Female Education, HIV/AIDS and the Education Vaccine in Sub-Saharan Africa: The Condom Piece of the Puzzle

ROBERT J. BRENT

ABSTRACT

There are many recent studies that indicate that female education and HIV are still positively related in sub-Saharan Africa (SSA). So education is not a vaccine as claimed and there exists a puzzle that needs to be resolved. In this study, by focusing on the use and non-use of condoms by females, we provide one piece of the puzzle. That is, we identify one transmission mechanism by which female education leads to higher rates of HIV infection. We find that educated females are less likely to use condoms with their spouses and regular partners than less education females.

JEL Classification: O1, I1

Keywords: Female Education, HIV/AIDS, Condoms, Sub-Saharan Africa, Tanzania

---

1 Professor, Department of Economics Fordham University, Bronx NY 10458. Email: brent@fordham.edu; website: www.fordham.edu/robertjbrent
1. INTRODUCTION

There exists a dominant / official view, first expressed by researchers at the World Bank - Over (1998) as background paper for 1999 World Development Report - that education used to be (in the 1980s and early 1990s) positively related to HIV, but that this relationship has subsequently been reversed over time. The alleged reversal has been so strong that investing in education was alleged to be a "vaccine" - see Vandemoortele and Delamonica (2002). On the basis of experience in Uganda, the Global Campaign for Education (2004) projected that universal primary education would save at least 7 million lives over a decade because of the education vaccine. Given the difficulty of developing an actual medical vaccine, this educational vaccine was thought to be the only one that will be available in the foreseeable future. However, there are many recent studies that indicate that female education and HIV are still positively related in sub-Saharan Africa (SSA), so there exists a puzzle that needs to be resolved. In this study, by focusing on the use and non-use of condoms by females, we provide one piece of the puzzle. That is, we identify one transmission mechanism, not using condoms, by which female education leads to higher rates of HIV infection. We find that educated females are less likely to use condoms with their spouses than less education females, unlike their behavior with regular and irregular partners, which is what researchers in this area usually highlight. Understanding this finding and bringing it to the attention of policy makers who are involved with HIV/AIDS interventions is important to ensure that female education will actually become a vaccine in the near future.

The relationship between education and HIV is important to understand for both male and females. We will focus just on female education because we will be discussing this relationship mainly in the context of SSA and in this part of the world the majority of people with HIV/AIDS are female, approximately 60% in 2010, UNAIDS (2011). Female choices in Africa are constrained by male-female power relationships. Trying to understand female decision-making is very important for policy purposes.

In World Bank (2002) they explained why there should be a negative relationship between female education and HIV. Increasing female general education would lead to prevention: "Better educated women are more likely, in comparison with their peers, to delay marriage and childbearing, have fewer children and healthier babies, enjoy better earning potential, have stronger decision making and negotiation skills as well as higher self-esteem, and avoid commercial sex." While there is nothing wrong with the World Bank's reasoning,
BRENT: THE CONDOM PIECE OF THE PUZZLE

unfortunately, the dominant view does not square with reality. Hargreaves et al. (2008) summarized the results of 35 studies dealing with the relationship between female education and HIV prevalence. Prior to 1996, 9 of the studies found a positive relationship, 6 had none, and 0 had a negative relation. Post 1996, 4 of the studies had a positive relationship, 13 had none, and 3 had a negative relation. While there has been some reversal over time, it is still the case that including both periods, the vast majority (32 out of 35) had a positive or zero relationship. We call this positive relationship between female education and HIV a “puzzle” as it is so counter-intuitive.

A study by Brent (2006) using HIV data for 2000, found that for 31 countries in SSA, the relation between HIV and female education was still positive. It used nine measures for female education, covering female literacy, gross and net enrollments in primary and secondary schooling, non-standard (overage) students, and including also all these measures expressed relative to male education (in the form of enrollment “gaps”). Talk of a vaccine is very premature. At this point in history the relationship can still be expected to be positive for many African countries. In this article we present one empirically supported possible explanation for the positive relationship between female education and HIV/AIDS prevalence in SSA.

There are a number of possible logical explanations for this positive sign, seeing that education is associated with higher socio-economic status, and the more educated are likely to live in urban areas, have more leisure time, travel more, and have greater mobility, all factors that are associated with HIV. But, it is important to understand how this positive sign can be reversed. Understanding female condom behavior is a necessary first piece in explaining the female education and HIV/AIDS puzzle, seeing that this disease is transmitted sexually and promoting condoms is the main intervention method that the Joint United Nations Programme on HIV/AIDS (UNAIDS) is relying on to prevent the spread of this disease. Expanding female education is a core development strategy in most countries of the world, so clearing up this puzzle is of prime importance for policy-makers.

In this article we concentrate on the experience of Tanzania. Tanzania’s HIV prevalence rate is typical of the seven countries in Central and East Africa that had rates above 5% (Cameroon, the Central Africa Republic, Gabon, Malawi, Mozambique, Uganda and the United Republic of Tanzania) that were not in the group of seven countries where the prevalence rates exceeded 15% (Botswana,
Lesotho, Namibia, South Africa, Swaziland, Zambia and Zimbabwe).\textsuperscript{2} Tanzania was a good country to study from the point of view of shedding light on the puzzle as there was a literature on female education, condoms and HIV rates to build on and the nationally representative survey for Tanzania, the Tanzanian HIV/AIDS Indicator Survey (THIS), confirmed that the positive gradient existed in Tanzania as recently as 2005.\textsuperscript{3} The puzzle is especially evident in Tanzania seeing that Ainsworth and Semali (1999) found that females completing primary education in Tanzania have been found to have the highest probability of dying of AIDS, even though for all non-AIDS deaths the probability of dying is greatest for those with no schooling or only a few years of primary education.

The positive relation between female education and HIV in Tanzania does not mean that investing in female education can not be socially worthwhile. In a Cost-Benefit Analysis of female primary education in Tanzania, Brent (2009a) found that the total effect of female enrolments was to decrease regional HIV rates. This came about because, although the direct effect of education on a region’s HIV rate was still perverse, the indirect effect working through education’s effect on regional income was in the opposite direction and large enough to reverse the net effect. So although income \textit{levels} raised HIV levels, changes in income brought about \textit{changes} (i.e., reductions) in HIV levels. As a result, educating a cohort of 20,507 females brought about a reduction of roughly 1,408 HIV cases. Valuing the lives saved by the present value of lifetime earnings, and valuing the costs by the present value of 7 years of tuition (and other costs), the benefit-cost ratio was in excess of 1 (i.e., between 1.3 to 2.9 in the best estimates), revealing that primary education was socially profitable in Tanzania.

Notwithstanding this result that female education was a worthwhile government investment in Tanzania, it is still necessary to try to understand why a positive relationship between female education and HIV could exist. So emphasis was given to the use, or non-use, by educated females of condoms. A study of condom use in Tanzania by Badru (2000) revealed that condom use very much depended on the choice of partner, i.e., whether he was a regular or a non-regular partner. So we will be examining condom use by education and choice of partner to see if this can help us unravel the education and HIV puzzle.

\textsuperscript{2} UNAIDS (2011), op.cit.
\textsuperscript{3} Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS), and ORC Macro (2005): \textit{Tanzania HIV/AIDS Indicator Survey} 2003-2004 (THIS). Calverton, Maryland, USA.
The outline of the rest of the article is this. Section II contains the theoretical framework. The purpose is to help identify the transmission mechanism by which female education determines condom use. Our method involves including an interaction term between education and her choice of sex partner in the regressions used to explain female condom behavior. This interaction term plus the education variable itself will combine to produce the total effect of education on condom use. Section III covers the data and specification of variables and section IV presents the estimation framework. The results are detailed in section V. Two policy implications of our analysis are examined in the discussion section VI. First we identify what is necessary to add to prevention messages to ensure that female education does in fact become a vaccine for HIV. Second, since Uganda was in the same HIV prevalence class as Tanzania, was the basis for the vaccine projections, and was one of the first countries to experience the reversal of the education gradient, a comparison of Tanzanian and Uganda findings will take place. We close with the summary and conclusions.

2. THEORETICAL FRAMEWORK

The determinants of the probability of female condom use will be embedded within the standard random utility model. Education and the choice of sex partner will be included as variables, together with measures of individual characteristics. The education variable will then be interacted with the sex partner variable. The total effect of education is decomposed into a direct and an induced effect. It is the total effect that reveals the education transmission mechanism. As the education variable will enter as a dummy variable, the marginal probabilities will be estimated using a discrete change specification. All the details are contained in the appendix. Here we just summarize the key ingredients in the analysis.

The dependent variable $Y_i$ is the choice by the female $i$ whether to use or not use a condom the last time she had sex and is represented by a dichotomous dummy variable: $Y_i = 1$ when a condom was used the last time the female had sex, and $Y_i = 0$ when a condom was not used the last time the female had sex. The determinants $X_i'$ $\beta$ of the female's condom use will be

---

4 The analysis will involve a binary choice as we exclude from our sample the females who are virgins and therefore abstain from sex. This latter group could not be using a condom as they do not have sex and so would not have any type of sex partner. In countries in SSA like Tanzania, where heterosexual sexual activity is the main transmission mechanism for HIV, virgins cannot be part of the
classified into two categories: the education variable $E_i$ and the $k$ individual characteristics $Z_{ki}$:

$$Y_i = X_i' \beta = \beta_0 + \beta_E E_i + \sum_{k} \delta_k Z_{ki}$$

where $\beta_0$ is the constant term, and $\beta_E$ and the $\delta_k$ are fixed coefficient.

The education transmission mechanism works through the coefficient $\beta_E$. Let $\beta_E$ be linearly related to the choice of $j$ sex partners $P_{ji}$ in the form: $\beta_E = \beta_1 + \beta_j P_{ji}$. Substituting for $\beta_E$ in (1) produces:

$$Y_i = X_i' \beta = \beta_0 + \beta_1 E_i + \beta_j [E_i P_{ji}] + \sum_{k} \delta_k Z_{ki}$$

From (2) we derive:

$$\frac{\partial Y_i}{\partial E_i} = \frac{\partial (X_i' \beta)}{\partial E_i} = \beta_1 + \beta_j P_{ji}$$

The total effect of changes in the level of education on the use of condoms in (3) is split up into a direct effect (the first term on the RHS) and an induced effect (the second term on the RHS) which works through the sex partner variable, where $E_i P_{ji}$ would typically be set equal to the sample mean $\overline{E_i P_{ji}}$.

For female education to operate as a HIV vaccine working through condom use, the total effect must be positive. By decomposing the total effect into a direct and induced effect, our analysis of the transmission mechanism by which education affects condom use will be able to check not only whether female education works as a vaccine, but also, if it is not a vaccine, we can identify what component needs to change in order for female education to become a vaccine.

**3. DATA SOURCE AND SPECIFICATION OF VARIABLES**

The specification for all the variables in our model was largely a function of the data source that we used. For our study, $Y$, $E$ and $P$ are vital for the analysis, so
specifications for these three main variables had to be found and this is why the
data set was selected in the first place. For the individual characteristics, the \( Z_i \)
there was some scope for experimentation as to what to include. We mention
only the ones that were found to be statistically significant.

A. Data

Population Services International (PSI) has existed since 1993 in Tanzania when it
was part of the Tanzania AIDS Project. Their aim was to promote behavior
change among groups at risk of contracting HIV. They provide mass media health
education as well as their own condom brand *Salama*. In 1999, PSI undertook a
male condom survey of 2,533 respondents randomly selected from condom
outlets (such as kiosks, hospitals/clinics, bars/lodgings, pharmacies, retail shops
and wholesale stores) in five major townships in Tanzania and the findings were
summarized in Badru (2000). Questions were asked about condom behavior, as
well as respondent characteristics in terms of region, outlet type where the
purchases were made, current partnership status, relationships with last sex
partner, education, religion and number of assets owned. Since we want to focus
on females, we concentrate on the 585 female respondents in the survey. Other
details of the PSI survey are given in Brent (2009b) and Brent (2010).

The PSI survey is very useful for our purposes as it allows us to analyze
educated females in greater depth and so helps us in our quest to make progress
in explaining the female education–HIV puzzle. Although in recent years in the
HIV/AIDS field there has been a general shift away from the use of purposive
samples towards population based Demographic and Health Surveys (DHS), the
PSI data source does have a distinct advantage over these surveys in the context
of what we are trying to achieve in this article, that is, explain the use or non use
of condoms by reference to the type of sex partner chosen. As pointed out by
Glick and Sahn (2008, p422) in their analysis of condom use based on DHS data
in eight African countries, these surveys do not generally identify the last partner
a person has had sex with. In the PSI data the question as to whether a condom
was used the last time there was sex was followed by a question asking for the
identity of the partner with whom last sex occurred.

B. Specification of variables

involves using condoms.
Our specifications of variables follow directly from the wording of the questions used in the PSI survey. We first present our variable specifications and then explain why they were chosen. The condom model that corresponds to equation (2), which has the two sets of variables of interest \( E \) and \( EP_j \) as well as the controls \( Z_k \), can be expressed as:

\[
last \ time = Y (education, \ education*sex \ partner, \ age \geq 40, \ and \ own \ TV) \tag{4}
\]

Where:
- \( last \ time = 1 \) if the last time the female had sex a condom was used, 0 otherwise;
- \( education = 1 \) if the female had some education, 0 was if the female had no education at all;
- \( sex \ partner \ is \ spouse = 1 \) if the last sex was with the spouse, 0 otherwise;
- \( sex \ partner \ is \ regular = 1 \) if the last sex was with a regular partner, 0 otherwise;
- \( sex \ partner \ is \ just \ met = 1 \) if the last sex was with a person just met, 0 otherwise;
- \( sex \ partner \ is \ casual = 1 \) if the last sex was with a casual partner, 0 otherwise;
- \( age \leq 20 = 1 \) if the female was 20 years of age or less, 0 otherwise;
- \( own \ TV = 1 \) if any member of the female’s household owns a TV, 0 otherwise;

We now explain the rationale for the specifications of these variables that are not the controls:

\( Y last \ time: \) The dependent variable \( Y \) relates to condom behavior. Since the value of condoms as a method of preventing HIV/AIDS is very much a function of their continued use, we defined \( Y \) in terms of the last time sex occurred rather than if a person ever used condoms.\(^{5}\)

In the PSI data for 527 sexually active females there were four categories for the education variable (excluding from examination the very small number of missing cases) and four types of sex partner. Table 1 summarizes the breakdown of these categories in the form of a contingency table:

\(^{5}\) This specification of the dependent variables in terms of whether a condom was used the last time sex took place means that there was only one observation on a partner per person. We have no record of condom use with each partner for multi-partnership females. We therefore could not control for non-observable characteristics of females that determine their use of condoms (such as the degree of trust in a partner) that Luke (2006) was able to control for by being able to apply a fixed effects model when there were multiple observations of condom use for each sex partner of a female.
Table 1: Cross tabulation of categories of education with categories of partner

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Spouse</th>
<th>Regular</th>
<th>Irregular</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>18</td>
<td>9</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>Primary</td>
<td>156</td>
<td>160</td>
<td>48</td>
<td>364</td>
</tr>
<tr>
<td>Secondary</td>
<td>45</td>
<td>44</td>
<td>5</td>
<td>94</td>
</tr>
<tr>
<td>Technical</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>231</strong></td>
<td><strong>231</strong></td>
<td><strong>65</strong></td>
<td><strong>527</strong></td>
</tr>
</tbody>
</table>

Given this breakdown, we can now explain our rationale for the specifications for $E$ and $P_j$. The specification of the interaction terms $EP_j$ directly follows from the specifications for its component parts.

$E$ some education: In the PSI survey, information about the number of years of education completed was not requested; instead the level of education reached was sought. So education was expressed simply as a dummy variable. The four categories of answer to the question as to the highest level of education completed were: no education, primary school, secondary, university/college and vocational training. In Tanzania, most students only complete primary school and this is reflected in our data with 69% being in this category. Given that primary school was the main level reached in Tanzania it makes most sense to include primary as an "educated" category and to define $E$ as having had some education, i.e., every education category except no education was 1, and no education (none) was 0.

Categories 3 and 4, "secondary" and "vocational", could not be used as separate specifications as these categories had so few observations when interacted with the types of partner. Nonetheless to give a more complete picture of the role of education in condom use in Tanzania we group categories
3 and 4 together to form an alternative specification of education, which we call "tertiary". In this alternative specification, which also will be tested in section V, \( E = 1 \) will be defined as more education and \( E = 0 \) for less education will include primary and no education.

**\( P \) sex partner:** In the PSI data set there were four categories of partner identified for the last time the female has sex: spouse, regular partner, someone one just met and casual partner. Many studies have shown that the utilization of condoms is very much related to the intended type of sex partner (regular or not). Luke (2006) has shown that one reason why females do not use condoms is because there is some informal exchange of money gifts for unprotected sex. Because this exchange relation may not exist, or be different, for married couples, she excluded marital relations from her sample. So clearly whether the partner was ones spouse or not would be an important determinant of condom use. We will focus on partner is spouse as the main specification for \( P_j \) and this will be designated as \( P \). Table 1 shows that there are as many observations for partner is regular \( P_k \) as for partner is spouse, so this will be the main comparison group. Finally, by combining partners who have just met and those who are causal to form partner is irregular, \( P_i \), we have sufficient observations to form a second comparison group.

The role of the \( Z \) variables is mainly to ensure that the magnitudes of the coefficients of interest are not affected by excluded variables and that, by improving the statistical fit of the regression, the standard errors of the included variables are kept low. However, there is one \( Z \) variable, i.e., age, which requires special mention as this was the crucial control that generated the vaccine hypothesis in the first place based on Uganda data (which we comment on further in the discussion section). The PSI data set recorded age and we use this with a discrete specification, i.e., \( Z_1 \) **age ≤ 20** is whether the female was aged 20 years or younger, or not. The rationale for this specification follows the logic of the De Walque (2007) study of the role of education in the Ugandan HIV epidemic. De Walque argued that one should expect a cohort effect with education. Those people in the younger generation have been brought up with

---

6 Although primary education was the dominant category, it did not make too much sense to define \( E \) only in terms of primary education as this would involve combining those with no education with those with a lot of education (secondary and technical).

7 See, for example, Deheneffe et al. (1999), Grieg and Koopman (2003) and Hearst and Chen (2004).

8 There was no question in the PSI survey on payments or gifts involved with sexual activity so we could not employ a version of Luke’s (2006) value of transfers variable in our analysis.
the knowledge and realities of HIV in their personal lives. We should expect the young to be more mindful of the harmful effects of not using a condom and therefore more likely to use one. So age ≤ 20 should have a negative sign in the condom use equation.

The summary statistics for all variables used in the estimation are listed in table 2. All of the variables, including the cross-product terms, are dummy variables, so the arithmetic means give the percentage representation in the sample. We see that only just more than a third of the females used a condom the last time they had sex. Also 93% of the sample had some education, 24% had more education and 44% of the time the females last had sex, the partner was their spouse.

Note that the data summary in table 2 is basically for 543 observations as 42 were dropped from the 585 females in the original sample because they were virgins (see footnote 3). A further 26 observations were excluded because data was missing for at least one of the variables that we used. This means that 517 was the sample used for all the main results.

**Table 2. Data Summary**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>0.3460</td>
<td>0.4762</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0.9263</td>
<td>0.2615</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0.2394</td>
<td>0.4271</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_s</td>
<td>0.4383</td>
<td>0.4967</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_r</td>
<td>0.4383</td>
<td>0.4967</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_i</td>
<td>0.1233</td>
<td>0.3291</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_s</td>
<td>0.4041</td>
<td>0.4912</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_r</td>
<td>0.4213</td>
<td>0.4942</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_i</td>
<td>0.1081</td>
<td>0.3109</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_s</td>
<td>0.1082</td>
<td>0.3109</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_r</td>
<td>0.1176</td>
<td>0.3225</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E*P_i</td>
<td>0.0171</td>
<td>0.1297</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Z_1</td>
<td>0.1636</td>
<td>0.3703</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Z_2</td>
<td>0.2056</td>
<td>0.4045</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
4. ESTIMATION FRAMEWORK

Any single equation estimation framework is open to the possibility that selection of the sample into the regression may not be random and this would lead to biased estimates. In addition, the main variables of concern, in our case education \( E \) and the interaction terms \( EP_j \), could be endogenous and hence any single equation estimates used would again be biased. We analyze these two estimation issues in turn.

A. Sample selection

The concern here is with the main focus of the paper, the role of education on condom use. The sample is taken from people who visited or were close to retail outlets where condoms were purchased. It would seem that sampling is directly related to the dependent variable, which is the classical sample selection problem. If for example, education strongly determined a person’s likelihood of going to buy condoms at one of these locations, but education had small effects on condom use conditional on being a person who went to one of these locations, then the regressions would be under-estimating the role of education.

Although plausible, this sample selection concern is unlikely to apply to our estimates. The condom use behavior we are analyzing in this paper relates solely to females. In our sample, the vast majority of the females surveyed did not purchase any condoms that were used. Only 18.4% of the females purchased condoms. Education could affect condom and partner choice, but not determine who is in the sample.

B. Endogeneity

In our equation (2), we have three types of variable: the controls \( Z_k \), the education variable \( E \) and the interactions \( EP_j \). The controls are demographic characteristics and thus have to be exogenous. It is the endogeneity of \( E \) and \( EP_j \) that needs to be examined.

There are two paths to endogeneity. The first is reverse causality whereby \( Y \) is determined by \( E \) and \( EP_j \) and then, in return \( E \) and \( EP_j \) are influenced by \( Y \). The second path is through a third variable \( T \) which affects both the dependent and independent variables and thereby causes them to be correlated. The variable \( T \) is going to be a latent variable. Since it is unobserved...
it will be assumed to be included as part of the error term. We will examine the
two paths for $E$ and $EP_j$ in turn.

We can rule out both paths for endogeneity of $E$ on a priori grounds. Regarding reverse causality, recall that our dependent variable is very time specific, i.e., the last time one had sex, was a condom used. One is not asking whether a condom was ever used. The education variable is also time specific. It is whether the female was educated in the past. Primary education takes seven years in Tanzania. It cannot occur instantaneously. There is no scope for current use of a condom to determine whether a female has been educated or not in the past.

The plausibility of possible endogeneity for $E$ through $T$ depends on whether education acts as a vaccine or not. If $E$ acts as a vaccine, i.e., the total effect of $E$ on condom use is to be positive, then endogeneity would come about by $T$, which is assumed to be ability in the education literature, which we will just call commonsense in our context, that leads to a female deciding in the past to be educated and then deciding now to use a condom the last time she had sex. However, if education has a perverse, negative effect on condom use, which is what we found, then this latent property for $T$ loses plausibility. Commonsense may lead to a female being educated, but why would this same property cause the female not to use a life-saving condom?

On the other hand, using either path, the choice of partner $P_j$ could be considered to be endogenous to whether a condom was used the last time the female had sex. In this case the choice of partner and the choice of whether to use a condom would be jointly determined. A female would choose a partner because of her preference to use, or not to use a condom. Although the choice of partner is not directly an independent variable in any of the estimation equations, $P_j$ does appear as an interaction with the education variable in the form of $EP_j$. If we are correct that $E$ is exogenous, then $EP_j$ might also be exogenous. But, endogeneity is still a possibility. So to allow for this possible endogeneity we will use instrumental variables to replace the interaction terms.

The main instrument used, in addition to the control variables, will be the serial number, serial no, that was used to identify the respondent when the survey was administered. This number was allocated randomly in advance of knowing the sex behavior of the respondent. Because (1) serial no was significantly related to $P_j$ and, (2) being a random variable, is expected to be
unrelated to the error term in the estimation equation, and thus satisfy the so-called exclusion restriction, it meets both of the two requirements necessary for the standard justification for a variable to be a suitable instrument for estimation purposes.

However, over and above the standard justification for an instrument, to be completely satisfied that the instruments we have selected are indeed exogenous, we will subject our instruments to the Walt Test for exogeneity. This test checks whether the estimated residuals from the regression with the instruments replacing $EP_j$ are correlated with the instruments. As long as $n$ times the $R^2$ for the residuals regression does not exceed the Chi-squared value critical value (which depends on the significance level and the number of instrumental variables from outside the equation minus the total number of endogenous variables, which is 1 in our case), the instruments are judged exogenous.

To summarize our estimation strategy to allow for the possibility of endogeneity of the interaction term: We will supply two sets of coefficient estimates. In the first set, which uses Probit for the estimation, we assume that all the right-hand side variables are exogenous. In the second set, which relies on Probit with endogenous regressors, what Stata software calls IVprobit, we use serial no and $Z_1$ and $Z_2$ as instruments for the partner interactions $EP_j$. We present both sets of estimates so that the two sets of results can be compared. After each IVprobit equation we present the results for the Wald test confirming the exogeneity of the instruments that we have used.

5. RESULTS

Table 3 presents the main results. The coefficient estimates are the Probit marginal effects, i.e., discrete changes in probabilities from the various dummy variables as expressed by equations (A7a) and (A7b) in the appendix. The direct effect corresponds to an estimate of $\beta_1$ in equation (3). The induced effect is the estimate of $\beta_j P_{ji}$ in the same equation, which is always equal to $\beta_j$ given that $P_{ji} = 1$ for the particular partner being chosen. And the total effect is the sum of the two, which corresponds to the LHS of equation (3). The top part of the table has the Probit estimates and the bottom part has the IVprobit estimates (i.e., Probit marginal effects with the interaction terms treated as endogenous regressors).
Table 3: Probit estimates of the marginal effects of some education on condom use (last time), \( p \) values in brackets

<table>
<thead>
<tr>
<th>Equation</th>
<th>Direct Effect</th>
<th>Induced Effect</th>
<th>Total Effect</th>
<th>Pseudo R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular</td>
<td>Irregular</td>
<td>Spouse</td>
<td></td>
</tr>
<tr>
<td>Probit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1a)</td>
<td>0.1098</td>
<td></td>
<td>+ 0.1098</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.158)</td>
<td>(0.158)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2a)</td>
<td>-0.0027</td>
<td>0.2618**</td>
<td>+ 0.2591**</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.975)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>(3a)</td>
<td>0.0467</td>
<td>0.5217**</td>
<td>+ 0.5684**</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>(0.547)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>(4a)</td>
<td>0.2312**</td>
<td>-0.4799**</td>
<td>-0.2487**</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
</tbody>
</table>

IVprobit

<table>
<thead>
<tr>
<th></th>
<th>Wald test ( p ) values</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(2b)</td>
<td>-0.1171</td>
<td>+ 0.4235**</td>
</tr>
<tr>
<td></td>
<td>(0.210)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>(3b)</td>
<td>-0.1253</td>
<td>+ 0.5195**</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>(4b)</td>
<td>0.2541**</td>
<td>-0.2908**</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

Notes: For Probit: All equations have variables \( Z_1 \) and \( Z_2 \) as controls. The total effect is the sum of the direct and induced effects and the \( p \)-value is for the Chi-square test that the sum is equal to zero. For IVprobit: All equations have serial no as the instrument for the interaction term included in the equation. In addition, all equations have variables \( Z_1 \) and \( Z_2 \) as instruments. So the Wald Test checks whether the set of instruments serial no, \( Z_1 \), and \( Z_2 \) are endogenous or not. The \( \rho \) values for the Wald Test relate to the Chi-square statistic testing exogeneity. * Significant at 5% level; ** Significant at 1% level.

Regression equation (1) in the top and bottom of table 3 shows the impact of education \( E \) on condom use on its own with just the two \( Z \) variables as controls. Regression equations (2) to (4) have the effects of \( EP_j \) as additional independent variables. Thus regression equation (2) has \( E \), \( EP_R \) and the two \( Z \) variables. Regression equations (3) and (4) have the same set of variables except that they replace \( EP_R \) by \( EP_I \) in regression equation (3), and by \( EP_S \) in regression equation (4). We first cover the Probit results and then the IVprobit results.

A. Probit estimates
On its own in regression (1a), where the direct effect and the total effects are one and the same, *some education* is not significantly positively related to condom use. Education and the two controls explain hardly any of the variation in condom use. The fit (the pseudo $R^2$) increases considerably once we allow for the transmission mechanism of *some education* to work through the choice of sex partner. The fit rises to 6% when $EP_r$ is added in regression (2a) and to 9% in regression (3a) with $EP_r$ added. The best fit of 22% comes when the partner is spouse $EP_S$ is involved as reported in regression (4a).

When the fit is not so high, i.e. regressions (1a) to (3a), the direct effect is not significant. The induced effect is positive in regressions (2a) and (2b), making the total effect significantly positive. A Chi-squared test that these positive total effects are not different from zero can be rejected at the 1% level. However, in regression (4a) where the fit is highest, the results are radically altered. The direct effect is now significantly positive, but the induced effect is significantly negative. The induced effect is so strong that it completely reverses the direct effect, making the total effect negative. This negative total effect of $-0.2487$ is significant at well within the 1% level.

What the results from the top part of table 3 tell us about the education as a vaccine hypothesis is this. When the fit is not high, there is no evidence that the direct effect of *some education* has any vaccine properties by increasing condom use. Only when the transmission mechanism is allowed for as proposed in this paper, does *some education* seem to act like a vaccine through the induced effect. But this is true only when the partner is regular or irregular. It is not the case when the partner is the spouse. When the spouse is the female’s partner, the direct effect is positive, but completely swamped by the negative induced effect, making the overall impact of education perverse. So the main finding based on regression (4a), which gives by far the best overall explanation of condom use, is that for the subset of females whose last sex is with their spouse, education is not a vaccine as it actually decreases condom use and this helps solve the female education-HIV puzzle.

Note that the sign of *age ≤ 20* variable, $Z_1$, was not positive in any of the regression regressions as was expected. In fact $Z_1$ was significantly negative at the 1% level in (4a). The younger cohort was not more likely to use condoms as was the case in Uganda. As we shall see in section VI, this is one of the main reasons why the Tanzanian HIV experience was different from that in Uganda.

**B. IVprobit estimates**
There is no IVprobit counterpart to regression (1a) as there is no interaction term in the regression which could be endogenous. Otherwise, the results for regressions (2b) to (4b) exactly mirror those of regressions (2a) to (4b) and so all the findings and conclusions carry over when IVprobit is used instead of Probit for the estimation. In all three regressions the Wald test that the instruments used to predict the interaction terms (i.e., serial no, Z₁ and Z₂) are exogenous cannot be rejected even at the 10% level and higher.

C. The direct effect, induced and total effects of more education on condom use.

As table 3 reveals, there were 119 (24%) of the females in our sample who completed tertiary education, so there was sufficient data to be able to test whether the alternative specification of E as more education would give different results as for some education. The testing strategy exactly mirrors that for some education in sections A and B, so we can go straight to the results. Table 4 is the more education counterpart to the some education results given in table 3.

Table 4: Probit estimates of the marginal effects of more education on condom use (last time), ρ values in brackets

<table>
<thead>
<tr>
<th>Equation</th>
<th>Direct Effect</th>
<th>Induced Effect</th>
<th>Total Effect</th>
<th>Pseudo R²</th>
<th>Wald test ρ values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Regular</td>
<td>Irregular</td>
<td>Spouse</td>
<td></td>
</tr>
<tr>
<td>Probit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1a)</td>
<td>0.1364**</td>
<td></td>
<td></td>
<td>+0.1364**</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>(2a)</td>
<td>–0.0437</td>
<td>0.3449**</td>
<td></td>
<td>+0.3012**</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.000)</td>
<td></td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>(3a)</td>
<td>0.0992</td>
<td></td>
<td>0.5112**</td>
<td>+0.6104**</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>(4a)</td>
<td>0.3105</td>
<td></td>
<td>–0.3250**</td>
<td>–0.0145</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
<td>(0.792)</td>
<td>(0.792)</td>
<td></td>
</tr>
<tr>
<td>IVprobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2b)</td>
<td>–0.3612</td>
<td>0.6748**</td>
<td></td>
<td>+0.3136**</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>(3b)</td>
<td>0.2237**</td>
<td></td>
<td>–0.5106**</td>
<td>–0.2869**</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
<td>(0.000)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>(4b)</td>
<td>0.4622**</td>
<td></td>
<td></td>
<td>–0.1031</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
<td></td>
<td>(0.388)</td>
<td></td>
</tr>
</tbody>
</table>
Notes: For Probit: All regressions have variables $Z_1$, and $Z_2$ as controls. The total effect is the sum of the direct and induced effects and the $p$-value is for the Chi-square test that the sum is equal to zero. For IVprobit: All regressions have serial no as the instrument for the interaction term included in the regression. In addition, all regressions have variables $Z_1$, and $Z_2$ as instruments. So the Wald Test checks whether the set of instruments serial no, $Z_1$, and $Z_2$ are endogenous or not. The $p$ values for the Wald Test relate to the Chi-square statistic testing exogeneity. * Significant at 5% level; ** Significant at 1% level.

Overall the model fit for Probit is much lower for more education than we found for some education. Otherwise the top part of table 4 reports remarkably similar results to the top part of table 3. The signs of all variables are exactly the same. Again regressions (1a) to (3a) suggest a positive total effect, which is completely reversed by the total effect in regression (4a).

For the IVprobit results, regressions (2b) and (4b) give coefficients signs and significance levels in table 4 that are exactly the same as for (2b) and (4b) in table 3. The only real difference is that the negative total effect for (4b) in table 4 is now no longer significant. The main differences come from comparing the estimates for (3b) in tables 3 and 4. The direct effect is significantly positive in table 4 when it was insignificant in table 3. But more surprising is the fact that the induced effect is now significantly negative in table 4 when it was significantly positive in table 3. So for more education under IVprobit, the total effects of both irregular partners and spouses are perverse, while for some education the perverse effect was only found for partners who were spouses.

6. DISCUSSION
In this section we present two main policy implications of our results as they relate to the female education and HIV/AIDS puzzle. We examine whether education, basically in the form of some education, can be viewed as a HIV/AIDS vaccine at this time and why the puzzle exists in Tanzania and not Uganda.

A Education as a vaccine for HIV/AIDS in Tanzania
The Global Campaign for Education (2004, section 2) which endorsed the idea of female education as a vaccine also explained how the vaccine works: “A complete primary education leads to increased ability to evaluate, understand and apply facts; gains in confidence; and greater decision-making power in relationships. It creates the context in which preventative messages about HIV can best be understood and acted upon.” A major preventative message has been the one to increase condom use to reduce the risk of being infected by HIV/AIDS. Hence,
they point out, "better educated young people have increased condom use and reduced the number of casual partners at a much steeper rate than those with little or no education." There is no doubt that the global campaign is generally correct on this point about condom use. A recent survey by Glick and Sahn (2008) confirms that females were more likely to use condoms with casual partners, and our results for equations (3a) and (3b) confirmed this relationship for Tanzania.

However, missing from the vaccine story is the flip side. What happens to condom use by educated females when they have sex with their spouses? For our Tanzania study, we had data on the sex partner when the last time the female had sex. So we were able to analyze condom use with spouses, regular partners who are not spouses, and irregular partners (just met and casual partners). We found that female use of condoms with their spouses was very different from their behavior with the other two types of partner. If we ignore the case when the partner is a spouse, as with regressions (1a) to (3a), and (2b) to (3b) in table 3, we get the impression that the claim that education is a vaccine is valid. However, the overall finding is reversed in regressions (4a) and (4b) when the partner is spouse variable is included. This result comes from a regression that has roughly double the explanatory powers of the other regressions where the spouse is not the sex partner. We found that overall, condom use goes down by 29% when a sexually active female is educated and has their spouse as the sex partner.

It is because \( \beta \) was so influential and strongly negative in our condom use study that we found in regressions (4a) and (4b) that, by \( E \) impacting it through the interaction term \( E^*\beta \), female education ended up reversing any positive direct effect of \( E \) on condom use. The total effect of \( E \) was to reduce condom use and thus ensure that female education was not an effective vaccine against HIV/AIDS in Tanzania. Clearly it is not a sufficient condition for a vaccine to exist that female education lead to more condom use with non-spouse partners; condom use with spouses needs to be a priority as well.

There is some additional information in the PSI data set that we can use to help us understand the findings we obtained. Interestingly, the effect of education on condom use via \( \rho_j \) was not due to educated females being more likely to be married. 41.63% of women with some education were married and this was almost identical to the 41.88% marriage rate for all women in the sample. Rather there is something within the marriage relationship itself that works against condom use. To help identify what this missing factor may be, we can
refer to a follow-up question that PSI used for those, nearly 50% of the sample, who said that they never used condoms. The number one reason (in 88 of 233 cases) why women said that they never used condoms was because they trusted their partner. When we break down this overall response by some education, we find that 39% of educated females did not use a condom because they trusted their partners, while only 26% of non-educated females were so trusting of their partners. As this sub-sample size is less than half of our total sample we cannot draw definitive conclusions as to why the vaccine did not exist in Tanzania. But it does seem that, if non-educated females did not use condoms because they trusted their partners, then educated females did not use condoms because they were even more trusting of their partners.

Our discussion so far has been when some education was the education specification. We added to our analysis the effects of more education on condoms use, defined as completing secondary school or technical college. The results largely mirrored what we found with some education as the E specification. More education increased condom use when the sex partner is a spouse is ignored and it decreased condoms use when the female's sex partner was her spouse. Condom use fell by 10% if a sexually active female has more education and her sex partner is her spouse. Again, one cannot claim that education is a vaccine even for this more educated group in Tanzania.

B The role of education in Tanzania and Uganda compared

It is useful to compare the findings of our study for Tanzania with that of the experience of Uganda. Uganda is often cited as a success story in controlling HIV/AIDS, see for example Pisani (2008), while Tanzania is typical of many of the SSA African countries where the epidemic is still rising. De Walque et al. (2005)'s finding of a positive gradient between education and HIV for 1989-90 turning negative 11 years later was obtained in a context where HIV was falling over time and condom use was rising. It was due to female education that they attributed the increased condom use. They argued that education helps individuals access and process health related HIV information that was a part of the infection intervention strategy of Uganda. The educated female was hypothesized to translate that information to a reduction in risk-taking behavior, i.e., increased use of condoms.

The transmission mechanism by which the rise in female education led to increased condoms use in Uganda was not tested empirically. Our study specified and quantified this missing transmission step. Unlike Uganda, condom
use in Tanzania was roughly the same at the beginning of the 1990s decade as it was at the end, see MEASURE (2001). Thus our study operated in a context different to that of Uganda. With female education increasing over the 1990’s in Tanzania, and condom use remaining constant, it should not have been surprising that we were analyzing a sample where female education and condom use were negatively related.

Given that HIV prevalence rates were declining over time in Uganda, it could be expected that De Walque et al. would find that age was the crucial control variable in their study. Thus it was only the age group 18-29 at the end of the 11 year period that had significant negative gradients between education and HIV prevalence and not those 30 and older. De Walque et al. put emphasis on the age effect because it was only the younger group who could have benefited from the information and prevention programs concerning HIV/AIDS. Older cohorts developed their sexual behavior prior to these campaigns. In Tanzania there was no decline in HIV prevalence over time. Consequently there should have been no such dynamic cohort age effect in this country.

De Walque (2007) in a second study related to Uganda’s education intervention revealed that he was aware of the literature that found a positive relationship between HIV and secondary education. His view was that this was a spurious correlation because these two variables, education and HIV, were determined by a third factor, that is, age. Thus young individuals were more likely to have some secondary education and, at the same time, were more likely to be HIV positive. This reasoning seemed to fit the experience of Uganda because, although he also found at first a perverse positive effect between secondary education and HIV in his data set, subsequently, when he controlled for age, the positive association between HIV and schooling disappeared.9

However, we found that an age effect could not explain away the education-HIV puzzle in Tanzania. In fact, age confounded the problem. We included being under 20 as a separate independent variable in all our regression equations explaining variations in condom use. Instead of increasing condom use among the young females, as De Walque et al. would have expected, the sign of age ≤ 20 was significantly negative in our main result. The inclusion of an age variable did not prevent the sign of the net relation between education and

---

9 See De Walque (2007, footnote 15). The literature he was referring to was the earlier survey by Hargreaves and Glynn (2002) which found that only one study existed where the relation between education and HIV prevalence was negative.
condom use being negative. So, once again we see that the Ugandan and Tanzanian experience is very different.

7. SUMMARY AND CONCLUSIONS

What is it about the education of females that leads to an increase and not the expected decrease in HIV/AIDS? This is a puzzle. An important first step in resolving this puzzle is to find out the sexual risk behavior of educated females. To the extent that some of those educated are less likely to use condoms one would have one piece of the puzzle. The next piece would be to explain why it is that these females do not use condoms. In this paper we indeed see that in Tanzania, many with some education had approximately a 29 percent lower probability of using condoms than those without such schooling. To help understand how education influences condom behavior, the coefficient on the sex partner $P_j$ was made a direct function of education. This produced an estimation equation with an interaction term $EP_j$. This interaction term enables us to specify the transmission mechanism from education to the use of condoms.

Decomposing the total effect of education into a direct effect via education on its own and an induced effect via the interaction terms was important because it showed that the sign of the overall effect was not a function simply of the direct effect. The direct effect often gave the impression that female education was acting as a vaccine, which was true only when the sex partner was regular or irregular, not the spouse. When the sex partner was the spouse the overall effect of education was negative.
The findings in this paper carry considerable significance for HIV policymakers. Clearly a country like Tanzania, experiencing a positive relationship between female primary education and HIV, now must have already realized that it has to target females in schools to get their behavior to change. What this study highlights is what to focus on to ensure that the necessary behavior change takes place. One reason why educated females are more likely to be infected by HIV is because they are less likely to use condoms with their spouses than other females. In a sample where 44% of the females last had sex with their spouse, this non-use of condoms with their spouses is quantitatively important. While females with some education already are more likely to use condoms with a non-spouse partner, this preference needs to be reinforced and encouraged further when the partner is a spouse. The message needs to be with condom usage that an educated female need not be more trusting when the partner is a spouse than are other females with a spouse.

The rationale for this paper fits exactly within the general empirical framework suggested by Deaton (2010) requiring that empirical studies be redirected to evaluating effect mechanisms and not just the effects themselves. He argues that too much emphasis has been placed on random controlled trials because they allegedly show ‘what works’; finding what works ‘is not informative about mechanisms, if only because there are always multiple mechanisms at work’ that generate the project effect. He adds: ‘the analysis of projects needs to be refocused toward the investigation of potentially generalizable mechanisms that explain why and in what contexts projects can be expected to work.’ Thus it may be interesting to know that some countries, such as Tanzania, have the education-HIV puzzle, while other countries, such as Uganda, do not have the puzzle, as we have shown in this article. But more importantly, this study fits in the new research agenda proposed by Deaton. Our aim was to explain why this perverse, positive relationship between female education and HIV would take place.

The transmission mechanism identified here involved the use, or non-use, of condoms with a female’s various sex partners. We found that educated females, like less educated females, were highly likely to use condoms with casual partners, or partners they had just met. However for partners that were their spouses, educated females were less likely to use condoms than their less educated counterparts, and this was one reason why the female education – HIV puzzle existed. Knowing this partner transmission mechanism should help lead to more effective HIV interventions for this particular demographic group. This is
especially true given that, as we have shown, the transmission mechanism consisting of age that operated in Uganda did not apply in the Tanzania context.
REFERENCES


Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS), and ORC Macro (2005): *Tanzania HIV/AIDS Indicator Survey 2003-2004 (THIS)*. Calverton, Maryland, USA
BRENT: THE CONDOM PIECE OF THE PUZZLE


APPENDIX

The appendix fills in the theoretical framework that is summarized in section II. There are three sections. First we present the random utility model as applied to condom use. Then we formulate the condom utility determinants which produces the estimation equation (1) in the text. The direct and indirect effects of education can be derived from the coefficients in the estimation equation and this leads to equation (2) in the text. Lastly we explain how the marginal effects that make up the direct and indirect effects in equation (2) can be estimated when there is an interaction term comprising two dummy variables.

A. Random Utility Model

Consider a sexually active female $i$ who faced two choices:

- $Y_i = 1$ was to use a condom the last time the female had sex, and
- $Y_i = 0$ was not to use a condom the last time the female had sex.

The utility from making either of these two choices is given by an index (utility function) $U_i$. Assume that $U_i$ depends linearly on a set of outcomes $X_i$ (characteristics) of the choices that are observable and a set of outcomes that are not observable by the researcher $\varepsilon_i$:

$$U_i = X_i'\beta + \varepsilon_i$$  \hspace{1cm} (A1)

Although $U_i$ is not fully observable, we are able to observe the behavior or choices $Y_i$. It is assumed that the female would only choose to use condoms if the utility was positive and would not use them if the utility was negative. We deduce that:

$$Y_i = 1 \text{ if } U_i > 0, \quad Y_i = 0 \text{ if } U_i \leq 0.$$  

The conditional probability that $Y_i = 1$ is:

$$\Pr(Y_i = 1 | X_i) = \Pr(X_i'\beta + \varepsilon_i > 0) = \Pr(\varepsilon_i > -(X_i'\beta))$$

If the probability density function $f(X_i'\beta)$ is normal (which is symmetric):

$$\Pr(Y_i = 1 | X_i) = \Pr(\varepsilon_i < X_i'\beta) = \Phi(X_i'\beta)$$  \hspace{1cm} (A2)

where $\Phi$ is the standard cumulative normal distribution. By inverting (A2):

$$Y_i = 1 \mid X_i = \Phi^{-1}(X_i'\beta) = X_i'\beta$$  \hspace{1cm} (A3)

The random utility model involves running the dichotomous dependent variable regression given by (A3), which provides the Probit score, and then applying $\Phi$
BRENT: THE CONDOM PIECE OF THE PUZZLE

to those scores to obtain the \( \text{Prob} \) values on which one can carry out any comparative statics exercises.

B. Condom Utility Determinants

The determinants of the individual’s condom use will be classified into two categories: the education variable \( E_i \) and the \( k \) individual characteristics \( Z_{ki} \):

\[
Y = X'_i \beta = \beta_0 + \beta_E E_i + \sum_k \delta_k Z_{ki}
\]  
(A4)

Equation (A4) is equation (1) in the paper.

The education transmission mechanism works through the coefficient \( \beta_E \). Let \( \beta_E \) be linearly related to the choice of \( j \) sex partners \( P_{ji} \) in the form:

\[
\beta_E = \beta_1 + \beta_j P_{ji}.
\]

Substituting for \( \beta_E \) in (A4) produces:

\[
Y = X'_i \beta = \beta_0 + \beta_1 E_i + \beta_j [E_i P_{ji}] + \sum_k \delta_k Z_{ki}
\]  
(A5)

Equation (A5) is equation (2) in the paper.

From (A5) we derive:

\[
\frac{\partial Y_i}{\partial E_i} = \frac{\partial}{\partial E_i} X'_i \beta = \beta_1 + \beta_j P_{ji}
\]

(A6)

Equation (A6) is equation (3) in the paper.

The total effect of changes in the level of education on the use of condoms in (A6) is split up into a direct effect (the first term on the RHS) and an induced effect (the second term on the RHS) which works through the sex partner variable, where \( E_i P_{ji} \) would typically be set equal to the sample mean \( \overline{E_i P_{ji}} \).
The marginal effects when the interaction terms comprises two dummy variables

Equation (A6) applies in the context of $E P_j$ being a continuous variable. In our application, the education variable $E_i$ and the sex partner variables $P_{ji}$, will take the form of dummy variables. This means that the interaction terms $E P_j$ will also be dummy variables. To convert the regression coefficients in (A6) into marginal probability effects we need to consider discrete changes in the value of the probability function for unit changes in the independent variables. The education variable will be changed in a discrete fashion from $E_i = 0$ to $E_i = 1$. In the absence of the interaction term, the effect on $Prob_i$ of using a condom for a discrete change in female education would be found from: $\Phi(X_i' \beta)_{Ei=1} - \Phi(X_i' \beta)_{Ei=0}$, where all the $Z_k$ instruments would be set at their sample means $\overline{Z_k}$. However, because changes in $E_i P_{ji}$ are inherently linked to changes in $E_i$ and $P_{ji}$, we need to specify how the latter two variables will alternatively take the values of 0 and 1 when the two discrete changes occur simultaneously.

Since the focus of this article is on the effect of the sex partner for an educated female on her condom use, we will set $E_i = 1$ irrespective of whether $P_{ji} = 1$ or $P_{ji} = 0$. For the direct effect when $E_i = 0$, we will also keep $P_{ji} = 1$. This means that the direct effect will be estimated by: $\Phi(X_i' \beta)_{Ei=1, P_{ji}=1} - \Phi(X_i' \beta)_{Ei=0, P_{ji}=1}$. For the induced effect, when $E_i P_{ji}$ is 1, $P_{ji}$ must also be 1 when $E_i = 1$. However, when $E_i P_{ji}$ is 0, $E_i$ can be 0 or 1. To resolve this ambiguity, for the interaction $E_i P_{ji} = 0$, we will set $P_{ji} = 1$ when $E_i = 0$. The induced effect will be: $\Phi(X_i' \beta)_{Ei=1, E_i P_{ji}=0} - \Phi(X_i' \beta)_{Ei=0, E_i P_{ji}=0}$. The result of all of this is that our estimation of the direct and induced effects for a discrete specification for changes in $X_i' \beta$ that is based on equation (A5) simply becomes:

Direct effect: $\Phi[\beta_0 + \beta_1 + \sum_k \delta_k \overline{Z_k}] - \Phi[\beta_0 + \sum_k \delta_k \overline{Z_k}]$

(A7a)

Induced effect: $\Phi[\beta_0 + \beta_1 + \beta_j + \sum_k \delta_k \overline{Z_k}] - \Phi[\beta_0 + \beta_1 + \sum_k \delta_k \overline{Z_k}]$

(A7b)
Thus the coefficients in equation (A6), which is equation (3) in the paper, were estimated on the basis of equations (A7a) and (A7b).