NBER WORKING PAPER SERIES

HIV STATUS AND LABOR MARKET PARTICIPATION IN SOUTH AFRICA

James A. Levinsohn
Zoë McLaren
Olive Shisana
Khangelani Zuma

Working Paper 16901
http://www.nber.org/papers/w16901

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2011

Thanks to John DiNardo, Taryn Dinkelman, Justin McCrary, Nicoli Natrass, Jeff Smith, and Duncan Thomas for especially helpful comments and suggestions. Levinsohn acknowledges support from the NICHD. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

HIV Status and Labor Market Participation in South Africa
James A. Levinsohn, Zoë McLaren, Olive Shisana, and Khangelani Zuma
NBER Working Paper No. 16901
March 2011
JEL No. O12

ABSTRACT

Because individuals with HIV are more likely to fall into poverty, and the poor may be at higher risk of contracting HIV, simple estimates of the effect of HIV status on economic outcomes will tend to be biased. In this paper, we use two econometric methods based on the propensity score to estimate the causal effect of HIV status on employment outcomes in South Africa. We rely on rich data on sexual behavior and knowledge of HIV from a large national household-based survey, which included HIV testing, to control for systematic differences between HIV-positive and HIV-negative individuals. This paper provides the first nationally representative estimates of the impact of HIV status on labor market outcomes for southern Africa. We find that being HIV-positive is associated with a 6 to 7 percentage point increase in the likelihood of being unemployed. South Africans with less than a high school education are 10 to 11 percentage points more likely to be unemployed if they are HIV-positive. Despite high unemployment rates, being HIV-positive confers a disadvantage and reinforces existing inequalities in South Africa.

James A. Levinsohn
Yale School of Management
PO Box 208200
New Haven, CT 06520
and NBER
James.Levinsohn@yale.edu

Zoë McLaren
School of Public Health
University of Michigan
Ann Arbor, MI 48109
zmclaren@umich.edu

Olive Shisana
Health Sciences Research Council
Pretoria, South Africa
oshisana@hsrc.ac.za

Khangelani Zuma
Health Sciences Research Council
Pretoria, South Africa
kzuma@hsrc.ac.za

1 Introduction

This paper employs an especially rich South African data set to estimate the causal impact of HIV status on an individual's labor market outcomes. The paper contributes to a better understanding of the economic impact of HIV/AIDS and the role HIV/AIDS plays in reinforcing inequality in South Africa.

The existence of an impact of HIV on labor market participation is hotly debated in South Africa. Some argue that HIV is a severe constraint on current economic growth because those who are too ill to work represent lost economic output. Others argue that because unemployment is so high – in the 30 to 40 percent range according to the broad definition which includes individuals who desired employment but had no job search activity within the past month (i.e. discouraged workers) – HIV has minimal economic impact today. Proponents of this view note that the sex-age cohorts with the highest unemployment are essentially the same as those with the highest HIV prevalence rates. Our estimates speak directly to this issue. This paper provides the first nationally representative estimates of the impact of HIV on labor market status, certainly for southern Africa but probably also more broadly for non-rich countries.

There are at least two reasons for the paucity or outright lack of evidence– stringent data requirements and econometric challenges.

Obtaining estimates of the impact of HIV on labor market participation clearly requires data on both labor market participation (which is generally available) and on an individual's HIV status (which is collected fairly rarely).[1] Our econometric methods also require extensive data on correlates of an individual's HIV status.

Even with our rich data set, econometric challenges arise due to econometric endogeneity.

In our context, endogeneity arises because an individual's labor market status might affect their HIV status. For example, migrating for employment opportunities may put individuals at greater risk for HIV (reverse causality bias). Alternatively, there may be factors that affect both HIV status and labor market status, such as an individual's age or their willingness to invest in the future (selection bias, omitted variable bias).[2]

Two of the standard approaches to addressing these issues are infeasible in this instance. The traditional solution is to apply an instrumental variables approach. For the application at hand, this would require an instrument that is correlated with the individual-level HIV status and is orthogonal to shocks to labor market status (conditional on observed covariates.) This is a tall order to fill. One class of correlates with HIV status – socioeconomic variables such as income, education, sex, age and even household size – are likely correlated with labor market status, while another class of correlates – health-related variables such as sexual practices and knowledge about HIV – are themselves likely to be correlated with the socioeconomic covariates and, by association, with labor market status.[3] (A closely related approach uses a control function with an exclusion restriction, and the same concerns that preclude good instruments similarly preclude a convincing exclusion restriction. See Heckman and Navarro-Lozano (2004).)

The other approach to addressing sample selection which has gained currency in development economics is to conduct interventions such that the bias vanishes by design.[4] In a study of HIV status and labor market participation, this would involve randomly assigning HIV status to individuals and comparing labor market status across the two groups. This is one context for which the experimental approach is impossible.

We turn to methods based on the propensity score to address the issues of selection and reverse causality. We investigate the impact of HIV status on labor market participation

using two methods– propensity score reweighting and control functions.[5] While both have been used to address the endogeneity issue in the literature, there have been few direct comparisons of multiple estimators for the same problem. We use both methods as a marginal methodological contribution and, more importantly, as a robustness check.

Propensity score reweighting and the control function approach both rely on the assumption that sample selection into being HIV+ is random, conditional on attributes that are observable to the econometrician (selection on observables). Given the richness of the data set we use, this assumption is weaker than it might be under other circumstances. The survey questionnaire included over 175 questions, many with multiple sub-sections, yielding over four hundred individual covariates. The data set contains detailed information on sexual practices and knowledge of HIV transmission in addition to household characteristics and other factors that might influence HIV status. We examine the validity of the assumption of selection on observables in section 5 by testing whether a parsimonious specification succeeds in balancing variables not used in the estimation (balancing tests) and experimenting with multiple specifications to confirm that the inclusion of additional controls does not change the results (robustness checks).

In the next section, we briefly discuss the related literature. Section 3 discusses the data that are employed. In section 4, the two approaches to estimation are presented. Section 5 presents results, while section 6 provides both caveats and conclusions.

2 Related Literature

While understanding the relationship between HIV status and labor market participation is important from multiple perspectives, there are few studies that present evidence on the impact of HIV and fewer still that address issues of endogeneity.

A number of studies develop macroeconomic simulation models to examine the effect of HIV/AIDS on economic growth, however these models may be extremely sensitive to assumptions about life-expectancy for HIV+ individuals (see Cuddington and Hancock (1994), Kambou, Devarajan and Over (1992), and Arndt and Lewis (2000)). Instead of assuming an HIV/AIDS mortality rate, Bloom and Mahal (1997) estimate the impact of AIDS on growth by exploiting cross-country variation in HIV prevalence rates. They find that HIV has an insignificant effect on per capita GDP growth for the sample of 51 countries.

The HIV pandemic not only affects current economic growth, but can have lasting effects on growth rates into the future. Kalemli-Ozcan (2006) and Bell, Devarajan and Gersbach (2003) model the impact of current HIV prevalence rates on future economic growth. Because HIV+ parents are likely to die before reaching old age, they may invest less in their children's human capital acquisition, lowering the stock of human capital and contributing to lower growth rates in the future. On the other hand, if the AIDS epidemic causes a reduction in fertility that dominates this human capital effect, higher future living standards may result (Young (2005)).

Most papers that examine the impact of HIV on economic growth look exclusively at the effect of AIDS mortality, overlooking the effects of the illness on employment and productivity. There are some exceptions. Murray, Sonnenberg, Nelson, Shearer, Bester, Begley and Glynn (2005) find that the rate of minor work-related injuries was 30 percent higher for HIV+ miners in South Africa than for HIV- miners. The correlation was observable within one year of seroconversion which suggests that the acute initial infection or the psychological shock has an effect long before AIDS symptoms appear. This is likely an underestimate of the impact of HIV because miners who are most affected are more likely to take on easier tasks at work or leave employment altogether.

Habyarimana, Mbakile and Pop-Eleches (2010) use ARV therapy inception date as an instrument for health status to examine the effect of health on productivity for mine workers in Botswana. They document an inverse-V-shaped pattern of absenteeism for HIV+ workers in the two years around the inception of anti-retroviral (ARV) therapy. Workers who subsequently enrolled in ARV therapy missed about five times as many days of work as non-enrolled workers in the year prior to ARV therapy inception, but absenteeism rates returned to pre-peak levels after a year of therapy. Non-enrolled workers appear to be a valid control group because the two groups had similar levels of absenteeism from 5 years to 1 year prior to therapy inception. These findings suggest that health can have large effects on employment outcomes, and that ARV therapy is effective in reducing the disparity in productivity between HIV+ and HIV- workers.[6] Fox, Rosen, MacLeod, Wasunna, Bii, Foglia and Simon (2004) examine differences in on-the-job productivity between workers who subsequently died of AIDS and other workers on tea plantations in Kenya. They find that in the last year before death, AIDS victims are less productive (in this case, measured by the quantity of tea leaves picked), are more likely to be reassigned to less strenuous but less lucrative tasks and are more often absent from work.

One way to estimate the causal effect of HIV status on employment is to use a plausibly exogenous instrument for HIV status. Variation in circumcision has been used to instrument for HIV status because circumcision has been found to be associated with a reduced risk of HIV both in regression analyses (Weiss, Quigley and Hayes (2000)) and in randomized controlled trials (Auvert, Taljaard, Lagarde, Sobngwi-Tambekou, Sitta and Puren (2005)). Werker, Ahuja and Wendell (2006) find that HIV/AIDS did not have a measurable effect on economic growth, savings, or fertility behavior in African countries but that there was weak evidence that HIV/AIDS reduced youth literacy, and increased malnutrition. McKelvey (2007) finds that across nine African countries and Haiti, HIV+

individuals are significantly less likely to have been employed or to have earned enough to contribute more than half of household expenditures. However, specification checks using two populations that would not be expected to benefit from circumcision (men under 20 years old and those who have never had sex) suggest that unobservable differences that vary with circumcision may be driving the results.

Transitory economic shocks can have short-term effects on an individual's propensity to engage in risky behavior, but potentially long-term effects on health. This is one avenue through which employment outcomes may impact HIV status, and one reason why reverse causality is a salient issue. Women who are economically vulnerable may become active with multiple sexual partners (Dinkelman, Lam and Leibbrandt (2007)), or turn to sex-for-gift exchanges to smooth consumption (Dunkle, Jewkes, Brown, Gray, McIntyre and Harlow (2004); LeClerc-Madlala (2002)). Migration is another potential response to transitory shocks that can put both men and women at greater risk for HIV (Zuma, Lurie, Williams, Mkaya-Mwamburi, Garnett and Sturm (2005)).

3 Data

Our data come from the nationally-representative South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey (SABSSM II) conducted in 2005 by the Human Sciences Research Council (HSRC), the Centre for AIDS Development, Research and Evaluation (CADRE) and the Medical Research Council.[7] The survey asked adult respondents questions about demographics, knowledge of HIV, sexual history, knowledge of voluntary counseling and testing (VCT) services, health, mental health, and drug and alcohol use. It also included a household module that asked for basic demographic data for all household members in addition to questions about household infrastructure

and participation in government programs. While there were no questions directly addressing income or expenditure, the individual survey did query labor market participation. Respondents were asked to classify their "present employment situation" as one of the following thirteen categories: homemaker not looking for work, homemaker looking for work, unemployed not looking for work, unemployed looking for work, informal sector not looking for permanent work, old age pensioner, sick/disabled and unable to work, student/pupil/learner, self-employed full time (40 or more hours per week), self-employed part time (less than 40 hours per week), employed part time (if none of the above), employed full time or other.

The sample consisted of 23,275 individuals in 10,584 households.[8] Not everyone in a household was sampled. Using the roster of all household members (including those who usually live in the household, whether or not they were present at the first interviewer visit, and any guests that stayed in the household the previous night), field workers randomly selected at most one person from each of three age groups (2-14 years, 15-24 years and 25 and up) to be interviewed. While this sampling strategy is an efficient way to obtain a measure of HIV prevalence, it is problematic for any analysis of household dynamics. For example, we are unable to directly measure the impact of having an HIV+ spouse on adult labor supply, since only one household member above the age of 25 was surveyed.

The survey also included an opt-out HIV test for respondents age 2 and older. The response rate for testing was 65.4 percent overall, and 73.3 percent in the adult sample (age 15 and older) used for analysis (see Table 1). HIV incidence data was collected from all HIV+ specimens using an enzyme immunoassay that measures the ratio of HIV antibodies to other antibodies to determine the elapsed time since HIV infection.

Table 2 presents HIV prevalence rates by race. Table 2 indicates that while HIV impacts

all races, Africans have, by far, the highest prevalence rates. In our sample, 17.34 percent of Africans are HIV+ while the population group with the next highest prevalence rate, Coloured, has a rate of 2.73 percent.[9]

For the remainder of this paper, we restrict our analysis to Africans. Africans comprise about 80 percent of the population. The HIV prevalence rate, the employment rate, the age profile of employment, and the education profile of employment are each markedly different for Africans.[10]

Figure 1 presents the age profile of HIV prevalence– a profile broadly comparable to those of other African countries. HIV prevalence peaks between age 25-30 for women and between 30-40 for men. The age profile for women is about 5 percentage points lower than UNAIDS ante-natal clinic statistics, which may be due to differences in the sampling frame (UNAIDS (2005)).

Table 3 reports employment status by HIV status separately for men and for women. For our analysis, individuals were assigned to a labor market status using the broad definition of unemployment, which includes discouraged workers. Respondents were classified as unemployed if they report that they are unemployed (either looking for work or not) or if they are a housewife or homemaker who is looking for work. They were classified as not economically active if they reported being a student, an old age pensioner, too sick or disabled to work, or a housewife or homemaker who is *not* looking for work. Otherwise, they were classified as employed.[11] Our main set of results excludes individuals who are not economically active and we explore the sensitivity of our results to this restriction in Appendix Table 2. A higher proportion of HIV+ individuals are unemployed relative to the employed, and this is especially pronounced for women. Almost 57 percent of HIV+ women are unemployed. Students and other not economically active individuals (mostly

old age pensioners) have lower than average prevalence rates. A similar proportion of each labor market group refused the HIV test.

4 Methodologies

4.1 Overview

Generating a plausible counterfactual is the core challenge to identifying a causal effect of HIV. For the situation at hand, this approach entails creating a counterfactual in which one could compare individuals who were virtually identical except for their HIV status and then compare differences in labor market status. The following subsections present two methods, each based on the propensity score, to generate unbiased estimates of the effect of being HIV+ on labor market status.

Our "treatment" (as the public health and program evaluation literatures refer to it) is HIV status, denoted by D, where $D = 1$ is HIV+ and $D = 0$ is HIV-. If HIV status were independent of the untreated value of our outcome of interest, labor market status (denoted Y_0), then there is no sample selection problem and we can use a simple "naive" estimator, a simple difference in means, to estimate the treatment effect. However, HIV status is not randomly assigned and is in all likelihood related to our outcomes of interest. We make the conditional independence assumption (CIA) that the untreated value of labor market status is independent of HIV status conditional on a vector of covariates (denoted X) :

$$Y_0 \perp\!\!\!\perp D|X. \tag{1}$$

4.2 The Propensity Score

In this context, the propensity score, denoted $P(HIV+)|X \equiv p(X)$, is the probability that an individual is HIV+ conditional on a vector of observable exogenous covariates (X). These covariates must be exogenous to HIV status in that they cannot be affected by HIV status (i.e. it would be inappropriate to use sexual behavior that may be changed if HIV status is known).[12] Rosenbaum and Rubin (1983) show that if the CIA is satisfied by conditioning on the vector X, it is satisfied by conditioning on the propensity score (i.e. the propensity score is a sufficient statistic for the X vector in the CIA). Hence, the propensity score makes the problem of finding a comparable control group tractable by reducing the dimensionality of the comparison, while still satisfying the CIA.

4.3 Propensity Score Reweighting

Propensity score reweighting uses the propensity score to create a counterfactual distribution of X in the HIV- (control) population so as to match the distribution of X in the HIV+ population. Essentially, HIV- observations with X characteristics that are most like HIV+ observations (i.e. that have a high $\hat{p}(X)$) receive the most weight, whereas HIV- observations that are very different from the HIV+ population receive less weight.

More formally, Dehejia and Wahba (1997) and DiNardo, Fortin and Lemieux (1996) show that

$$\Delta^{ATET} = \frac{1}{N^T} \Sigma_{i=1}^{N^T+N^C} \left(D_i Y_i - (1-D_i) Y_i \frac{p(X_i)}{1-p(X_i)} \right) \tag{2}$$

is a consistent estimator for the average treatment effect on the treated (ATET), provided the CIA and the common support condition hold. In the calculation, HIV+ observations receive a weight of 1 (the first term in parentheses) and HIV- observations are weighted

with $\omega_i = \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}$ (the second term in parentheses). This estimator does not impose a functional form on the relationship between HIV status and the outcome of interest (Dehejia and Wahba (1999)).

This approach requires overlap in the support of the propensity scores for the HIV+ and HIV- groups (the common support condition). In the Results section, we present density plots to verify the validity of this assumption.[13] Busso, DiNardo and McCrary (2010) show that reweighting outperforms propensity score matching in settings likely to be encountered in empirical work.

4.4 Control Functions

The control function approach is an alternative econometric strategy for addressing selection bias. The reason one obtains biased estimates from a simple regression of labor market status on HIV status is that the disturbance term in such a regression is correlated with HIV status (the independent variable). The essence of the control function approach is to control for the portion of the disturbance term that is correlated with HIV status. Once the portion of the disturbance term that is responsible for the correlation is expunged, the new error term is uncorrelated with HIV status, and the regression yields unbiased estimates of the impact of HIV on employment status. The control function approach was developed in Heckman and Robb (1985) and has been used to estimate the impact of training on earnings (Heckman and Hotz (1989)), the returns to education (Card (1999)) and the capitalization of pollution into housing values (Chay and Greenstone (2005)) among other applications. As with propensity score reweighting, we maintain the assumption of selection on observables.[14]

The data generating process is given by:

$$Y = \Lambda\Big(\beta_0 + \gamma D + f(X) + \epsilon\Big) \tag{3}$$

where $f(X)$ is a function of observables X.

Under the assumption of selection on observables, conditioning on $f(X)$ results in a disturbance term, ϵ, that is independent of D (HIV status) and hence, the estimate of the parameter of interest, γ is unbiased. In practice, a polynomial in the estimated propensity score as well as linear terms in X are used to flexibly model $f(X)$. We estimate:

$$Y = \Lambda\Big(\beta_0 + \gamma D + X'\phi + \sum_{i=1}^{k} \beta_i \hat{p}(X)^i + \epsilon\Big) \tag{4}$$

separately for HIV- and HIV+ groups and obtain predicted values, $\hat{Y}_0$ and $\hat{Y}_1$ respectively, from each regression. We calculate the ATET by averaging the difference in predicted values across HIV+ observations.

5 Results

Estimating the impact of HIV status on labor market participation is a three-step process. The first step is to estimate the propensity score. The second step is to empirically examine the validity of the CIA and the common support condition to ensure that the appropriate observable variables are included in the propensity score regression. If the underlying assumptions hold, then one can proceed to the third step, estimating the impact of HIV on labor market participation using propensity score reweighting and a control function approach.

5.1 The Propensity Score

There are competing philosophies behind what constitutes a properly specified propensity score regression. One approach is to adopt a relatively parsimonious specification, albeit one still rich enough to plausibly satisfy the CIA, while another approach is to include most all plausible regressors. We adopt the former, but experiment with the latter in sensitivity analyses.

The plausibility of the selection on observables assumption clearly rests on having data that can, in our context, account for selection into HIV status. The SABSSM II data set has several *hundred* variables for most respondents. In addition to the usual demographic information, the survey also collected extensive information on sexual practices and knowledge about HIV transmission, which is exactly the type of information needed to account for selection into HIV status. Information on either sexual practices or knowledge about HIV transmission can be misleading, so it is important to have information on *both* for predicting HIV status. For example, a respondent who has multiple partners but was well informed on how HIV is transmitted before their sexual debut and therefore practices safe sex may have a low probability of being HIV+. Similarly, a respondent who knows very little about how HIV is transmitted but is abstinent will have a low likelihood of being HIV+. Conditional on comprehensive information about an individual's sexual practices and knowledge regarding HIV transmission, unobservables such as attitudes towards risk or moral beliefs may have little explanatory power. Hence, the selection on observables assumption seems especially appropriate given the specifics of our data. The selection on observables assumption is testable, and we investigate the reasonableness (or not) of this assumption by conducting balancing tests on covariates in the propensity score specification as well as plausible covariates that were not included in the specification (see Section

5.2).

We estimate the propensity score as the predicted value of a logit regression. Table 4 reports the estimated coefficients (not marginal effects) and standard errors from our base case. The covariates include basic demographic characteristics, educational attainment, information about a respondent's sexual debut, knowledge of HIV transmission, and whether there is a pensioner living in the household.[15] We do not discuss the estimated coefficients because the main focus of this paper is on using the propensity score to correct for selection bias rather than on the correlates of seropositivity. Furthermore, it is quite difficult to have much intuition about the marginal impact of a single regressor conditional on the other 44.[16]

5.2 Examining the Validity of the Propensity Score Method

We next examine the validity of the CIA and common support assumption under our preferred (base case) propensity score specification. The former is done with balancing tests and the latter by examining empirical distributions.

The intuition behind balancing tests is appealingly clear. The idea of propensity score reweighting is to reweight the distribution of the observables (X's) of the HIV- population so as to match the distribution of the X's for the HIV+ population (see Equation 2). Balancing tests simply examine whether the difference in means of X's between HIV+ and HIV- populations is reduced when the observations are reweighted. If the reweighted means are similar (i.e. not significantly different) for HIV+ and HIV- populations, then the reweighting has achieved its goal, the data are balanced, and the CIA is appropriate.

Table 5 reports the results of balancing for those variables that enter the propensity score regression (i.e. internal balancing). We also conducted balancing tests for each of the indi-

cator variables for missing data, but these are not reported in the table. All were balanced. The results in Table 5 apply to the entire sample. Balancing tests were also conducted for each of the four subsamples with comparable results. Row 1 of Table 5 presents the results for age; HIV+ individuals were on average 2.5 years younger than the HIV- population. This difference had a t-statistic of -4.86 so the difference was highly significant. After reweighting, the difference falls to 0.255 years and is not significantly different from zero. None of the variables in the propensity score regression have a statistically significant difference between the HIV+ and HIV- populations in the reweighted data.

The propensity score weight, by design, attempts to minimize differences between HIV+ and HIV- groups for the variables included in the propensity score estimation (the X vector). A more stringent balancing test criterion is whether the propensity score weight also succeeds in balancing covariates that were not used in the estimation. External balancing tests were conducted on 56 variables. These external variables included measures of sexual activity and almost 50 variables that are plausibly related to economic well-being (e.g. source of water, type of cooking fuel, type of toilet, and measures of privation.) In the unweighted data, 20 of these variables had means that were significantly different (at the 95 percent significance level) between the HIV+ and HIV- populations. After reweighting, only three of those differences were still significant.

Based on internal and external balancing tests, we conclude that the propensity score specification is adequately reweighting the data to justify the CIA. While selection on observables is a strong assumption, we find that the richness of our data provides support for the assumption.

The balancing tests provide support for the CIA. The other assumption underlying the propensity score reweighting approach is the common support condition.[17] We examine

the appropriateness of the support condition by comparing the empirical distributions of the propensity scores for the HIV+ and HIV- populations. These distributions are shown in Figure 2. In that figure, it is clear that the densities have a common support.

5.3 The Impact of HIV

5.3.1 Estimates

Table 6 presents estimates of the causal impact of HIV on labor market participation. The table is organized such that each column presents estimates resulting from a different estimator and each row presents estimates using different samples of the data.

The first column presents naïve estimates of the impact of HIV status on labor market participation. Estimates in this column are simply the (unconditional) difference in the mean employment status for the HIV+ and HIV- populations, without the inclusion of any controls. Variables are defined such that the 0.079 figure in the first cell implies that, on average, HIV+ individuals are 7.9 percentage points more likely to be unemployed.

The second column presents the coefficient on HIV status when the variables in the propensity score (listed in Table 4) are included as controls in a simple logit regression of employment status on HIV status. This is a simple regression to run and is the same as the control function approach but it excludes the higher order terms of the propensity score. The logit result implies that for the full sample being HIV+ lowers the probability of employment by 6.5 percentage points.

The third and fourth columns present estimates of the causal impact of HIV on employment using propensity score reweighting and the control function approaches, respectively. For internal consistency of our results, both the propensity score weights and the control

function are based on the propensity score which is calculated using the set of control variables included in the logit in column 2. Each parameter estimate in columns 3 and 4 is the marginal effect of HIV status from a logit regression. For column 3, it is a simple logit regression of labor market status on HIV status using propensity score reweighting while in column 4, the reported estimate is the coefficient on HIV status (γ) in the regression given by Equation 4. Using the entire sample, propensity score reweighting indicates that being HIV+ raises the probability of unemployment by 6.2 percentage points while the control function approach indicates an increase of 7.0 percentage points. Each of these impacts is precisely estimated.[18] Our estimates of the causal impact of HIV on employment imply that, all else equal, being HIV+ raises the probability of unemployment by between 6 and 7 percentage points.

The estimates in the first row of Table 6 apply to the entire sample and, as such, potentially hide substantial underlying heterogeneity in the impact of HIV on labor market status. We investigate this heterogeneity in the remainder of Table 6 by restricting the sample to particular sub-populations. The second through seventh rows in Table 6 present results obtained from an analysis of sub-samples of the data, using propensity score estimates calculated within the sub-sample alone. For men, propensity score reweighting and the control function approach give estimates of 0.061 and 0.080, respectively. The former is not precisely estimated while the latter still is. For females, the point estimates from propensity score reweighting and the control function approach are 0.055 and 0.060. These are slightly lower than the estimates obtained with the entire sample.

The next two rows in Table 6 restrict the analysis to respondents aged 25 and older, and divide the sample between individuals whose education level is a Matric or higher and those whose education level is less than a Matric. Respondents under 25 may not have completed their schooling. A Matric is about equivalent to a high school education. The

message here is clear: the impact of HIV on labor market status is severe for those with lower levels of education and is negligible for those with higher levels. The causal impact of HIV status on labor market status for those with less than a Matric is an increase in the likelihood of unemployment of 10 percentage points (with propensity score reweighting) and 10.9 percentage points (with the control function approach.) These are large and precisely estimated impacts. We repeated the analysis for subgroups defined by race as well as by groups defined by education and sex, age, and education and age. These results are not reported here, but are available upon request from the authors. The results in Table 6 capture the gist of these divisions. Women, and especially women with lower levels of education, experience larger causal impacts of HIV on labor market status.

The last two rows highlight urban/rural differences in the impact of HIV on unemployment. Although both HIV and unemployment are more prevalent in rural areas, the causal impact of HIV on unemployment is larger in urban areas where being HIV+ lowers the probability of employment by between 7 and 8 percentage points.

5.3.2 Discussion

A concern about the role of selection bias motivated our choice of methodologies. Comparing the naïve estimates in column 1 of Table 6 to the causal estimates in columns 3 and 4 speaks to this issue. The naïve estimates are generally two to three percentage points higher than the causal estimates, and this highlights the importance of selection. Our results are consistent with the hypothesis that individuals who are HIV+ are more likely to be unemployed than the average South African, irrespective of their HIV status.

Because this study is probably the first to examine the causal impact of HIV on employment outcomes, it is difficult to place the magnitude of the estimated impact in context.

There are no other estimates available for comparison. There are at least two economic arguments for why one might have expected no causal impact. First, if unemployment were so pervasive that HIV+ individuals would be unemployed even in the absence of HIV, one would expect no impact. Second, if ARVs were sufficiently widely used, one might expect either no impact or a tiny impact. Our estimates indicate that these arguments, while perhaps *ex ante* plausible, are on average simply incorrect. Being HIV+ lowers the probability of employment.

A counter-argument to the notion that HIV confers a negligible penalty in the face of extremely high unemployment is that in the presence of high unemployment, even a small disadvantage (e.g. the stigma sometimes associated with HIV) much less a large disadvantage (adverse physical effects of HIV) mean the difference between keeping a job and losing it. The results in Table 6 are consistent with this counter-argument.

One way to gauge the magnitude of our estimates (beyond noting that they are not zero) is to compare the estimated marginal effect of HIV status (D) with the estimated marginal effect of other respondent characteristics (X) in Equation (4). For men, the magnitude of the labor market advantage of being HIV- is approximately equal to the impact of 3 years of age, a Matric qualification (compared to no education), or the absence of a female pensioner in the household. For women it is equal to the impact of 1.5 years of age, some secondary education (compared to no education), or the absence of a male pensioner in the household. It is worth noting that our results are likely an underestimate of the true magnitude of the effect of HIV status on labor market outcomes because they do not account for intra-category changes in employment outcomes. As their disease progresses, HIV+ workers may shift into less physically demanding occupations at reduced wages, which has implications for inequality.

We have estimated the causal impact of HIV status on the probability of unemployment, and in this sense HIV contributes to unemployment in South Africa. It would be wrong, though, to think that HIV was substantially *responsible* for South Africa's dire unemployment. The numbers simply don't add up.

Our results are conditional on the availability of ARVs as of 2005. We do not have data on which HIV+ respondents were on ARV therapy. Access to ARVs in South Africa is far from universal. One estimate is that in 2005, the year of our sample, only about 18 percent of those who needed ARVs were actually using them. (See Dorrington, Bradshaw, Johnson and Daniel (2006).) As the availability of ARVs changes, the impact of HIV on labor market status, as we estimate it, will change. Because ARVs are more widely available today than they were in 2005, *ceteris paribus*, the impact of HIV on unemployment is lower today than it was in 2005. In addition, ARV usage is non-random, and this may in part contribute to the pattern of results in Table 6. For example, ARVs are much more likely to be employer-provided in the formal sector than they are in the informal sector. Our findings are consistent with the fact that women and the less-educated tend to be more heavily represented in the informal sector and in domestic (housekeeper) work and hence less likely to receive employer-provided ARVs.

We also found that being HIV+ had virtually no employment impact for better educated workers (and recall this result already accounts for the fact that highly educated individuals are less likely to be HIV+). ARVs may also be contributing to this finding. Employers have greater incentive to invest in ARVs for workers who are more difficult to replace such as highly educated workers.

In sum, HIV appears to reinforce the already existing inequalities in South Africa.

5.4 Sensitivity Analyses

We investigate the sensitivity of our results both to alternative definitions of unemployment and to alternative specifications of the propensity score regression.

As described in section 3, individuals were divided into three groups in our base case specification – unemployed, employed, and not economically active (NEA). The first two groups are included in our analysis while the third is not. In the base case results using the broad definition of unemployment, so-called discouraged workers – workers who were not actively seeking a job but desired employment – were classified as unemployed. The narrow definition of unemployment classifies discouraged workers as NEA. We repeat the analysis using this narrow definition of unemployment. Results are reported in Appendix Table A1. The results are very similar to those in Table 6. The estimated coefficients are about one-half to two-thirds the size of those reported in Table 6. We also repeated the analysis by combining the NEA with the employed to create a group that might (awkwardly) be called not-unemployed. Results are reported in Appendix Table A2. Compared to Table 6, coefficients tend to increase, usually by about one and one-quarter to one and a half times. Results tend to be more precisely estimated, and this is driven in part by the increase in the sample size when NEA individuals are included in the analysis. From Appendix Tables 1 and 2, we conclude that our findings are robust to alternative definitions of labor market status.

We also experimented with alternative specifications of the propensity score regression. The results obtained from regressions using additional variables in the calculation of the propensity score are virtually identical to results obtained from the preferred specification. These results are not reported here. When including ten variables with information on the household, no point estimate changed by more than two percentage points, and only three

of the 15 point estimates changed by more than one percentage point. The additional household variables chosen to indicate socioeconomic level were: type of toilet facility, source of energy for cooking, access to electricity, presence of a land line, and a dummy for whether the household information was missing for the observation. Adding additional behavioral variables did not change any of the point estimates by as much as one percentage point. The additional behavioral variables were condom use at last sex, number of current sexual partners, number of partners in the last year, whether the respondent had been tested for HIV before, whether they had received their test result, and whether they had heard of ARVs.

6 Conclusions and Caveats

6.1 Conclusions

Identifying the causal impact of HIV on labor market status requires addressing the issue of selection into being HIV+. In the absence of plausible instruments, we exploit the richness of the data and assume that selection is on observables. External balancing tests support the validity of this assumption. Employing two estimation strategies, we find that being HIV+ causes, on average, an increase in the likelihood of unemployment of six to seven percentage points. This penalty exists despite very high unemployment rates. The average impact hides important heterogeneity. HIV's causal impact on unemployment is larger (10 to 11 percentage points) for less educated South Africans. The results are robust to multiple alternative econometric specifications.

6.2 Caveats

These results are the first nationally representative estimates of the causal impact of HIV on employment in South Africa. While informative, they are not dispositive. Rather, the results should be interpreted with caution for at least five reasons.

First, our analysis does not account for any general equilibrium effects. In particular, it would be misleading to think that if ARVs were made universally available or a cure for HIV/AIDS were found, that HIV+ individuals would see their likelihood of employment rise on average by about seven percentage points. Rather, the labor market would adjust and these adjustments would depend on supply and demand elasticities. Complicating this analysis, the ability of the labor market to immediately absorb additional healthy workers is questionable.

Second, data limitations preclude an analysis of the indirect labor market impact of having multiple HIV+ adult household members. Recall that the structure of the SABSSM II survey is such that only one adult age 25 or older is sampled from each household. It is unclear in which direction our results may be biased. An HIV- worker could be unemployed because they are caring for an HIV+ spouse, resulting in downward bias (i.e. the true impact is larger than our estimates suggest), or an HIV+ worker might be more motivated to obtain employment to financially support another HIV+ household member, resulting in upward bias.

Third, our results are conditional on the time profile of HIV and prevalence rates as of 2005. Given an approximately nine-year period (on average) of latent HIV infection before AIDS conversion, and HIV prevalence rates that increased from 5 percent in 1996 to 12 percent in 2001,[19] we would expect the "stock" of individuals with AIDS to increase quite sharply between 2003 and 2008. This implies that, *ceteris paribus*, the impact of HIV on

unemployment would rise in coming years as the number of HIV+ individuals who are too ill to work increases. However, it is unclear how this effect would interact with any increase in the availability of ARVs.

Fourth, our results are not structural. As such, we are unable to convincingly address the particular avenues through which HIV impacts labor market status. Relatedly, we are unable to conduct detailed policy analysis. For example, increased access to ARVs and successful programs to de-stigmatize HIV might each increase the likelihood of employment, but our approach cannot conduct the counterfactual experiments to estimate the likely impacts of these potential policies.

Lastly, our results are generally not applicable to other countries. South Africa has a stunningly high rate of unemployment, high HIV prevalence rates and a troubled history with the distribution of ARVs – three factors that suggest that it may be misguided to generalize the results of this study to other countries.

Notes

[1]There are other data sets available for which the techniques employed in this paper are potentially applicable. These include some of the Demographic Health Surveys (DHS) and the Botswana AIDS Impact Survey (see Levinsohn and McCrary (2008))

[2]These challenges are discussed in detail in Strauss and Thomas (1998).

[3]With country-level data, the fraction of the male population that has been circumcised has been used as an instrument (see Werker et al. (2006)). At the individual-level, this instrument is problematic. The protective effect may not transfer to women (Wawer, Kigozi, Serwadda, Makumbi, Nalugoda, Watya, Buwembo, Ssempijja, Moulton and Gray (2008)) and there is often virtually no variation within large categories of the male population like religion, race and ethnic/tribal affiliation (see McKelvey (2007)).

[4]For example, Thomas and et al. (2006) use experimental methods to obtain a causal estimate of the effect of improved health (in this case, from increased iron intake) on labor market outcomes.

[5]These methods are drawn mostly from the program evaluation literature, but they have also been used in other fields of economics. For example, a similar control function approach is used in Olley and Pakes (1996) for solving the endogeneity problem in the context of production functions. Levinsohn and McCrary (2008) use the propensity score re-weighting technique to address a missing data problem in the context of HIV prevalence.

[6]Thirumurthy, Zivin and Goldstein (2005) also found that in western Kenya, labor supply increased within six months of initiating ARV therapy.

[7]A detailed description of the survey methodology is found in Shisana and et al. (2005), available online at: http://www.hsrcpress.ac.za/product.php?productid=2134&cat=0&page=1

[8]When matching individuals to households based on household identification numbers, approximately 14 percent of the individual observations (3413) could not be matched to household data. Household variables were imputed to zero for these unmatched observations. A specification check suggests that the data were consistent with a pattern of "missing at random"; a dummy variable for being unmatched was not significant in any specification. As with all variables that are imputed, an indicator variable was created that took a value of 1 if the value was imputed and 0 otherwise. These indicator variables are included in the empirical specifications. Individual survey records were matched to HIV incidence and viral load data using the barcode number that identified each specimen.

[9]The prevalence rates in Table 2 do not correct for sample selection due to non-random opt-out of the HIV test. We have replicated our analysis on the impact of HIV on labor market participation with weights that do correct for non-random opt-out (as computed by HSRC) and the results are virtually identical. This suggests that selection into testing is independent of the effect of HIV on employment. These rates apply to our sample which only includes individuals aged 15 and over. Hence they differ from the national prevalence rates reported in Shisana and et al. (2005).

[10]A version of this paper that includes all races in all the results is available on Levinsohn's website.

[11]Those who responded being in the "other" category are grouped with the employed to generate conservative estimates.

[12]It is important to note that these are not the same variables we would use as instruments in an instrumental variables (IV) regression. In that case we would want Zs correlated with HIV status but not with labor market status; in our case these Zs should not be included in the propensity score regression. Because actual HIV status is known by the econometrician and can therefore be controlled for in the regression, the covariates in X do not need to be good predictors of HIV status. In fact, selecting the attributes for X depending on predictive power can actually increase bias. See Heckman and Navarro-Lozano (2004).

[13]Frölich (2004) demonstrates that an estimated propensity score close to 1 can cause problems for estimating the ATET. Our values of the estimated propensity scores are not large enough for this to be a concern.

[14]A similar control function method can be used if there is selection on *un*observables (see Heckman and Navarro-Lozano (2004)). However, allowing for selection on unobservables requires an exclusion restriction– a variable that is correlated with HIV status but uncorrelated with labor market status.

[15]Elements of the X vector of covariates were imputed to zero for item non-response, and a dummy variable for imputation was included in the specifications. We include indicator variables for each province in South Africa. These coefficients are not reported in the table.

[16]We also examine particular subsamples of the data (e.g. by sex, by education, by area of residence). For each of these sub-samples, we re-estimate the propensity score. These results are available on request.

[17]The support condition is not required for the control function approach, however confirming that it holds ensures that we are not relying on solely on functional form assumptions for any values of the propensity score.

[18]Both the reweighting and the control function approach use the *estimated* propensity score. Hirano, Imbens and Ridder (2003) show that using the estimated propensity score rather than the true propensity score produces efficient estimates. They suggest bootstrapping to obtain standard errors. We do so.

[19]ASSA demographic model predictions cited in Natrass (2004) (p. 42). The model on which these estimates are based predicted a 14 percent prevalence rate for 2004 – about the same as that of our 2005 data.

References

Arndt, Channing and Jeffrey D. Lewis, "The Macro Implications of HIV/AIDS in South Africa: A Preliminary Assessment," *South African Journal of Economics*, December 2000, *68* (5), 856–87.

Auvert, Bertran, Dirk Taljaard, Emmanuel Lagarde, Joelle Sobngwi-Tambekou, Remi Sitta, and Adrian Puren, "Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial," *PLoSMed*, 2005, *2* (11), e298.

Bell, Clive, Shantayanan Devarajan, and Hans Gersbach, "The Long-Run Economic Costs of AIDS: Theory and an Application to South Africa," *World Bank Policy Research Working Paper*, October 2003, (3152).

Bloom, David E. and Ajay S. Mahal, "Does the AIDS epidemic threaten economic growth?," *Journal of Econometrics*, 1997, *77* (1), 105–124.

Busso, Matias, John DiNardo, and Justin McCrary, "New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators," 2010. Working Paper.

Card, David, "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 1999, *3*, 1801–1863. Editor: O. Ashenfelter and D. Card; Publisher:Elsevier Science B.V.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, April 2005, *113* (2), 376–424.

Cuddington, J. and J. Hancock, "Assessing the impact of AIDS on the growth path of the Malawian economy," *Journal of Development Economics*, 1994, *4*, 363–8.

Dehejia, Rajeev H. and Sadek Wahba, "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," in Rajeev Dehejia, ed., *Econometric Methods for Program Evaluation*, Ph.D. Dissertation, Harvard University, 1997, chapter 1.

___ **and** ___ , "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 1999, *94* (448), 1053–1062.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 1996, *64* (5), 1001–1044.

Dinkelman, T., D. Lam, and M. Leibbrandt, "Household and community income, economic shocks and risky sexual behavior of young adults: evidence from the Cape Area Panel Study 2002 and 2005," *AIDS*, 2007, *21* (Suppl 7), S49–S56.

Dorrington, R., D. Bradshaw, L. Johnson, and T. Daniel, "The demographic impact of HIV/AIDS in South Africa: National and Provincial Indicators for 2006," Technical Report, Cape Town 2006. Joint publication of the Centre for Actuarial Research, the Burden of Disease Research Unit, and the Actuarial Society of South Africa (ASSA).

Dunkle, K., Rachel K. Jewkes, Heather C. Brown, Glenda E. Gray, James A. McIntyre, and Sioban Harlow, "Transactional sex among women in Soweto, South Africa: Prevalence, risk factors and association with HIV infection.," *Social Science and Medicine*, 2004, *59*, 1581–92.

Fox, Matthew P., Sydney Rosen, William B. MacLeod, Monique Wasunna, Margaret Bii, Ginamarie Foglia, and Jonathon L. Simon, "The impact of HIV/AIDS on labour productivity in Kenya," *Tropical Medicine and International Health*, March 2004, *9* (3), 318–24.

Frölich, Markus, "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, 2004, *86* (1), 77–90.

Habyarimana, James, Bekezela Mbakile, and Cristian Pop-Eleches, "The Impact of HIV/AIDS and ARV Treatment on Worker Absenteeism: Implications for African Firms," *Journal of Human Resources*, 2010, *45* (4), 809–839.

Heckman, James and Salvador Navarro-Lozano, "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models," *The Review of Economics and Statistics*, February 2004, *86* (1), 30–57.

Heckman, James J. and Richard Robb, "Alternative Methods for Estimating the Impact of Interventions," *Longitudinal Analysis of Labor Market Data*, 1985. Ed: James J. Heckman and Burton Singer; Pub: Cambridge University Press.

___ **and V. Joseph Hotz**, "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 1989, *84* (408), 862–874.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, July 2003, *71* (4), 1161–1189.

Kalemli-Ozcan, Sebnem, "AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa," *NBER Working Paper*, April 2006, (12181).

Kambou, G., S. Devarajan, and M. Over, "The economic impact of AIDS in an African country: simulations with a computable general equilibrium model of Cameroon," *Journal of African Economies*, 1992, *1*, 109–130.

LeClerc-Madlala, S., "Youth, HIV/AIDS and the importance of sexual culture and context," *Social Dynamics - A Journal of the Centre for African Studies, University of Capetown*, 2002, *28*, 20–41.

Levinsohn, James and Justin McCrary, "Estimating Correct HIV Prevalence Rates: The Case of Botswana," 2008. Working Paper.

McKelvey, Christopher, "Circumcision and the Labor Market Consequences of HIV in Developing Countries," October 2007. Unpublished.

Murray, Jill, Pam Sonnenberg, Gill Nelson, Stuart Shearer, Andre Bester, Arthur Begley, and Judith R. Glynn, "Effect of HIV on work-related injury rates in South African gold miners," *AIDS*, 2005, *19*, 2019–2024.

Natrass, Nicoli, *The Moral Economy of AIDS in South Africa*, Cambridge University Press, 2004.

Olley, S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1298.

Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

Shisana, Olive and et al., *South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey, 2005*, Human Sciences Research Council, 2005. Published in Cape Town.

Strauss, John and Duncan Thomas, "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, 1998, *36* (2), 766–817.

Thirumurthy, Harsha, Joshua Graff Zivin, and Markus Goldstein, "The Economic Impact of AIDS Treatment: Labor Supply in Western Kenya," *NBER Working Paper*, 2005, (11871). Forthcoming in JHR (XXX).

Thomas, Duncan and et al., "Causal Effect of Health on Labor Market Outcomes: Experimental Evidence," *California Center for Population Research On-Line Working Paper Series*, September 2006, (070).

UNAIDS, "UNAIDS/WHO AIDS Epidemic Update," Technical Report, UNAIDS/WHO December 2005.

Wawer, Maria, G Kigozi, D Serwadda, F Makumbi, Nalugoda, S Watya, D Buwembo, V Ssempijja, L Moulton, and R Gray, "Trial of Male Circumci-

sion in HIV+ Men, Rakai, Uganda: Effects in HIV+ Men and in Women Partners," February 2008. Paper presented at Conference on Retroviruses and Opportunistic Infections.

Weiss, Helen A., Maria A. Quigley, and Richard J. Hayes, "Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis," *AIDS*, 2000, *14*, 2361–70.

Werker, Eric, Amrita Ahuja, and Brian Wendell, "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis," *Harvard Business School Working Paper*, October 2006, (07-025).

Young, Alwyn, "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," *Quarterly Journal of Economics*, 2005, *120* (2).

Zuma, K., M.N. Lurie, B.G. Williams, D. Mkaya-Mwamburi, G.P. Garnett, and A. W. Sturm, "Risk factors of sexually transmitted infections among migrant and non-migrant sexual partnerships from rural South Africa," *Epidemiol.Infect.*, 2005, *133*, 421–428.

Table 1: Sample sizes

	Full sample	African only
Total sample	23,275	13,935
Aged 15 and up	16,398	9,664
Agreed to HIV test	12,032	7,443
Tested HIV+	1,351	1,239
Infected for more than 6 months	1,172	1,068

Table 2: Naïve HIV prevalence rates by race

Race	HIV prevalence rate (%)	N
African	17.34	7,443
Coloured	2.73	1,041
Indian	1.27	2,468
White	0.53	1,058
Total	14.14	12,010

[1] Sample includes all individuals aged 15 and over.

[2] Rates are calculated using sample weights only.

Table 3: Employment status by HIV testing status (for African men and women)

Employment status	HIV+	HIV-	Not tested	Total	N
Men					
Employed (and other)	44.53	34.82	37.64	36.54	1,329
Unemployed (broad definition)	38.19	27.92	27.93	28.99	986
Student (NEA)	5.02	25.25	26.54	23.47	921
Other NEA	5.15	9.60	7.18	8.53	301
Missing data	7.12	2.41	0.71	2.47	65
Total	100	100	100	100	3,602
Women					
Employed (and other)	21.19	20.00	21.82	20.57	1,309
Unemployed (broad definition)	56.61	35.67	38.93	39.66	2,260
Student (NEA)	8.50	16.26	17.14	15.22	1,059
Other NEA	13.20	26.97	21.34	23.62	1,374
Missing data	0.50	1.10	0.77	0.93	60
Total	100	100	100	100	6,062

[1] Sample includes all individuals aged 15 and over. NEA: not economically active.

Table 4: Propensity score regression for HIV+ status (base case)

Variable	Coefficient
Age	0.138***
	(0.051)
Female	0.470***
	(0.112)
Urban resident	0.305**
	(0.127)
Never married	0.371***
	(0.126)
Completed primary education	0.347
	(0.216)
Completed secondary education	0.338*
	(0.201)
Holds a matric qualification	0.096
	(0.232)
Has some post-matric education	-0.514*
	(0.277)
Has had sex	2.454**
	(1.066)
Age at first sex	-0.165*
	(0.096)
Age at first sex squared	0.004*
	(0.002)
Used a condom at first sex	-0.285*
	(0.169)
Knows HIV transmitted through vaginal sex	-0.169
	(0.209)
Believes HIV not transmitted through witchcraft	-0.139
	(0.169)
Knows condoms prevent HIV transmission	-0.121
	(0.175)
Knows reducing number of partners reduces risk	0.199*
	(0.119)
Male of pension age in household	-0.002
	(0.479)
Female of pension age in household	-0.233
	(0.193)
Pseudo-R^2	0.07
Observations	4,620

Table 5: Balancing tests for variables used in estimation of propensity score.

	Unweighted			Reweighted		
Variable	**Diff.**	**Std. E**	**t-Stat**	**Diff.**	**Std. E**	**t-Stat**
Age	-2.587	0.533	-4.86	-0.255	0.496	-0.51
Female	0.106	0.026	4.02	0.000	0.027	-0.01
Eastern Cape	-0.019	0.015	-1.23	-0.002	0.015	-0.16
Northern Cape	-0.003	0.002	-1.82	0.000	0.002	-0.08
Free State	0.003	0.019	0.17	-0.001	0.018	-0.03
Kwa-Zulu Natal	0.069	0.022	3.10	0.003	0.024	0.11
Northwest Province	-0.011	0.018	-0.65	0.001	0.017	0.07
Gauteng	0.019	0.023	0.81	0.004	0.024	0.18
Mpumalanga	0.034	0.014	2.50	-0.003	0.015	-0.19
Limpopo	-0.064	0.013	-4.77	-0.002	0.012	-0.19
Urban dweller	0.071	0.026	2.67	0.002	0.027	0.09
Never married	0.088	0.026	3.34	0.014	0.027	0.50
Has had sex	0.016	0.013	1.30	0.002	0.012	0.14
Age at first sex	-0.150	0.420	-0.36	0.021	0.432	0.05
Age at first sex-squared	-8.242	9.125	-0.90	-0.366	9.420	-0.04
Completed primary education	0.006	0.023	0.27	-0.007	0.024	-0.29
Completed secondary education	0.057	0.026	2.18	0.008	0.027	0.30
Holds a matric qualification	0.007	0.021	0.36	0.001	0.022	0.03
Completed some post-matric educ	-0.041	0.011	-3.80	0.002	0.010	0.20
Used a condom at first sex	-0.021	0.018	-1.15	0.003	0.018	0.17
Knows HIV transmitted through vaginal sex	-0.006	0.013	-0.49	-0.002	0.014	-0.18
Believes HIV not transmitted through witchcraft	-0.010	0.019	-0.54	0.002	0.020	0.10
Knows condoms prevent HIV transmission	-0.004	0.016	-0.25	0.003	0.016	0.16
Knows reducing number of partners reduces risk	0.027	0.025	1.10	0.001	0.025	0.04
Male of pension age in household	-0.015	0.018	-0.83	-0.003	0.017	-0.16
Female of pension age in household	-0.025	0.013	-1.86	-0.001	0.014	-0.10

[1] Difference between HIV+ and HIV- subsamples. Includes all individuals aged 15 and older.

Marginal effect of being HIV+ from naïve (unconditional) difference in means, logit with controls, propensity score reweighting (RW) and control function (CF).

Table 6: Marginal effect of being HIV+ on likelihood of being *un*employed.

Sample	Naïve		Logit		RW		CF		N
Full sample	0.079	***	0.065	**	0.062	***	0.070	***	4620
	(0.026)		(0.029)		(0.024)		(0.022)		
Male	-0.003		0.064		0.061		0.080		1761
	(0.044)		(0.051)		(0.046)		(0.041)		
Female	0.095	***	0.059		0.055		0.060	**	2859
	(0.028)		(0.031)		(0.028)		(0.026)		
Matric and up	0.044		-0.027		-0.006		0.014		839
	(0.059)		(0.073)		(0.072)		(0.059)		
Less than Matric	0.103	***	0.109	***	0.100	***	0.109	***	2581
	(0.035)		(0.035)		(0.031)		(0.031)		
Rural	0.082	**	0.051		0.046		0.067	**	1973
	(0.041)		(0.041)		(0.035)		(0.032)		
Urban	0.091	***	0.080	**	0.076	**	0.070	**	2647
	(0.033)		(0.039)		(0.033)		(0.029)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score estimation.

[3] Bootstrapped standard errors reported for RW and CF. *** significant at 99 percent level, ** significant at 95 percent level.

Appendix Table A1

Sensitivity check using narrow definition of unemployment.
Marginal effect of being HIV+ from naïve (unconditional) difference in means, logit with controls, propensity score reweighting (RW) and control function (CF).

Sample	Naïve		Logit	RW	CF		N
Full sample	0.068	***	0.025	0.034	0.038	***	4620
	(0.026)		(0.029)	(0.026)	(0.024)		
Male	-0.012		0.021	0.031	0.049		1761
	(0.042)		(0.047)	(0.045)	(0.043)		
Female	0.086	***	0.031	0.027	0.040	**	2859
	(0.031)		(0.034)	(0.032)	(0.029)		
Matric and up	0.017		-0.049	-0.007	0.015		839
	(0.058)		(0.067)	(0.064)	(0.058)		
Less than Matric	0.091	***	0.044	0.051	0.063	***	2581
	(0.035)		(0.037)	(0.033)	(0.032)		
Rural	0.063		0.002	0.009	0.039	**	1973
	(0.043)		(0.047)	(0.042)	(0.034)		
Urban	0.082	**	0.038	0.049	0.040	**	2647
	(0.033)		(0.036)	(0.033)	(0.030)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score estimation.

[3] Bootstrapped standard errors reported for RW and CF. *** significant at 99 percent level, ** significant at 95 percent level.

Appendix Table A2

Sensitivity check including not economically active (NEA) individuals in the sample.
Marginal effect of being HIV+ from naïve (unconditional) difference in means,
logit with controls, propensity score reweighting (RW) and control function (CF).

Sample	Naïve		Logit		RW		CF		N
Full sample	0.180	***	0.086	***	0.093	***	0.104	***	7443
	(0.022)		(0.023)		(0.022)		(0.021)		
Male	0.103	***	0.048		0.063		0.101	***	2672
	(0.039)		(0.040)		(0.044)		(0.039)		
Female	0.209	***	0.096	***	0.100	***	0.087	***	4771
	(0.026)		(0.028)		(0.027)		(0.025)		
Matric and up	0.056		-0.007		0.013		0.023		932
	(0.056)		(0.064)		(0.067)		(0.056)		
Less than Matric	0.178	***	0.090	***	0.096	***	0.096	***	3804
	(0.029)		(0.030)		(0.029)		(0.029)		
Rural	0.177	***	0.073	**	0.084	**	0.099	***	3538
	(0.034)		(0.035)		(0.035)		(0.031)		
Urban	0.185	***	0.100	***	0.106	***	0.110	***	3905
	(0.029)		(0.030)		(0.030)		(0.028)		

[1] Full sample and results by sex include individuals aged 15 and older. Results by education level are restricted to individuals aged 25 and older.

[2] Naïve is OLS without controls. Controls used in logit are same as those used in propensity score estimation.

[3] Bootstrapped standard errors reported for RW and CF. *** significant at 99 percent level, ** significant at 95 percent level.

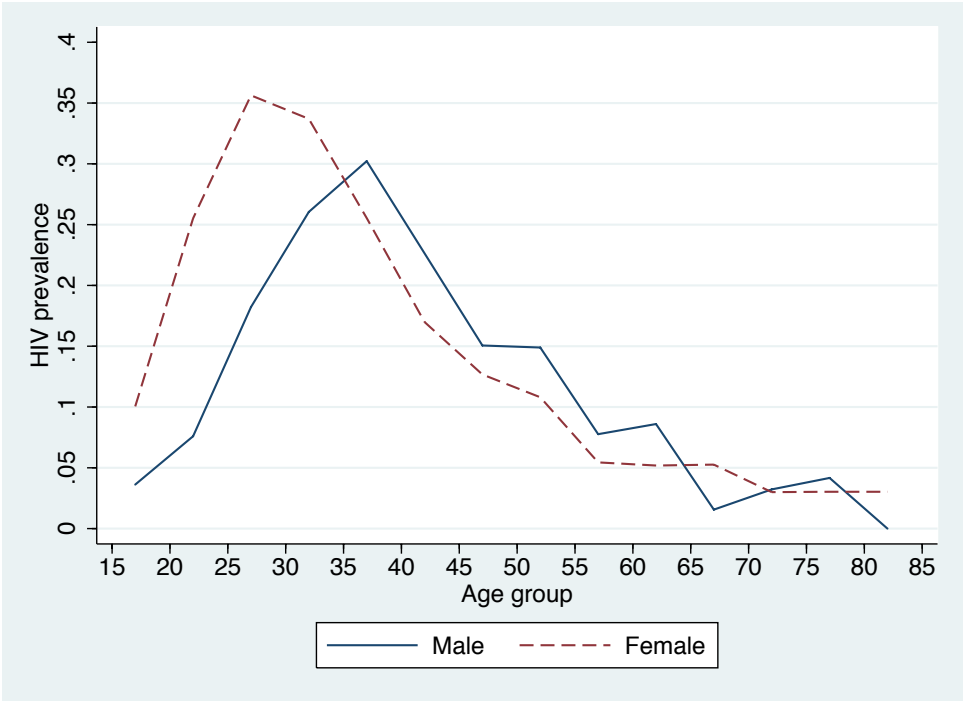

Figure 1: Age profile of naïve HIV prevalence– Africans only

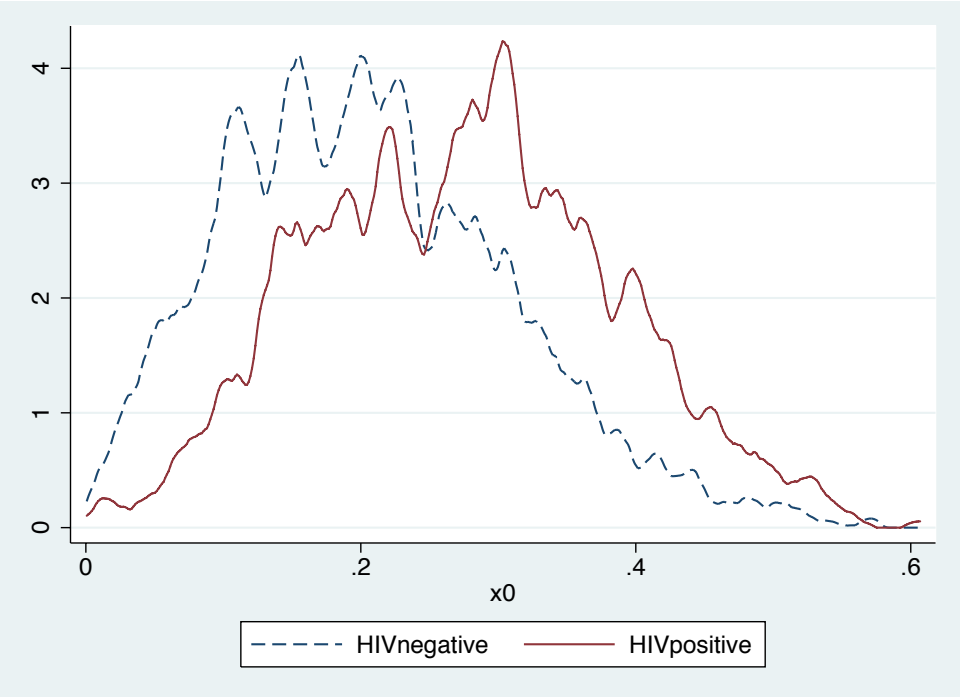

Figure 2: Empirical distributions of the propensity scores for the full African sample (N=4620.)