acquired HIV by non-sexual transmission routes. HIV infections in virgins however, reflect only the lower bound for $E(Y|A^* = 0)$ since sexually active individuals may have been also

exposed to non-sexual risks. A consistent estimator for the total prevalence of HIV from non-sexual transmission therefore needs to take nonsexual transmission in sexually active respondents into account.

To estimate the probability of non-sexual HIV transmission, the potential outcome notation is used. Let Y^1 denote the outcome for an individual that is sexually active and let Y^0 be the outcome for the same individual if she was not sexually active. If A^* is an indicator for sexual activity, the realized (observed) outcome Y for an individual equals

$$Y = Y^0(1 - A^*) + Y^1A^*$$

HIV prevalence from non-sexual transmission is $E(Y^0)$, which is the HIV prevalence that would occur if no one was sexually active [$E(Y|A^* = 0) = E(Y^0)$]. To consistently estimate $E(Y^0)$, two problems need to be solved: (1) One can in principle observe a person in only one state (sexually active of inactive) so that the counterfactual outcome can never be observed, and (2) misreporting of sexual activity is not observable, so that it is unclear whether the observed outcome Y is the outcome for an individual who is truly sexually active or inactive.

To solve these two problems, the paper modifies the ideas from (Battistin and Sianesi 2006), who estimate bounds for the causal effect of education on labor outcomes, given that education is misreported. This approach is based on the following assumptions:

(i) Suppose we had *prior information* on respondents who provide inaccurate answers. This information allows to observe subgroups, who provide accurate answers (*verification of observed subgroups*, VOS). In the current example I assume that respondents who admit sexual intercourse do not misreport, and that misreporting is only relevant among women who self-report being a virgin. The assumption simplifies the estimation procedure and can be easily justified by common traditional cultural values for young respondents to remain virgins until marriage, which set an incentive for misreporting for sexually active respondents but not for virgins (Cowan, et al. 2002). Let A be an indicator that a woman *declares* being sexually active, Z indicates whether a respondent provides inaccurate information, and A^* is the accurate sexual activity status, *VOS* can be written as:

$$P(Z = 1|A^* = 1) = P(A = 0|A^* = 1) = \lambda \geq 0$$
$$P(Z = 1|A^* = 0) = P(A = 1|A^* = 0) = 0$$

(ii) The *Conditional Independence Assumption* -CIA- states that the true sexual activity status A^* is independent of the potential outcomes conditional on the value of suitably chosen covariates:

$$Y^1, Y^0 \coprod A^* | X = x$$

CIA - alternatively called "ignorable treatment assignment" (Rosenbaum und Rubin 1983) or "selection on observables" in the regression framework (Barnow, Cain und Goldberger 1980) - is a standard assumption in empirical research used to solve the selection bias. The selection problem arises even if the true sexual activity status was known. The identification of the effect would be precluded if virgins and sexual active women systematically differ along several dimensions which are relevant to the health outcome. For instance it may be the case that sexually active women may be more exposed to other risk factors that are not directly related to sexual HIV transmission (such as iatrogenic HIV transmission for example). CIA assumes that all these confounding factors are observable and can be controlled for. In this case, it is as if randomization had happened within cells defined by these confounding variables.

In praxis, CIA is not testable. It requires however, to control for all variables that jointly affect the potential outcomes (HIV infection status) and true sexual activity. To make this assumption creditable, a good institutional knowledge on relevant confounding variables is relevant. Section 3.2 provides a detailed discussion on relevant confounding variables that had been identified in the empirical literature.

(iii) The *Non-Differential Misclassification Assumption* -NDM- solves the problem that the confounder distribution of respondents who misreport sexual activity is unknown. NDM assumes that the reported sexual behavior does not contain information to predict the health outcome of interest conditional on the true sexual activity and a set of relevant confounding variables:

$$Y^1, Y^0 \coprod A | X = x, A^*$$

This assumption would be violated if the respondents were aware of their HIV status and tried to hide the infection by lying about their true sexual activity status. This is unlikely to be the case in the current application for two reasons: (1) HIV has a long latency period (Pantaleo, Grazios and Fauci 1993) so that adolescents who acquired HIV by sexual HIV transmission should not show any symptoms yet. Therefore, the only way how adolescents can be aware of their serostatus is by conducting HIV test. Testing prevalence however, is usually very low. During the interview, respondents had

been asked whether they ever had an HIV test. Only 12% of women in the sample used in the application reported that they ever had an HIV test, less than 6% of self-reported virgins reported that they ever had an HIV test. Only six self-reported virgins was subsequently tested to be HIV-positive and reported that she ever had a HIV test. (2) Consent to the interview and the subsequent HIV test is independent. It is unlikely that a women tries to hide a known HIV infection by lying about her sexual activity status but then agrees to a HIV test that would detect the infection.

Assumption (i) to (iii) can now be used to construct an estimator for the proportion of non-sexual HIV transmission. Note that HIV risk conditional on sexual behavior ($A = i$ with $i = 0,1$) can be formulated as

$$E(Y|A = i, x) = E(Y|A^* = 1, A = i, x)P(A^* = 1|A = i, x) + E(Y|A^* = 0, A = i, x)P(A^* = 0|A = i, x)$$

Using the potential outcome notation, the conditional risk for self-reported virgins can be written as

$$E(Y|A = 0, x) = E(Y^1|A^* = 1, A = 0, x)P(Z = 1|A = 0, x) + E(Y^0|A^* = 0, A = 0, x)P(Z = 0|A = 0, x)$$

If NDM hold, this equals

$$E(Y|A = 0, x) = E(Y^1|A^* = 1, x)P(Z = 1|A = 0, x) + E(Y^0|A^* = 0, x)P(Z = 0|A = 0, x) \quad (2)$$

The first part of equation (2) is identified by NDM and the assumption that respondents who admit sexual activity have no incentive to misreport sexual activity status (VOS). It can therefore be directly estimated from data on reported sexual activity: $E(Y|A = 1, x) = E(Y^1|A^* = 1, x)$.

Proof:

$$E(Y|A = 1, x) = E(Y^1|A^* = 1, A = 1, x)P(A^* = 1|A = 1, x) + E(Y^0|A^* = 0, A = 1, x)P(A^* = 0|A = 1, x) = E(Y^1|A^* = 1, A = 1, x) = E(Y^1|A^* = 1, x)$$

and

$$P(A^* = 0|A = 1, x) = P(Z = 1|A^* = 0, x)\frac{P(A^* = 0|x)}{P(A = 1|x)} = 0$$

Thus, the conditional HIV prevalence from self-reported sexual active respondents can be introduced into equation (2), which can be reformulated to get the conditional HIV prevalence from non-sexual transmission:

$$E(Y^0|A^* = 0, x) = \frac{E(Y|A=0,x) - P(Z=1|A=0,x)E(Y|A=1,x)}{1 - P(Z=1|A=0,x)} \quad (3)$$

with

$$P(Z = 1|A = 0, x) = \frac{P(A = 1|X)}{1 - P(A = 1|X)} \frac{\lambda}{1 - \lambda}$$

Thus, the probability that a self-reported virgin misreports is an increasing function of the probability of declaring sexual activity, and the probability of misreporting $\lambda = P(Z = 1|A^* = 1)$ given she is in reality sexually active.

If CIA holds, potential outcomes are mean independent: $E(Y^0|A^* = 0, x) = E(Y^0|x)$. Thus, the equation above denotes the conditional HIV prevalence from nonsexual HIV transmission for respondents with $X = x$.

Consequently, all components of $E(Y^0|x)$ is identified from the observed data and from assumptions regarding the probability of misreporting λ. There are binding constraints on assumptions that can be made on λ because equation (3) need to be between zero and one (see Appendix). The total HIV prevalence of non-sexual transmission can be estimated by integrating the conditional HIV prevalence of equation (3) over the full distribution of X.

2.2. Estimation

The different components of the estimator for the population attributable fraction (PAF) under misreporting defined by equation (1) to (3) can be identified from observed data and can be estimated as following:

HIV prevalence $\mu_Y = E(Y)$ is estimated by the weighted sample average of observed HIV infections:

$$\hat{\mu}_Y = \sum_i Y_i w_i$$

where w_i denotes the respective selection probability. Note that in the application, pooled data from three different countries are used. The corresponding selection probabilities (inverse of the survey weights) are therefore standardized. Each selection probability is divided by the sum of the initial selection probabilities for each country. The standardized selection probabilities are then multiplied with the respective relative population size for each country provided by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population

Prospects: The 2008 Revision. (The inverse of these adjusted selection probabilities denotes the adjusted sample weight.)

To estimate the conditional HIV prevalence from non-sexual HIV transmission we need a consistent estimator for the probability of being HIV positive conditional on self-reported sexual behavior $\mu_Y(j,x) = E(Y|A=j,x)$, the probability to report sexual activity $\mu_A(x) = P(A=1|x)$, and the assumption regarding the probability of misreporting λ_i.

The two conditional probabilities of being HIV positive $\mu_Y(j,x)$ are estimated from a standard parametric Probit model (applying adjusted survey weights):

$$\hat{\mu}_Y(A_i, X_i) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 A_i + \hat{\beta}_2 X_i)$$

Likewise, the conditional probability to report sexual activity $\mu_A(x)$ is estimated from a standard parametric Probit model (applying adjusted survey weights):

$$\hat{\mu}_A(X_i) = \Phi(\hat{\theta}_0 + \hat{\theta}_1 X_i)$$

Finally, the individual probability of misreporting λ_i is allowed to vary between 0 and 1, while trimming this probability to a maximum such that the binding constraints defined in the appendix (C1 and C2) are satisfied.

Combining these three components yields an estimator for the conditional HIV prevalence from non-sexual HIV transmission $\mu_{Y^0}(x)$.

$$\hat{\mu}_{Y^0}(x) = \frac{\hat{\mu}_Y(0,X_i) - \hat{\mu}_Z(0,X_i)\hat{\mu}_Y(1,X_i)}{1 - \hat{\mu}_Z(0,X_i)}$$

with $\hat{\mu}_Z(0,X_i)$ being the estimator for the probability that a self-declared virgin misreports sexual activity $P(Z=1|A=0,x)$:

$$\hat{\mu}_Z(0,X_i) = \frac{\hat{\mu}_A(X_i)\lambda_i}{[1-\hat{\mu}_A(X_i)](1-\lambda_i)}$$

To obtain the unconditional probability of non-sexual HIV transmission, the conditional probabilities are averaged over the empirical distribution of the covariates:

$$\hat{\mu}_{Y^0} = \sum_i \hat{\mu}_{Y^0}(x) w_i$$

Finally, the population attributable fraction is then estimated by

$$\widehat{PAF} = \frac{\hat{\mu}_Y - \hat{\mu}_{Y^0}}{\hat{\mu}_Y}$$

Note that the procedure outlined above seems somewhat restrictive since a parametric model for the probability of HIV infection conditional on self-declared sexual activity is

assumed. In principle, one could estimate the unconditional probability of non-sexual HIV transmission directly using a semi-parametric inverse probability weighting estimator, that does not rely on a parametric model for the probability of HIV infection. This is not done in the current application for the following reason: In standard applications one may have some information on the extent of misreporting. This however, is not the case in the current context. We do not know the extent of misreporting and the estimation is done under different (sometimes extreme) assumptions on the extent of misreporting. In this situation, one cannot assume that the binding constraint C2 is automatically satisfied for all individuals. Therefore, it is necessary to trim the probability of misreporting for individuals for whom the condition is not fulfilled. To so that, a parametric model for the conditional probability of HIV infection is needed.

3. Empirical application

3.1. Data

The data is from the Demographic and Health Surveys (DHS) in Lesotho, Zimbabwe, and Swaziland. DHS include interviews on maternal and child health, family planning, nutrition and related issues, and also collect biological and clinical data such as HIV testing. A description of DHS can be found on ORC Macro's webpage http://www.measuredhs.com/ or in Mishra, et al. (2006). The countries are chosen because they have sizeable HIV epidemics with HIV prevalence rates (measured as average HIV prevalence in adults age 15 to 49) exceeding 10%.

The surveys are conducted by face-to-face interviews, where women are questioned by female interviewers, and male are questioned by male interviewers. The interviewer strives to interview the respondent alone, since a third person during an interview can prevent respondents from giving honest answers. If this is not possible, the presence of another person is coded. The surveys are designed such that the questionnaire on socio-economic variables and individual behavior is independent from HIV testing. After finishing the questionnaire, all men and women who live in a randomly selected sub-sample of households (usually one third of survey households) are asked whether they accept a HIV test. Rates of consent for HIV testing are high, where more than 70% of all selected respondents accepted being tested for HIV. Acceptance is usually higher among women and younger respondents where acceptance rates exceed 80%. In all three surveys, HIV prevalence among men and women who self-report virginity is

relatively high ranging between 2.1% and 3.7% in male virgins and 3.9% and 5.0% in female virgins.

In the following analysis, only female respondents will be considered. Differences in behavior and biological susceptibility between men and women require to separately estimating risk-profiles for men and women. However, overall HIV prevalence in men is relatively low. Therefore too few male HIV cases are reported in the data to allow a detailed analysis.

The analysis is furthermore restricted to adolescents (age 15-19) who declare that they have never been married. These restrictions are necessary because most adults (age > 19) and women in fixed relationships declare to be sexually active (mo common support). To maximize the number of HIV cases, data for the three countries is pooled. In total, 3,098 women are considered. This sample represents about 80% of the female population age 15 to 19.

TABLE 1: HIV prevalence among unmarried adolescent women (age 15-19)

	Full sample	Sexually inactive	Sexually active	Fraction of sexually inactive HIV cases
Total		66.2%	33.8%	
HIV prevalence	7.5%	4.4%	13.8%	38.3%
Lesotho				
Total		67.2%	32.8%	
HIV prevalence	6.8%	5.3%	9.9%	52.5%
Swaziland				
Total		62.6%	37.4%	
HIV prevalence	8.2%	4.4%	14.7%	33.3%
Zimbabwe				
Total		88.4%	11.6%	
HIV prevalence	4.3%	3.1%	12.9%	64.7%

Note: Average values are estimated applying the adjusted sample weights.

Our main variable of interest is self-declared sexual activity. This indicator is derived from the question *"How old were you when you had sexual intercourse for the very first time?"*, where the indicator for sexual activity is coded as $A = 0$ if the women answered with *"never"* and $A = 1$ else (if the question was answered). Note that this question does not relate to the intactness of the hymen. This disregards from a situation, where women practice noncoital behaviors to maintain virginity, which could potentially explain the

high prevalence in self-reported inactive women. Unfortunately, the question does not clearly define sexual intercourse. Quantitative studies have highlighted that most adolescent respondents consider penile-vaginal intercourse and penile-anal intercourse as sex, while many adolescent respondents do not think that oral-genital contact constitutes having sex (Sanders und Reinisch 1999). I do not believe that the common failure to consider oral-genital contact as sex causes a major bias in this application, as it is agreed that HIV transmission through oral-genital sexual intercourse is rare and has a much lower HIV-infection risk than penile-vaginal or penile-anal sex (Rothenberg, et al. 1998). Therefore, I believe that possible interpretation errors are unlikely to explain the puzzle of high infection rates among self-declared virgins.

Table 1 shows that approximately 66% of adolescent women in the considered population report that they are sexually inactive (in the further application we call these women "self-reported virgins"). The highest proportion of self-reported virgins can be found in Zimbabwe (88%); the lowest proportion can be found in Swaziland (63%). Approximately 7.5% of women are HIV positive, with HIV prevalence rates highest in Swaziland and lowest in Zimbabwe. A significant share of HIV infections (38%) occurred in women, who reported to be virgins. This proportion is highest in Zimbabwe (65%) and lowest in Swaziland (33%). This demonstrates that a significant share of HIV infections in this sample is unexplained by *reported* sexual behavior, which could be either a sign for misreporting or for non-sexual HIV transmission. The remaining descriptive statistics are presented in the appendix (Table A1).

3.2. Identification strategy

To estimate the population attributable fraction of sexual activity defined in equation (1) to (3), we rely on the Conditional Independence Assumption and the Non-Differential Misclassification Assumption. These assumptions require, that relevant confounders for the onset of sexual activity, misreporting, and the risk of non-sexual HIV infection need to be controlled for. Unfortunately, CIA and NDM cannot be directly tested. To make these assumptions creditable, it is therefore essential to understand how HIV is transmitted and which factors determine the reliability of the data concerning sexual behavior. From such an analysis, relevant confounders can then be identified.

3.2.1. HIV transmission modes

HIV can be transmitted through direct exposure to contaminated body fluids, such as blood, semen, cervical and vaginal secretions, and breast milk. The transmission routes are (1) mother-to child transmission during pregnancy, labor, delivery, and breast feeding, (2) sexual transmission, and (3) blood-to-blood transmission (blood transfusions, exposure to HIV-contaminated needles, syringes, and other sharp objects).

Mother-to-child transmission

The highly active antiretroviral therapy (HAART) was introduced in 1997, so that effective HIV treatment had not been available to women belonging to this sample at infancy. Because average life expectancy for untreated HIV-positive infants is only 3.75 years (Marston, et al. 2005), long term survivors of mother-to-child transmission should be rare in this age-cohort. Therefore, it is relatively unlikely that HIV cases in this sample are caused by mother-to-child transmission; observed cases must be either due to sexual transmission or due to blood-to-blood transmission.

Sexual transmission

Various studies document that HIV infections are associated with sexual activity. For example a systematic overview of 68 Epidemiological Studies conducted in Sub Saharan Africa identified sexual risk factors such as multi-partner sex, paid sex, and STIs infections as important risk factors for HIV transmission (Chen, et al. 2007). These observations have led to the conclusion that HIV is predominantly sexually transmitted in Sub Saharan Africa accounting for more than 95% of HIV infections (Ezzati, et al. 2006, Schmid, et al. 2004).

Since the current paper considers only unmarried women, the risk factor of interest is premarital sexual activity. Representative surveys reveal that reported median age of first sex is less than 20 in almost all Sub Saharan African countries (Zaba, et al. 2004). Since self-reported age of first sexual activity has either remained constant or increased slightly, while the age at first marriage increased dramatically over the past 15 years, the proportion of young women who engage in premarital sex has increased substantially (Mensch, Grant and Blanc 2006). A higher likelihood to engage in premarital sexual relations is related to socio-cultural differences (D. Smith 2004), economic constraints, which force women to initiate sexual partnerships such as “sugar daddy relationships”

(Luke 2006), and the low bargaining power of women, which leaves them little choice over their own sexual behavior (Clark, Bruce and Dude 2006).

Blood-to-blood transmission

Blood-to-blood transmission occurs through direct contact with contaminated blood mainly through the usage of sharp objects. The two most common blood-to-blood transmission modes are needle sharing (i.e. injecting drug users) and unsafe medical procedures (i.e. contaminated blood transfusions, re-use of contaminated injection and surgery equipment). Unsafe medical procedures are effective in transmitting HIV/AIDS as the outbreak of HIV in children attending the Al-Fateh Hospital in Benghazi, Libya has demonstrated (de Oliveira, et al. 2006).

Relatively little is known about injection drug use in Sub Saharan Africa but it is suggested that injection drug use is an emerging HIV risk as a result of the expansion of international drug trafficking (Beckerleg, Telfer and Hundt 2005, Kloos and Mariam 2007). Needle sharing is very common, and high risk behaviors such as drawing blood back in a syringe and passing it to other drug users (which is used as a method to avoid pains from withdrawal) is frequently observed (McCurdy, et al. 2006, McCurdy, et al. 2007). To my knowledge, risk factors that lead to injection drug use and needle sharing have not been studied in the context of Sub Saharan Africa but European data suggest that among psychological factors and family background, low socio-economic status (Reinherz, et al. 2000, Poulton, et al. 2002), and low education (Johnson, et al. 1995, Stronski, et al. 2000) are associated with illicit substance abuse.

Health care safety is a major concern for public health authorities (WHO 2002). Several studies document the lack of basic supplies and infrastructure as been demonstrated in various Health Service Provision Assessments (GSS 2003, MoHR 2003), the re-use of injection equipment (Hutin, Hauri and Armstrong 2003), or the usage of untested blood transfusions (WHO 2002). Thus, exposure to (potentially unsafe) health care could be a relevant non-sexual risk factor for an HIV infection. The individual demand for health care is typically influenced by the costs of health care, financial and geographic accessibility, barriers that arise from low bargaining power, and individual socio-economic status (Hjortsberg 2003, Sahn, Younger and Genicot 2003).

3.2.2. Reliability of self-reported data

Social science researchers are usually skeptical concerning the accuracy of self-reported sexual behaviors. Unfortunately, self-reports such as the information on the initiation of sexual activity cannot be externally validated. Researchers have used three different strategies to analyze the consistency and reliability of self-reported sexual behavior: (1) The reliability of data is examined within data reported by the same individual or the same age-cohorts at different times, which demonstrated inconsistencies in self-reported age of first sexual intercourse (Gersovitz 2005, Glick and Sahn 2008, Rodgers, Billy and Udry 1982). Inconsistent answers are correlated with observable characteristics (such as race and gender), which suggest that errors in the measurement of sexual activity are not random (Rodgers, Billy and Udry 1982). (2) Examining the impact of the data collection mode, it was noticed that women admit a higher number of partners if the interviewer is female (Wilson, et al. 2002), or that the presence of other people affected the respondent's ability to respond honestly (Aquilino, Wright and Supple 2000). (3) Several studies use biomarkers (i.e. tests for various sexually transmitted infections) to understand whether respondents who reported that they had never had sexual intercourse had misreported sexual activity status (Buvé, et al. 2001, Mensch, Hewett, et al. 2008).

These studies report that misreporting is an issue in surveys on sexual behavior but cannot quantify misreporting. The first two methods can only reveal inconsistencies in reports, but cannot predict the accuracy of the responses; the third method crucially relies on the assumption that sexual transmission is the only transmission mode, which can lead to misleading results if non-sexual transmission modes are also relevant. This is particularly the case since the studies by Mensch, et al. (2008) and Buvé, et al. (2001) use positive HIV tests to detect "inaccurate reports".

3.2.3. Selection of control variables

Blood-to-blood transmission, sexual behavior and data inaccuracy are likely to be influenced by the same factors. CIA and NDM require controlling for factors that jointly influence non-sexual risks, sexual activity, and misreporting. These factors can be categorized into the following groups: (1) socio-economic variables, (2) socio-cultural variables, (3) socio-geographic variables, (4) bargaining power, and (5) survey characteristics.

The Demographic and Health Surveys provide a large set of covariates that can be used to approximate these categories such as age, education, household wealth, occupation, religion, regional indicators, a variable that indicates the presence of other people during the interview, and a set of self-reported barriers to health care, which approximate geographic accessibility, individual bargaining power, and moral values.

Unfortunately, Demographic and Health Surveys are cross-sectional and it is not possible to control for confounding variables before women decide to initiate in sexual activity (which would rule out that variables are influenced by sexual activity or resulting HIV infections). For two reasons I believe that this does not cause a major bias: (1) The period of clinical latency (where patients have no symptoms) is about 10 years (Pantaleo, Grazios and Fauci 1993), and is much longer than the time span of reported sexual activity in this age cohort (max 8 years, median 1 year). Therefore, it is unlikely that education, occupation choice, productivity etc. are influenced by resulting HIV infections because women who acquired HIV by sexual transmission should not show symptoms of AIDS yet. (2) Other variables are measured on the household level (such as wealth, residence, etc.) which is unlikely to be affected by adolescent women's decision to initiate sexual activity or subsequent HIV infections.

A second problem may arise because important information, such as current income and ethnicity, is not available from the Demographic and Health Surveys. Income may increase bargaining power to negotiate sexual activity and may also be an important factor that impacts exposure to non-sexual risks. However, wealth, education, and occupation most likely approximate income. Ethnicity could be associated with certain cultural habits and rituals (ritual circumcision for example), that may be associated with HIV infections. Cultural differences however, are likely to be captured by the indicators for regional differences, which are closely related to differences in ethnic composition.

A further limitation arises from the implicit assumption that the risk increase in HIV prevalence associated with sexual activity is due to sexual HIV transmission. This does not need to be the case because sexually active women are more likely to be exposed to non-sexual risk factors (hormonal injections are a common contraception mode; pregnant women receive antenatal health care; sexually active women receive treatment for sexually transmitted infections). It is not possible to adjust for variables that approximate exposure to health care, because available variables on access to health care (injections in the previous months, visits of health facilities) are constrained by

reverse causality. The paper controls for self-reported barriers to access health care, which is likely to partly capture this effect.

3.3. Predicting the probability of self-reported sexual activity and HIV prevalence

To predict the population attributable fraction, one need to consistently estimate the conditional HIV probability to report sexual activity $\hat{\mu}_A(X_i)$ and the conditional HIV prevalence $\hat{\mu}_Y(A_i, X_i)$. This is done by estimating two Probit models, where the dependent variables are variables indicating that a person reported sexual activity and that a person is HIV positive. The control variables are all variables presented in table A2, dummy variables for 24 regions, and interactions terms between wealth and country dummy variables (not presented). The interaction terms are included because the wealth indicator is constructed for each country separately (Rutstein and Johnson 2004), and thus measures the relative wealth distribution within a country and not across the three countries.

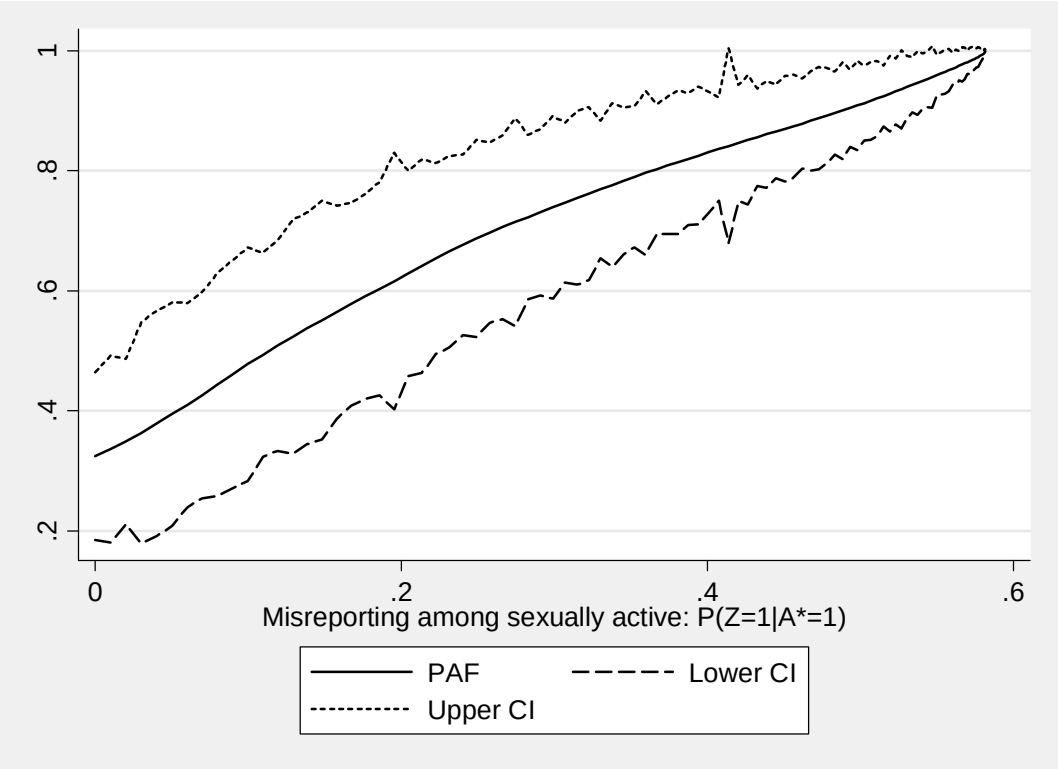
The regression results are presented in the appendix (Table A2). The first column reveals that older age, working (compare to not-working), and not wanting to go alone to a health provider is positively associated with self-reported sexual activity; higher education, being Christian (Roman Catholic and Protestant), the fear that no female health worker may be available, as well as rural residence is negatively associated with self-reported sexual activity. The second column shows that HIV prevalence is significantly higher if a person reports sexual activity. Furthermore, we find a lower HIV infection rate among women with higher education, and among Roman catholics. Higher infection rates are found among elder adolescents (age=19), women who work in household & domestic and other professions, women living in poorer households (compare to women living in the poorest households) , and among women who reported that getting permission is a major barrier to access health care (which could be an indicator for bargaining power).

3.4. Estimating the proportion of sexual HIV transmission

Estimates from the Probit model are used to construct an estimator for the proportion of sexual HIV transmission. HIV prevalence conditional on declaring sexual activity are obtained by a Probit extrapolation. If one is willing to assume that all respondents have correctly specified their sexual activity status ($\lambda = 0$), total HIV prevalence by

nonsexual transmission is equal to 5.1%. The corresponding population attributable fraction can then be estimated by setting the excess HIV prevalence into relation with the total HIV prevalence: $PAF = (7.5 - 5.1)/7.5 = 0.33$ [95% CI: $0.19 - 0.46$]. Thus, the scenario with no misreporting would predict that about 30% of HIV cases in the sample are attributable to sexual transmission, 70% are attributable to non-sexual routes (either before or after the onset of sexual activity).

FIGURE 1: Proportion of sexually transmitted HIV infections under different degrees of misreporting



Note: Confidence intervals are based on bootstrap sampling with 99 repetitions. Different proportion of misreporting among self-reported virgins are displayed on the x-axis.

If misreporting is relevant however, one need to use the model defined by equation (1) to (3) to estimate the population attributable fraction. This is done in Figure 1, which presents the full range of population attributable fractions in case λ is constant (but trimmed to the maximum value to satisfy the binding constraints outlined in the appendix). From this figure, we can see that we need to assume that more than 55% of the sexually active population need to misreport sexual activity in order to conclude that at least 95% of HIV cases in the considered population is due to sexual transmission. The figure document also the full range of possible population attributable fractions, depending on alternative assumptions on misreporting. If for example one wants to assume a moderate range for misreporting (between 0 and 10%),

the proportion of sexually transmitted HIV cases would be bounded between 33% and 48%. With a misreporting between 20 and 30%, the population attributable fraction would be bounded between 62% and 74%.

The results for the different countries are presented in the Table 2, which demonstrates that bounds for the proportion of sexual HIV transmission are not constant across the three countries. With moderate misreporting, the percentage of sexually transmitted HIV cases is bounded between 34% and 54% in Swaziland, but only between 15% and 22% in Lesotho. In a scenario, where more than half of the sexually active population misreports sexual activity, PAFs are bounded between 56% and 100% in Zimbabwe and 95% and 100% in Swaziland.

TABLE 2: Bound for the proportion of sexually transmitted HIV cases for different countries

	0-10%		20-30%		>50%	
	Lower	Upper	Lower	Upper	Lower	Upper
Lesotho	15%	22%	33%	48%	75%	100%
Swaziland	34%	54%	69%	80%	95%	100%
Zimbabwe	22%	26%	32%	39%	56%	100%

Note: Bounds are estimated based on Probit models that are estimated for each country.

So far it is assumed that all sexually active women have identical probability of misreporting. This may not be realistic since empirical studies have demonstrated that inconsistencies in reports are correlated with observable characteristics (Rodgers, Billy and Udry 1982). To investigate the sensitivity of the results to different assumptions regarding factors associated with misreporting, the paper uses different scenarios: (1) Misreporting could be associated with religions or cultural values, particularly if the tradition places a high value on virginity. The Catholic Church and certain evangelical Protestant churches, heavily promote premarital abstinence, which may not only have an effect on the decision to initiate a premarital partnership, but also on the likelihood to admit any premarital sexual relations. Population attributable fractions are therefore estimated assuming the probability of misreporting to be twice as high if the woman reports to be Christian (Catholic or Protestant). (2) Young women may be less mature about their own sexual behaviors and thus, may be more likely to misreport sexual activity. Thus, PAFs are estimated assuming the probability of misreporting to be twice as high if the woman is younger than 17. (3) The presence of a third person during an interview may prevent respondents getting honest answers. The third scenario therefore

assumes that women who were interviewed while another person was present misreport sexual activity twice as often. The results of these sensitivity checks are presented in the appendix (Figure A1). It can be seen that the results are not sensitive to the chosen scenario, as predicted PAFs are almost identical.

As a further sensitivity check, the Probit models are re-estimated using different sets of control variables. This analysis helps to understand, if the failure to control for all relevant factors causes a major bias in the result. Figure A2 demonstrates that the results presented in this paper are driven by age and regional characteristics. Population attributable fraction from a restricted model that controls only for age and regional characteristics are almost identical to the full model, but differ to the naïve model that does not control for any confounding variables. This shows that other variables have only small impact, and the failure to control for all possible confounders is likely to be small.

4. Conclusions

This paper proposes a method to empirically uncover the implicit assumptions underlying the discussion on "virgin" HIV cases and applies this method to data from Lesotho, Swaziland, and Zimbabwe. The results demonstrate that a large share of HIV infections remains unexplained by sexual HIV transmission if one assumes that women have correctly reported sexual activity. Misreporting can explain the high proportion of "virgin" HIV cases, but at the "cost" of data inaccuracy. A substantial proportion of unmarried sexually active women age 15 to 19 needs to have misreported sexual activity status (>55%) to be able to conclude that HIV is predominantly sexually transmitted in this population.

Two possible implications for research and policy arise from this study: (1) If one assumes that misreporting is moderate, non-sexual risk factors account for more HIV infections than previously assumed. This raises the question whether non-sexual risk factors are sufficiently covered by the current prevention paradigm and whether prevention interventions aimed to prevent non-sexual HIV need to be intensified. (2) The hypothesis that HIV is almost exclusively sexually transmitted implicitly requires that misreporting of sexual behavior is severe. If this is true, self-reported sexual behavior is an inappropriate measure for monitoring the epidemic.

Due to the lack of accurate information on the importance of misreporting as well as on exposures to alternative risk factors, one cannot predict the importance of different transmission modes. Better data can therefore greatly improve our current knowledge on the importance of different HIV transmission modes. Future surveys on HIV/AIDS in Sub Saharan Africa should collect detailed information on all relevant transmission modes, such as sexual behaviors, health care, and illicit drug use. Biomarker studies aimed to detect response bias should make sure that positive test result cannot be due to non-sexual transmission, which excludes HIV as a possible biomarker. Gaining better understanding is however, necessary to design combination prevention by combining different prevention approaches to achieve maximum impact on HIV prevention.

Appendix

When making assumptions on the probability of misreporting, two binding constraint need to be considered:

(C1) The individual probability of misreporting cannot exceed the probability to declare sexual activity:

$$\mathrm{P}(Z=1 \mid A=0,x) \leq 1 \quad \Rightarrow \lambda(x) \leq 1 - P(A=1 \mid x)$$

(C2) The individual expected HIV risk from non-sexual transmission modes need to be less than one:

$$E(Y^0 \mid x) \geq 0 \quad \Rightarrow P(Z=1 \mid A=0,x) \geq \frac{E(Y \mid A=0,x)}{E(Y \mid A=1,x)}$$

For $E(Y^0 \mid x) \leq 1$, it also needs to be the case that the conditional prevalence of people who report sexual activity is larger than the conditional prevalence of people who do not report sexual activity: $E(Y|A=1,x) \geq E(Y|A=0,x)$. However, since CIA and NDM assume that all relevant variables, which jointly affect HIV, sexual activity, and misreporting, are observed, this condition is automatically fulfilled as long as true sexual activity is positively associated with HIV acquisition. This is likely to be the case as HIV is a sexually transmitted virus, and condoms do not provide a 100% protection against infection.

Table A1: Descriptive statistics

	Full sample			Sexually inactive			Sexually active		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
HIV infection	3098	0.08	0.26	2318	0.04	0.20	780	0.14	0.34
Sexual activity	3098	0.34	0.47						
Age (omitted category: Age=15)									
Age = 16	3098	0.24	0.42	2318	0.28	0.45	780	0.14	0.35
Age = 17	3098	0.21	0.41	2318	0.20	0.40	780	0.24	0.43
Age = 18	3098	0.18	0.38	2318	0.14	0.34	780	0.27	0.44
Age = 19	3098	0.16	0.37	2318	0.10	0.30	780	0.28	0.45
Wealth (omitted category: Poorest)									
Poorer	3098	0.18	0.38	2318	0.17	0.37	780	0.19	0.39
Middle	3098	0.20	0.40	2318	0.20	0.40	780	0.22	0.41
Richer	3098	0.23	0.42	2318	0.24	0.43	780	0.23	0.42
Richest	3098	0.24	0.43	2318	0.25	0.43	780	0.21	0.41
Education (omitted category: none)									
Complete primary	3098	0.13	0.34	2318	0.13	0.33	780	0.14	0.35
Incomplete secondary	3098	0.54	0.50	2318	0.58	0.49	780	0.47	0.50
Complete secondary, higher	3098	0.04	0.20	2318	0.04	0.19	780	0.06	0.23
Occupation (Omitted category: not working)									
Agricultural	3098	0.02	0.15	2318	0.02	0.15	780	0.03	0.16
Household & Domestic	3098	0.02	0.14	2318	0.02	0.14	780	0.02	0.14
Other	3098	0.09	0.28	2318	0.05	0.21	780	0.17	0.37
Self reported barriers to access health care (big vs. small problem)									
Getting permission to go	3098	0.98	0.16	2318	0.97	0.17	780	0.99	0.12
Getting money needed	3098	0.73	0.44	2318	0.73	0.45	780	0.74	0.44

Distance to health facility	3098	0.74	0.44	2318	0.74	0.44	780	0.75	0.43
Having to take transport	3098	0.78	0.42	2318	0.77	0.42	780	0.79	0.41
Not wanting to go alone	3098	0.86	0.35	2318	0.84	0.37	780	0.89	0.31
No female health provider	3098	0.89	0.31	2318	0.90	0.30	780	0.88	0.32
Residence: Rural	3098	0.78	0.42	2318	0.78	0.41	780	0.77	0.42
Presence of other person during interview	3098	0.10	0.30	2318	0.09	0.29	780	0.11	0.32
Religion (omitted category: none, other)									
Roman Catholic	3098	0.12	0.32	2318	0.12	0.33	780	0.10	0.30
Protestant Churches	3098	0.61	0.49	2318	0.65	0.48	780	0.54	0.50

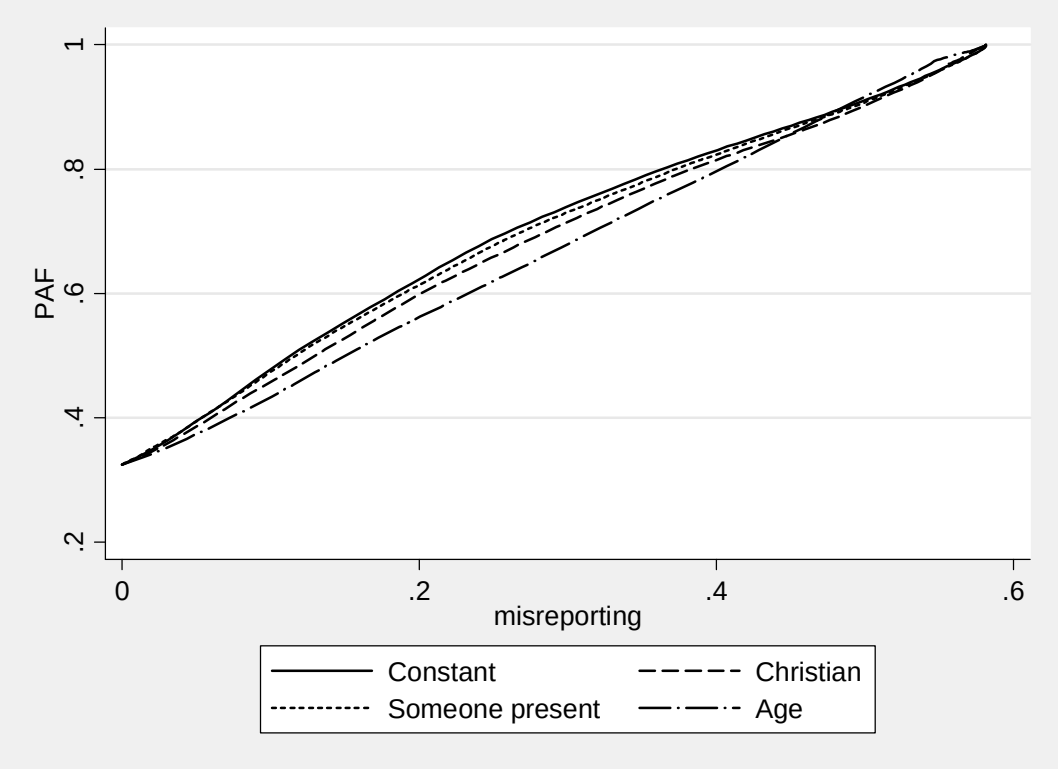
Note: Average values are estimated applying the adjusted sample weights.

Table A2: Regression results for the Probit models

	Sexual activity		HIV infection	
	Coef.	z	Coef.	z
Sexual activity			0.45 ***	4.09
Age (omitted category: Age=15)				
Age = 16	0.50 ***	4.19	-0.04	-0.26
Age = 17	1.09 ***	9.29	0.19	1.22
Age = 18	1.41 ***	11.62	0.21	1.37
Age = 19	1.66 ***	12.46	0.45 ***	2.76
Wealth (omitted category: Poorest)				
Poorer	0.33	1.39	0.78 ***	2.57
Middle	0.06	0.24	0.48	1.47
Richer	0.14	0.56	0.30	0.85
Richest	-0.03	-0.10	0.51	1.46
Education (omitted category: none)				
Complete primary	-0.19	-1.58	-0.12	-0.84
Incomplete secondary	-0.40 ***	-4.45	-0.33 ***	-2.76
Complete secondary, higher	-0.61 ***	-3.02	-0.66 **	-2.34
Occupation (Omitted category: not working)				
Agricultural	0.65 ***	2.90	0.26	0.98
Household & Domestic	0.27	1.21	0.47 **	2.06
Other	0.69 ***	5.20	0.29 *	1.94
Self reported barriers to access health care (big vs. small problem)				
Getting permission to go	0.26	1.09	0.63 ***	2.94
Getting money needed	-0.04	-0.43	-0.11	-0.89
Distance to health facility	0.09	0.84	0.03	0.23
Having to take transport	-0.03	-0.28	0.11	0.73
Not wanting to go alone	0.36 ***	2.71	-0.12	-0.77
No female health provider	-0.24 **	-2.01	0.12	0.68
Residence: Rural	-0.27 **	-2.36	-0.09	-0.59
Presence of other person during interview	0.01	0.12	-0.25	-1.45
Religion (omitted category: none, other)				
Roman Catholic	-0.34 **	-2.25	-0.63 ***	-3.16
Protestant Churches	-0.27 ***	-3.02	-0.17	-1.35
_cons	-1.28 ***	-3.09	-2.22 ***	-4.50
N		3098		3098
Pseudo R2		0.21		0.11

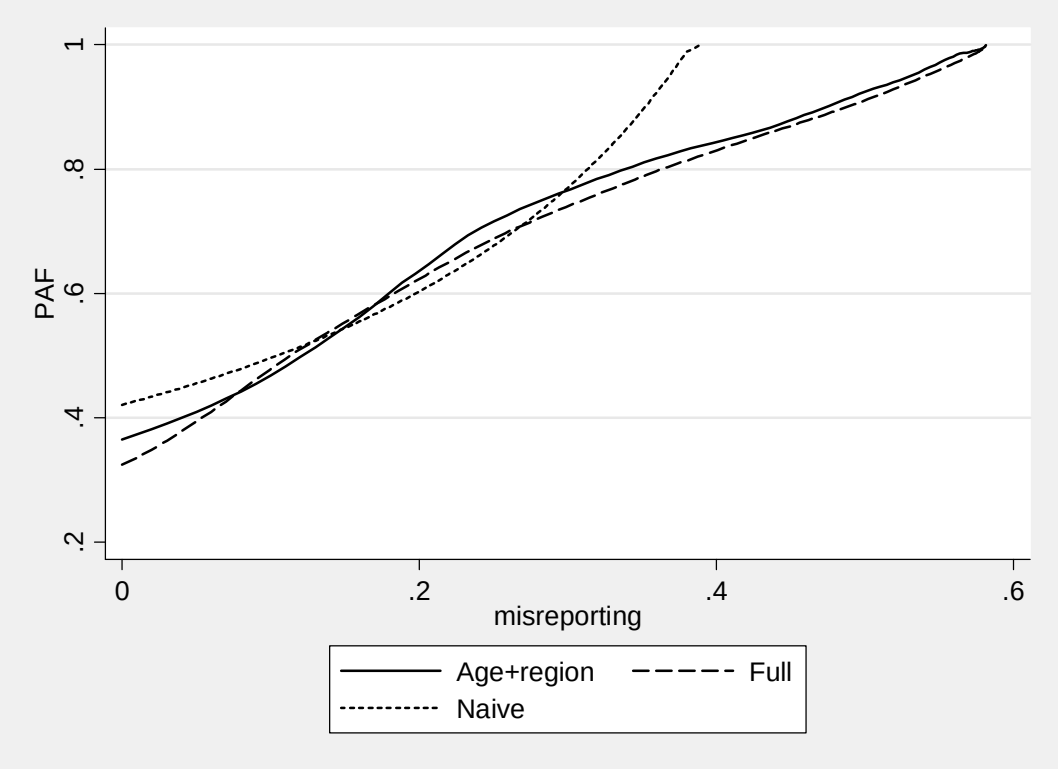
Note: Significance levels: *** p<0.01, ** p<0.05, * p<0.1; adjusted survey weights are applied, the regression controls for 24 regional dummy variables and interaction terms between wealth and country dummies (coefficients are not reported but available from the author on request)

Figure A1: Proportion of sexually transmitted HIV infections under different scenarios for factors associated with misreporting



Constant: λ is constant for all x, Age: λ is twice as high if age<17; *Someone present*: λ is twice as high if someone is present during the interview; *Christian*: λ is twice as high if respondent belongs to the Roman Catholic or Protestant Church.

Figure A2: Proportion of sexually transmitted HIV infections based on different set of control variable



Full: as in Table A2, *Age+region*: Age and regional dummy variables only, *Naïve*: no control variable.

REFERENCES

Aquilino, WS, DL Wright, and AJ Supple. "Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use." *Subst Use Misuse*, 2000: 845-67.

Awusabo-Asare, K, and SK Annim. "Wealth Status and Risky Sexual Behaviour in Ghana and Kenya." *Applied Health Economics and Health Policy*, 2008: 27-39.

Barnow, B. S., G. G. Cain, und A.S. Goldberger. «Issues in the Analysis of Selectivity Bias.» In *Evaluation Studies*, von E Stromsdorfer und G. Farkas. San Francisco: Sage, 1980.

Battistin, E, and S Sianesi. "Misclassified Treatment Status and Treatment Effects: An Application to Returns to Education in the UK." *Review of Economics and Statistics*, 2006, Forthcomming ed.

Beckerleg, S, M Telfer, and GL Hundt. "The rise of injecting drug use in East Africa: a case study from Kenya." *Harm Reduct J*, 2005: 12.

Bene, C, and Sonja Merten. "Women and Fish-for-Sex: Transactional Sex, HIV/AIDS and Gender in African Fisheries." *World Development*, 2008: 875-99.

Bongaarts, J, T Buettner, G Heilig, and F Pelletier. "Has the HIV Epidemic Peaked?" *Population and Development Review*, 2008: 199-224.

Brewer, DD, JJ Potterat, SQ Muth, and S Brody. "Converging evidence suggests nonsexual HIV transmission among adolescents in sub-Saharan Africa." *J Adolesc Health*, 2007: 290-1.

Buvé, A, et al. "Interpreting sexual behaviour data: validity issues in the multicentre study on factors determining the differential spread of HIV in four African cities." *AIDS*, 2001: S117-26.

Catania, JA, et al. "Risk factors for HIV and other sexually transmitted diseases and prevention practices among US heterosexual adults: changes from 1990 to 1992." *Am J Public Health*, 1995: 1492-9.

Chen, L, et al. "Sexual risk factors for HIV infection in early and advanced HIV epidemics in sub-Saharan Africa: systematic overview of 68 epidemiological studies." *PLoS ONE*, 2007: e1001.

Clark, S, J Bruce, and A Dude. "Protecting Young Women from HIV/AIDS: The Case Against Child and Adolescent Marriage." *International Family Planning Perspectives*, 2006: 79-88.

Cowan, FM, et al. "School based HIV prevention in Zimbabwe: feasibility and acceptability of evaluation trials using biological outcomes." *AIDS*, 2002: 1673-8.

de Boer, MA, DD Celentano, S Tovanabutra, S Rugpao, KE Nelson, and V Suriyanon. "Reliability of self-reported sexual behavior in human immunodeficiency virus (HIV) concordant and discordant heterosexual couples in northern Thailand." *Am J Epidemiol*, 1998: 1153-61.

de Walque, D. "Sero-discordant Couples in Five African Countries: Implications for Prevention Strategies." *Population and Development Review*, 2007: 501-23.

Eide, GE, and I Heuch. "Attributable fractions: fundamental concepts and their visualization." *Stat Methods Med Res*, 2001: 159-93.

Ezzati, M, et al. "Comparative Quantification of Mortality and Burden of Disease Attributable to Selected Risk Factors." In *Global Burden of Disease and Risk Factors*, by AD Lopez, CD Mathers, M Ezzati, Jamison DT and Murray CJ, 241-268. New York: Oxford University Press, 2006.

Gavin, L, et al. "Factors associated with HIV infection in adolescent females in Zimbabwe." *J Adolesc Health*, 2006.

Gersovitz, M. "The HIV Epidemic in Four African Countries Seen through the Demographic and Health Surveys." *Journal of African Economies*, 2005: 191-246.

Gisselquist, D, R Rothenberg, J Potterat, and E Drucker. "HIV infections in Sub Saharan Africa not explained by sexual or vertical transmission." *Int J STD AIDS*, 2002, 13 ed.: 657-66.

Glick, P, and DE Sahn. "Africans Practicing Safer Sex? Evidence from Demographic and Health Surveys for Eight Countries." *Economic Development and Cultural Change*, 2008: 397-439.

GSS. *Ghana Service Assessment Survey 2002.* Final Report, Calverton, Maryland: Ghana Statistical Service, and ORC Macro, 2003.

Hjortsberg, C. "Why do the sick not utilise health care? The case of Zambia." *Health Econ*, 2003: 755-70.

Hutin, YJ, AM Hauri, and GL Armstrong. "Use of injections in healthcare settings worldwide, 2000: literature review and regional estimates." *BMJ*, 2003: 1075.

Johnson, EO, CG Schütz, JC Anthony, and ME Ensminger. "Inhalants to heroin: a prospective analysis from adolescence to adulthood." *Drug Alcohol Depend*, 1995: 159-64.

Kloos, H, and DH Mariam. "Some neglected and emerging factors in HIV transmission in Ethiopia." *Ethiop Med J*, 2007: 103-7.

Luke, N. "Exchange and condom use in informal sexual relationships in urban Kenya." *Exchange and condom use in informal sexual relationships in urban Kenya.*, 2006: 319-48.

Marston, M, B Zaba, JA Salomon, H Brahmbhatt, and D Bagenda. "Estimating the net effect of HIV on child mortality in African populations affected by generalized HIV epidemics." *J Acquir Immune Defic Syndr*, 2005: 219-27.

McCurdy, S, GP Kilonzo, M Williams, and S Kaaya. "Harm reduction in Tanzania: an urgent need for multisectoral intervention." *Int J Drug Policy*, 2007: 155-9.

McCurdy, SA, MW Ross, GP Kilonzo, MT Leshabari, and ML Williams. "HIV/AIDS and injection drug use in the neighborhoods of Dar es Salaam, Tanzania." *Drug Alcohol Depend*, 2006: S23-7.

Mensch, BS, MJ Grant, and AK Blanc. "The Changing Context of Sexual Initiation in sub-Saharan Africa." *Population and Development Review*, 2006: 699-727.

Mensch, BS, PC Hewett, R Gregory, and S Helleringer. *Sexual behavior and STI/HIV status among adolescents in rural Malawi: An evaluation of the effect of interview mode on reporting.* Poverty, Gender, and Youth Working Paper No. 8., New York: Population Council, 2008.

Mishra, V, et al. "HIV testing in national population-based surveys: experience from the Demographic and Health Surveys." *Bull World Health Organ*, 2006: 537-45.

MoHR. *Rwanda Service Assessment Survey 2004.* Final Report, Calverton, Maryland: National Population Office, and ORC Macro, 2003.

Oster, E. "Sexually Transmitted Infections, Sexual Behavior, and the HIV/AIDS Epidemic." *Quarterly Journal of Economics*, 2005: 467-515.

Pantaleo, G, C Grazios, and AS Fauci. "New concepts in the immunopathogenesis of human immunodeficiency virus infection." *N Engl J Med*, 1993: 327-35.

Poulton, R, et al. "Association between children's experience of socioeconomic disadvantage and adult health: a life-course study." *Lancet*, 2002: 1640-5.

Reinherz, HZ, RM Giaconia, AM Hauf, MS Wasserman, and AD. Paradis. "General and specific childhood risk factors for depression and drug disorders by early adulthood." *J Am Acad Child Adolesc Psychiatry*, 2000: 223-31.

Rockhill, B, B Newman, and C Weinberg. "Use and misuse of population attributable fractions." *Am J Public Health*, 1998: 15-9.

Rodgers, JL, JO Billy, and JR Udry. "The rescission of behaviors: Inconsistent responses in adolescent sexuality data." *Social Sci. Res*, 1982: 280-96.

Rosenbaum, P., und D. Rubin. «The Central Role of the Propensity Score in Observational Studies for Causal Effects.» *Biometrika*, 1983: 41-55.

Rothenberg, RB, M Scarlet, C del Rio, D Reznik, und C. O'Daniels. «Oral transmission of HIV.» *AIDS*, 1998: 2095-2105.

Rutstein, SO, and J Johnson. *The DHS Wealth Index.* DHS Comparative Reports No. 6, Calverton, MD: ORC Macro, 2004.

Sahn, D, S Younger, and G Genicot. "The Demand for Health Care Services in Rural Tanzania." *Oxford Bulletin of Economics and Statistics*, 2003: 241-60.

Sanders, SA, und JM Reinisch. «Would you say you "had sex" if...?» *JAMA*, 1999: 275-277.

Schmid, GP, et al. "Transmission of HIV-1 infection in sub-Saharan Africa and effect of elimination of unsafe injections." *Lancet*, 2004: 482-8.

Schneider, WH, and E Drucker. "Blood transfusions in the early years of AIDS in sub-Saharan Africa." *Am J Public Health*, 2006: 984-94.

Smith, DJ. "Premarital sex, procreation, and HIV risk in Nigeria." *Stud Fam Plann*, 2004: 223-35.

Smith, TW. «Discrepancies between men and women in reporting number of sexual partners: a summary from four countries.» *Soc Biol*, 1992: 203-11.

Stronski, SM, M Ireland, P Michaud, F Narring, and MD Resnick. "Protective correlates of stages in adolescent substance use: a Swiss National Study." *J Adolesc Health*, 2000: 420-7.

UNAIDS. *UNAIDS Practical Guidelines for Intensifying HIV Prevention Towards Universal Access.* UNAIDS/07.07E / JC1274E, Geneva: Joint United Nations Programme on HIV/AIDS, 2007.

WHO. *Blood Safety - Strategy for the African Region.* ADR/RC51/9 Rev1, Brazzaville; Regional Office for Africa : World Health Organization, 2002.

WHO. *Quality of care; patient safety.* World Health Assembly Resolution WHA55.18, Geneva: World Health Organization, 2002.

Wilson, S, N Brown, C Mejia, and P Lavori. "Effects of Interviewer Characteristics on Reported Sexual Behavior of California Latino Couples." *Hispanic Journal of Behavioral Sciences*, 2002: 38-62.

Zaba, B, E Pisani, E Slaymaker, and JT Boerma. "Age at first sex: understanding recent trends in African demographic surveys." *Sex Transm Infect*, 2004: ii28-35.