Do the More Educated Know More about Health? Evidence from Schooling and HIV Knowledge in Zimbabwe

Jorge M. Agüero
University of Connecticut

Prashant Bharadwaj
University of California, San Diego

I. Introduction
The relationship between education and health has received a lot of attention from researchers (Grossman 2006). While the positive association between health and education is well noted, recent work by economists has examined whether there exists a causal relationship (Lleras-Muney 2005; de Walque 2007a; Clark and Royer 2010; McCrary and Royer 2011). A recent review by Cutler and Lleras-Muney (2010) suggests that there is little consensus on whether such a causal relationship exists. Moreover, even in instances in which researchers have found a causal link, we know little about the mechanisms behind this relationship. In the developing world, policy focus has been on improving education and health. In this context, it is imperative to understand whether an important externality of education results in better health investments.

To fill this gap, this article explores a fundamental channel linking education and health in the important setting of HIV in Africa: knowledge of HIV.1 Reducing HIV prevalence and increasing the proportion of people with comprehensive and correct knowledge of HIV are major issues in public policy and are two of the specific targets of the Millennium Development Goals. Our article explores whether the educated engage in HIV-preventing behavior and whether they know more about how HIV is transmitted.2 In addition,
we examine where the more educated might obtain information pertaining to HIV transmission. We study these questions in Zimbabwe, one of the five countries in sub-Saharan Africa with the largest HIV epidemic and where the estimated adult HIV prevalence rate is a staggering 14.3%.³

Obtaining a causal interpretation for the positive correlation between education and health is fraught with difficulty, and as a result, very few papers exist that examine mechanisms that might link the two.⁴ In order to examine this within a causal framework, we apply an instrumental variables approach that uses age-specific exposure to an education reform in Zimbabwe. In 1980, after independence, the education system in the country was reformed and the restrictions limiting the advancement of black Zimbabweans toward secondary school were eliminated, leading to a dramatic increase in secondary school enrollment (see fig. 1). The transition rate from primary to secondary school jumped from 27% in 1979 to over 80% in 1980. Since secondary school enrollment begins at age 14, we observe a large difference in completed education for people who were 15 and below as of 1980 as compared to people who were just 16 and over in 1980 (and as a consequence were too old and therefore less likely to be enrolled in secondary school).

It is important to note that plenty of changes occurred in Zimbabwe at the time of independence. Our identification strategy relies on the fact that the education reform created a discontinuity that disproportionately affected those age 15 or less in 1980 compared to children age 16 or more. Thus, while independence might have also brought about other social reforms, the underlying assumption for our article is that these other reforms did not affect children differentially around the age cutoffs that we examine. We provide evidence in favor of this argument by showing that health outcomes determined before the reform (such as height) do not show a discontinuity at age 15. Furthermore, we do not observe a discontinuity in completed years of education by age in 1980 in other neighboring countries.⁵

---

⁴ Lange (2011) is one of the few recent papers that examines how and why educated women in the United States might make different investment decisions with regard to breast cancer screenings (although the variation in education in Lange’s context is not exogenous). He finds evidence that the educated are more likely to include variation in risk factors when reporting personal cancer risk. This is supported by the fact that the more educated appear more likely to believe in scientific evidence regarding breast cancer. See also Clark and Royer (2010), Cutler and Lleras-Muney (2010), and Altindag, Cannonier, and Mocan (2011) for more details.
⁵ These findings complement the results from Agüero and Ramachandran (2014), who use the 2002 Zimbabwe population census and show that the there is no discontinuity in the educational at-
Given the greater vulnerability of women to HIV and data considerations that we discuss in Section III, this article focuses on the impact of education on HIV behavior and knowledge for women. We find that girls who were 15 and below in 1980 obtain 1.8 extra years of schooling and are 50% more likely to have attended some secondary school than their slightly older counterparts. Using this discontinuity as an instrument for completed education, we find strong evidence that the educated adopt less risky sexual behaviors by having fewer sexual partners. Data on actual HIV status are available only for a small subset of the observations, and while our two-stage least-squares results suggest that the more educated are less likely to be HIV positive, these results are not statistically significant. Our data allow us to explore the linkages between education and health, and we find that the more educated know more about HIV and also know more about how to mitigate its transmission. Having gone to secondary school increases the probability of having comprehensive knowledge about HIV (i.e., knowing that a healthy person can have HIV, that HIV cannot be spread via mosquitoes, and that HIV transmission can be mitigated by using condoms and having one sexual partner) by nearly 10%.

Figure 1. Secondary school enrollments in Zimbabwe, 1973–95. Note: Elaborated based on data from UN Statistical Yearbook, various years.
People who had an extra year of education are 4.6 percentage points more likely to know that a condom reduces the chances of getting HIV and are also 5 percentage points less likely to have misconceptions about how HIV is spread (e.g., knowledge that HIV cannot be spread via mosquitoes or food).

How might the more educated obtain knowledge about HIV? Secondary school (grades 8 and up in our context) is typically the time when sex education is taught in schools in many parts of the world; however, it is unlikely that sex education, especially with regard to HIV, was being taught in secondary schools in Zimbabwe in the 1980s. By many accounts, the National AIDS Coordination Program was only set up in 1987, and the country’s first HIV and AIDS policy was announced in 1999. We explore this issue by examining the use of various media outlets by educational status. We find that attending secondary schooling increases the likelihood of reading newspapers by 18%. This is an important finding, as during the late 1980s and early 1990s, most of HIV/AIDS awareness was via newspapers. Thus, we have suggestive evidence indicating that access to media outlets plays a role in how the educated manage to get more information on HIV. This finding is consistent with previous results in the literature that explore the mechanisms explaining why mothers’ education improves children’s health. For example, Thomas, Strauss, and Henriques (1991) showed in the context of Brazil, that almost all the observed association can be explained by mothers’ access to information, such as reading papers, watching television, and listening to the radio. In the case of Morocco, Glewwe (1999) argues that despite the fact that health knowledge was not directly learned in school, it is the most important skill that mothers (indirectly) obtain from their schooling.

This article contributes to the broader economic literature on education and health by examining an important mechanism behind the relationship. Perhaps most closely related is the work of de Walque (2009), which highlights trends in the relationship between education, HIV status, and knowledge in five African countries. De Walque highlights that schooling is a consistent predictor of health behavior and knowledge. Similar to some of the

---

6 Writing in 1993, Pitts and Jackson (1993, 225) note that “national newspapers are the most important source of information reliably reaching the largest number of people. Newspapers have, in fact, been cited by social workers as the major source of information on AIDS and HIV in Zimbabwe.”

7 While somewhat limited by data given the setting and the empirical design, we also explore other channels of obtaining information such as spousal education, access to antenatal clinics and family planning services, and receipt of health insurance. While we find some support for the spousal education channel, we find little support for other ways of obtaining information. We realize that there could be other channels that drive the link between education and health knowledge and in future work plan to explore these channels in greater depth.
evidence presented in this article, de Walque finds that “education level predicts protective behaviors such as condom use, use of counseling and testing, discussion of AIDS between spouses, and knowledge about HIV/AIDS” (209). While broad in its scope, de Walque (2009) does not use exogenous variation in education to establish causal relationships. While many recent papers have examined the role of information in inducing behavior change with regard to HIV (de Walque 2007b; Thornton 2008; Dupas 2011), these papers do not tackle the question of whether the educated possess more information about HIV. Understanding this link is a first step toward uncovering the ways in which education can affect health. In addition, we are able to provide some evidence of how the more educated obtain more knowledge about HIV. Linking the use of various media outlets to educational status within our instrumental variables framework is a novel addition of our article.

II. The 1980 Education Reform in Zimbabwe

In April 1980, the newly elected government of Zimbabwe reformed the education system to break with the apartheid-like regime that prevailed in Rhodesia. Before 1980, at least 25% of black school-age children failed to enter primary school due to a lack of places (Riddell 1980). For example, in 1976, for every 1,000 black school-age children, 250 never started school. Of those who went to school, 377 graduated from primary school, but only 60 of them transitioned into secondary education. Thereafter, 37 reached Form IV and less than 3 reached lower Form VI (Nhundu 1992, 79).

The 1980 education reform has been widely documented in the literature (e.g., Dorsey 1989; Edwards and Tisdell 1990; Edwards 1995). As described in Nhundu (1992), there were four key initiatives undertaken by the new government: (1) the introduction of free and compulsory primary education, (2) the removal of age restrictions to allow overage children to enter school, (3) community support for education, and (4) automatic grade progression, in particular from primary to secondary school. An immediate impact of these steps was an enormous increase in school enrollment. Between 1979 and 1985, total enrollment rose from 885,801 to 2,698,878: an unprecedented 205% increase (Nhundu 1992, 82). The greatest expansion took place in

---

8 For a history of Rhodesia’s education system and the policies dictating the quantity and quality of schooling Africans received, see Atkinson (1972) and O’Callaghan and Austin (1977).

9 Zimbabwe’s education system consists of primary education, secondary education, and tertiary education. The primary level is a 7-year cycle, and the official entry age is 6 years. It runs from grade 1 through grade 7. Primary education leads to a grade 7 certificate. Secondary education is divided into three 2-year levels: junior, middle, and high/advanced. Entering high/advanced secondary school requires the student to pass the “O”-level examinations.
secondary education, where enrollment grew by 628% during the same period. As figure 1 shows, gross enrollment in secondary schools climbed from 66,215 in 1979 to 482,000 in 1985, peaking at a little over 700,000 in 1991.

To accommodate the increased demand, the government built new schools and undertook extensive reconstruction and expansion of existing facilities. This increase is shown in figure 2. Between 1979 and 1983, the overall number of schools grew by 90%. Primary schools increased by 65%. Again, the largest increase is found in secondary schools: they grew by 575% since 1979. These figures are consistent with an increase in the budget allocated to education. In the fiscal year 1979–80, the share of education was 11.6% of the national budget. It almost doubled in 1980–81 (22.1%) and remained at about 17% until 1986–87 (Dorsey 1989). The early years of the reform focused on opening new secondary schools, especially in rural areas. The target was to provide a secondary school within walking distance of all rural pupils, especially where geographic and demographic factors were conducive.10

Mirroring the massive response in enrollments are the transition rates from primary to secondary in Zimbabwe’s schools. As figure 3 shows, the transition rate from grade 7 (last grade of primary education) to Form I (first grade of secondary education) remained below 30% throughout the ’70s. Beginning in 1980, the year of the reform, it jumped to 87% and averaged 70% for the rest of the decade.

Children start primary school at age 7, thus on-time completion of all primary grades would enable them to start secondary school at age 14. As shown in figure 3, there is a clear discontinuity in the probability that a child (boys and girls together) would go to secondary school in 1980. A child graduating from primary school in 1979 had a one-in-five chance of enrolling in secondary school. The same child, but who graduated 1 year later in 1980, was four times as likely to enroll in secondary education. Also, the elimination of age restrictions allowed many overage children to remain in or return to school. For example, while there were 112,890 children enrolled in grade 6 in 1980, the number of children enrolled in grade 7 the following year was 15% larger (over 129,000). Thus, the benefits of the reform extended to children age 15 in 1980 in addition to those age 14 or less in 1980. Therefore, the educational reform of 1980 provides a natural experiment,

10 Agüero and Ramachandran (2014) explored whether the reform, by opening a large number of schools, created a spatial rollout similar to the Indonesia experiment studied by Duflo (2001, 2004). While some districts indeed opened schools sooner than others, by 1983, only 3 years after the reform, the disparities were eliminated. Comparing the educational attainment of those born in districts where secondary schools opened “earlier” to those born in districts that opened schools “later” shows no difference.
where for reasons exogenous to her choice, an adolescent could eventually acquire more schooling.

III. Data

The data used in this article come from the Demographic and Health Surveys (DHS) of Zimbabwe. The DHS are standardized nationally representative (cross-sectional) household surveys in developing countries, where the main unit of analysis comprises women between age 15 and 49. These women answer a long questionnaire about their birth history, fertility preferences, family planning, and socioeconomic and marital status, among other characteristics. Since the mid-1990s, the DHS have asked women and their partners about their knowledge of HIV as well sexual practices. Men are interviewed only if they are the spouse or partner of the respondents. Therefore, the DHS provide us with a random sample of women but not men. In the case of Zimbabwe, only the 1999, 2005–6, and 2010–11 DHS contain information on HIV knowledge, and we focus on these surveys. Specifically, for each survey year, the individual recode component of the survey (the component that pertained to females) was used. Blood sample data, which were tested for HIV, were collected in 2005–6 and 2010–11; hence, the sample for examining actual HIV status is smaller.
Since the data sets contain year of birth and completed education as of the survey year, we are able to construct age as of 1980 (the beginning of the reform) and relate it to completed schooling for men and women. We find that people 15 or below as of 1980 obtain significantly more years of education, as well as having a higher probability of attending secondary school. The discontinuity is apparent in figures 4a and 4b.

Table 1 provides some of the details regarding sample construction. Combining the three relevant rounds of the DHS, and restricting our attention to women who were between age 8 and 21 in 1980, gives us a sample of around 6,034 women. After omitting certain observations due to outliers in years of schooling and completeness on covariates like urban status and region of residence, we are left with a sample of 5,985 women across the 3 survey years.11 Since the HIV blood sample was only conducted in the two most recent surveys, our sample size with relevant HIV data is 3,391 women.

IV. Empirical Strategy
We consider the following relation between education and HIV-related outcomes:

11 We drop people with strictly greater than 16 years of schooling. This constitutes about 0.79% of the total data. Including them does not make a difference to our results.
Figure 4. Years of education in Zimbabwe: a, full sample; b, omitting ages 14 and 15. Note: The fitted lines are local polynomials estimates.
where $H_i$ is a measure of HIV behavior (number of sexual partners or HIV status) or knowledge (having comprehensive knowledge of HIV, knowing whether HIV is transmitted via mosquitoes, etc.) for person $i$, and $S_i$ represents person $i$’s educational attainment (i.e., years of completed schooling or the probability that the person attended secondary school). Thus, $\beta$ is the parameter of interest. Equation (1) includes a vector of characteristics ($x_i'$), such as age in 1980, survey fixed effects, region fixed effects, and a binary variable equal to one if the person lives in an urban area and zero for rural.

The primary concern is that ordinary least squares (OLS) estimates of $\beta$ are biased. First, higher levels of ability lead to more education and better knowledge of health and other subjects. Thus, the observed correlation between education and knowledge could be an artifact of ability rather than education. This implies an upward bias in the estimates of $\beta$ using OLS. Second, if health improves education rather than the other way around, it will also bias our estimates of $\beta$. One solution to resolve these biases is to have a source of exogenous variation in schooling, which can be then used to obtain a causal impact.

As mentioned earlier, the 1980 reforms dramatically altered the schooling opportunities available to men and women young enough to take advantage of them. Figure 3 illustrates the sudden discontinuity in the transition rates in-
duced by the reform allowing automatic admission to Form I after grade 7. In itself, this implies a clearly different schooling experience for pupils age 15 or younger in 1980 compared to their seniors. However, the discontinuity in the probability of secondary school enrollment is more fuzzy than sharp (also seen in fig. 4a). Hence, in some specifications we drop the ages around the cutoff to show the robustness of our results. Our preferred specification drops ages 14 and 15 from the analysis (and hence is analogous to fig. 4b, where the discontinuity is clearer) and keeps a window of 9–20 years old in 1980. As we show in the next section, however, the results are largely robust to several variations of the regression discontinuity (RD) design, including the change in the window length.

The age-specific component of the reform provides the source of exogenous variation in schooling to test for the causal effect on knowledge about health. In essence, our estimation strategy compares the HIV-related behavior and knowledge of individuals age 15 in 1980 to the corresponding knowledge of those age 16 in 1980. As in van der Klaauw (2002), the indicator for the age cutoff serves as an excluded instrument in a two-stage least-squares (2SLS) regression. Formally, our first-stage equation estimates the following relationship:

$$S_i = \beta_1 \text{DumAge}_i + \beta_2 \text{DumAge}_i \times (\text{Age80}_i - 15) + \beta_3 (1 - \text{DumAge}_i) \times (\text{Age80}_i - 15) + x_i' \theta + \epsilon_i,$$

where DumAge is a dummy variable taking on the value of 1 if age in 1980 is less than or equal to 15, and 0 otherwise. We use linear approximations on either side of the cutoff, although we consider polynomials of the slopes in online-only appendix table 4. Visual representation of the RD (fig. 4) suggests that linear slopes might be a good fit. Standard errors are clustered at the region age level (around 80 clusters in our preferred specification). Predicted values of $S_i$ from this regression are then used to estimate our second stage:

12 Results are not sensitive to the use of higher-order polynomials for our preferred specification (dropping 14- and 15-year-olds). When those ages are included, the results do show sensitivity when using higher-order polynomials.
13 The age-region cluster is used to account for potentially different intensities of school construction across regions; hence, it is capturing the age-specific spatial correlations across regions. In app. table 5, we show that the main results of the article remain for alternative cluster constructions. Panel 1 shows the results for the main outcome variables without any clustering, and panel 2 replicates the results with our preferred age-region cluster. Panel 3 shows much smaller standard errors when clustering by age only. Finally in panel 4, we implement the wild-cluster bootstrap proposed by Cameron, Gelbach, and Miller (2008) but in reduced form.
\[
H_i = \beta S_i + \beta_2 \text{DumAge}_i \times (\text{Age80}_i - 15)
+ \beta_3 (1 - \text{DumAge}_i) \times (\text{Age80}_i - 15) + x'_i \theta + \epsilon_i.
\]  

We interpret ρ from the above equation as the causal effect of education on HIV-related behavior and knowledge.

Validity of the design rests on the assumption that children just below 15 years of age in 1980 are similar in unobservable ways to children just above age 15 in 1980. Our instrumental variables approach will be invalid if other variables (except for education) exhibit a discontinuity around the cutoff point. To some extent we can test the validity of our design. For example, since the education reform in Zimbabwe was targeted toward the black population who suffered an apartheid-style regime before 1980, we should expect a discontinuity at age 15 in 1980 for black but not for white Zimbabweans. Unfortunately, the DHS did not collect information about ethnicity. However, Agüero and Ramachandran (2014), using a 10% microsample of the 2002 population census in Zimbabwe, validate this point. They show that for whites there is no discontinuity at the cutoff point.

The increase in educational attainment due to the reform should affect only those in Zimbabwe and not those in other African countries. Agüero and Ramachandran (2014), again using the 2002 census, show that foreigners living in Zimbabwe in 2002 also failed to show a discontinuity at age 15 in 1980. We expand this idea and explore whether other neighboring countries exhibit a discontinuity for men and women at age 15 in 1980. In figure 5a we reproduce figure 4a for Zambia using the 1996 DHS. Observing a discontinuity at the cutoff for Zambia will reduce the validity of our identification and suggest that factors beyond the local education reform in Zimbabwe are driving the results in figure 4a. As observed in figure 5a, there is no evidence of such discontinuity in Zambia. Figure 5b shows a lack of discontinuity for the 1998 South Africa DHS. These cases allow us to reject the possibility that regional factors in southern Africa explain the gains in education for those age 15 or less in 1980 in Zimbabwe.¹⁴

Finally, it is clear that there were other reforms in the country taking place at the same time (Palmer 1990; Kumaranyayake et al. 2000). Thus, it is possible that our estimation strategy is capturing the confounding effect of all other changes in addition to the educational reform. For example, it has been noted that the provision of health services increased with Independence (Thomas 2000).

¹⁴ We also restricted the analysis to regions in Zambia and South Africa closer to Zimbabwe and found similar results (figures not shown but available on request). We thank a referee for this suggestion.
and Maluccio 1996). Thus, the creation of new hospitals or medical centers might seem to confound the effect of the reform. However, this is unlikely to be the case because a hospital would not be able to discriminate those age 15 and below in 1980 compared to those age 16 and over. Therefore, it is the fact that the reform affected those age 15 or less in 1980 in a disproportional way relative to those age 15 or more that allows us to identify the effects of education on knowledge. Nevertheless, we examine a proxy for health status around the age cutoff to ensure that our results are not driven by changes other than education reform.

The DHS have objective measures of height for women. If women age 15 or less were disproportionately benefiting from health- or nutrition-related services relative to those age 16 or more in 1980, we should expect to observe height as an adult to exhibit a discontinuity around the cutoff. Figure 5c plots mean height (implied 1 decimal point in centimeters) by age as of 1980 for women in Zimbabwe. While there appears to be a drop in height for people exactly age 15 years in 1980, the trends before and after age 15 do not show any sign of differential access to health services. Furthermore, given that women’s height is explained by the availability of nutrients during childhood and adolescence, we should expect that under “normal” conditions women have their adolescent growth spur before age 15. Thus, figure 5c also serves as a placebo test, as we should not observe a discontinuity at age 15 in 1980. Such a discontinuity would indicate that the reform might be confounding other possible changes in Zimbabwe around that time and harming our results. The fact that we do not observe such a discontinuity, we argue, reinforces the validity of our empirical strategy.

Finally, we also show the age distribution as of 1980 (app. fig. 1). The graph indicates that age in 1980 appears balanced across the cutoff of 15 years of age and suggests that there was no systematic misreporting around the age cutoff. Since this density is computed from people we observe in the surveys, we can think of this density as being composed of survivors. The density graph shows that there are just as many women to the left of the cutoff as there are to the right of the cutoff (15 years). So this should mitigate any concerns one might have about differential attrition due to death around the cutoff.

V. Results
A. First-Stage Results
Before turning to the results on behavior and knowledge, we first present evidence that the education reform affected the relevant age groups (15 and below in 1980) in terms of increasing years of schooling as well as the probability of having attended secondary school. We have already shown some
Figure 5. Robustness checks by age in 1980 years of education in Zambia using the 1996 DHS (a) and in South Africa using the 1998 DHS (b) and mean height for Zimbabwean women (c). Note: The fitted lines are local polynomials estimates.
of this evidence in the figures discussed in Sections II and IV; thus, we only discuss regression estimates in this section. As mentioned in Section IV, the RD approach dictates that we use the cutoff variable (1 if age 15 or less in 1980, 0 otherwise) as the instrument in our second-stage regressions (eq. [3]).

Table 2 shows that the first stage is rather strong. We explore three different windows—ages 8–21 in 1980 (cols. 1 and 4), ages 9–20 (cols. 2 and 5), and ages 10–19 (cols. 3 and 6)—as well as different versions of the RD design (panels A–C). In addition, columns 4–6 show that even changing the dependent variable to whether someone attends secondary school does not change the results. As is usually the case with RD designs, for larger windows, we get greater precision, but it also leads to the inclusion of points that are likely not comparable due to other factors. However, the point estimates are robust in all specifications and also across panels where we drop certain ages at the cutoff. Column 1 from panel C (our preferred specification drops ages 14 and 15 from the analysis) of table 2 suggests that people who were 15 and below as of 1980 gained approximately 1.8 years of additional schooling, and column 4 of the same panel suggests that they were 25 percentage points more likely to have attended secondary school. Within our preferred window of ages 9–20, the results are largely similar.
These effects are large. For example, Duflo (2001) finds that each new school constructed in Indonesia per 1,000 children was associated with an increase of 0.12 to 0.19 in years of education of males from a mean of 8 years of school. Our findings suggest that the reform in Zimbabwe increased by 25% the years of education of younger women relative to their slightly older counterparts. Younger women are also 50% more likely to have attended secondary school compared to 16-year-olds or older in 1980.

In appendix table 1 we show that it is only this relevant age cutoff (14 or 15 in 1980) that matters for the empirical design. As a check on the overall RD design, we show that at age cutoffs such as 17, 18, or 12 we find no discontinuities in educational attainment.

B. Education and HIV-Related Outcomes

Figures 6a and 6b present the visual, reduced-form descriptions of our results on the effect of education on number of sexual partners and HIV status, respectively. Women age 15 or less in 1980 have fewer sexual partners com-

---

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>Years of Schooling</th>
<th>Attended Secondary School</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8–21 (1)</td>
<td>9–20 (2)</td>
</tr>
<tr>
<td>A. All women:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age less than 15 in 1980</td>
<td>1.279***</td>
<td>1.170***</td>
</tr>
<tr>
<td>Observations</td>
<td>(213)</td>
<td>(233)</td>
</tr>
<tr>
<td>R²</td>
<td>.28</td>
<td>.27</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>7.08</td>
<td>7.01</td>
</tr>
<tr>
<td>B. Excluding women age 15 in 1980:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age less than 15 in 1980</td>
<td>1.580***</td>
<td>1.506***</td>
</tr>
<tr>
<td>Observations</td>
<td>(220)</td>
<td>(251)</td>
</tr>
<tr>
<td>R²</td>
<td>.29</td>
<td>.28</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>7.10</td>
<td>7.03</td>
</tr>
<tr>
<td>C. Excluding women age 14 and 15 in 1980:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age less than 15 in 1980</td>
<td>1.822***</td>
<td>1.837***</td>
</tr>
<tr>
<td>Observations</td>
<td>(256)</td>
<td>(312)</td>
</tr>
<tr>
<td>R²</td>
<td>.30</td>
<td>.29</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>7.07</td>
<td>6.99</td>
</tr>
</tbody>
</table>

**Note.** Column headings are dependent variables by age in 1980 bandwidth. All regressions include controls for urban/rural location, region fixed effects, survey fixed effects, and a linear spline in age in 1980. Robust standard errors clustered by age in 1980 and region in parentheses. *** Significant at 1%.
Figure 6. HIV-related behavior in Zimbabwe by age in 1980, number of sexual partners (a), and HIV status (b). Note: The fitted lines are local polynomials estimates.
pared to their counterparts age 16 and above. The results for HIV-positive status, however, do not show a similar discontinuity in the graphs. The regression analog of these graphs is shown in tables 3 and 4, where we present the OLS and 2SLS estimates for a variety of age windows following the same structure as table 2 in terms of accounting for the RD design. As discussed in the previous section, OLS estimates could be biased due to the presence of unobserved characteristics simultaneously affecting schooling and sexual behavior decisions. For example, promiscuity and schooling could be explained by income levels, preferences, and social values, making it difficult to sign the bias in OLS.

We trim the top 5% of data on number of sexual partners. There are outliers in these data, with some women reporting having 70–80 partners. The 95th percentile of number of partners is four.
Table 3 displays the effect on the number of sexual partners. The 2SLS estimate indicates (from col. 5, panel C) that each additional year of schooling is associated with a statistically significant decline in the number of sexual partners of 6.3% ($-0.0976/1.549$). This table also shows that the 2SLS results are quite robust in the RD design and with respect to the choice of window around the cutoff.

In Table 4, we repeat this exercise and find that the educated are less likely to have HIV, using an objective blood test included in the DHS, although these results are not statistically significant. Note that the sample is smaller, as the test was not included in all the surveys, and in each survey it is only administered to a random subsample of interviewed women.\textsuperscript{16} However, we only consider these results to be suggestive of lower HIV rates for the more

\textsuperscript{16} Furthermore, a power calculation shows that the sample size needed is around 7,000 with our effect size of 0.03 and the standard deviation of the dependent variable of 0.44, with standard levels of power of at least 80%.
C. Education and Knowledge about HIV
1. Do the Educated Know More about HIV?

We now explore whether the more educated have more knowledge about how HIV is transmitted. In figure 7 we show in reduced form a sampling of the results. For example, in figure 7a, women who were younger than 15 in 1980 are much less likely to have common misconceptions about HIV (knowledge that HIV cannot be spread via mosquitoes or food). Figure 7b shows a striking discontinuity in knowledge about whether condoms reduce the chances of HIV transmission.

Table 5 examines the impact of education on knowledge about HIV. For these results, we restrict the age window to 9–20 in 1980. As before, our results are robust to different window lengths. We present results for questions that were the same across surveys, to increase sample size (for smaller sample sizes, we obtain a rather small first-stage $F$-statistic). For each question about HIV knowledge, we present 2SLS results in the same format as before to show robustness to an RD design (OLS results are presented in app. table 2). Table 5 shows that within the 2SLS framework there appears to be a statistically significant relationship between education and HIV knowledge. While for some questions like, “Have you heard of HIV?” the size of the effect is predictably small (since most people have heard of HIV), for some questions like, “Have comprehensive knowledge of HIV?” every extra year of education raises the probability of being correct by almost 5 percentage points.17 This is a large effect since only 47% of women seem to have comprehensive knowledge of HIV. When it comes to comprehensive knowledge or having fewer misconceptions about HIV, education seems to play a very important role. Indeed, the 2SLS estimates are much larger than OLS estimates (at least for these two important questions), and even the OLS estimates are quite large compared to the mean. Overall, table 5 suggests that for a majority of knowledge measures, education plays an important role. The more educated, within the 2SLS framework, have more knowledge about how HIV is spread.18

---

17 Someone who correctly answers the following questions is considered to have "comprehensive" knowledge according to the DHS: (1) Can a healthy person have HIV? (2) Does a condom reduce the chance of getting HIV? (3) Can HIV be passed on via mosquitoes? (4) Is HIV transmission limited by having only 1 sexual partner?

18 In app. table 6 we consider a binary variable capturing whether participants attended at least some secondary school instead of years of schooling. As shown there, these findings remain unchanged with this alternative measure of schooling.
Figure 7. HIV knowledge in Zimbabwe by age in 1980: misconception that HIV can be spread via mosqui- toes (a) and knowledge that condoms reduce the chances of HIV transmission (b). Note: The fitted lines are local polynomials estimates.
## TABLE 5
EDUCATION AND KNOWLEDGE ABOUT HIV (TWO-STAGE LEAST SQUARES)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hear of HIV?</strong></td>
<td>0.027***</td>
<td>0.065***</td>
<td>0.008</td>
<td>0.056**</td>
<td>0.044**</td>
<td>-0.007</td>
<td>0.050</td>
<td>-0.089***</td>
</tr>
<tr>
<td><strong>Comprehensive Knowledge of</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV?</td>
<td>0.009</td>
<td>0.023</td>
<td>0.018</td>
<td>0.022</td>
<td>0.021</td>
<td>0.020</td>
<td>0.035</td>
<td>0.024</td>
</tr>
<tr>
<td><strong>Can a Healthy Person Have</strong></td>
<td>5,033</td>
<td>5,033</td>
<td>4,922</td>
<td>5,033</td>
<td>4,633</td>
<td>4,548</td>
<td>3,554</td>
<td>3,375</td>
</tr>
<tr>
<td>HIV?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condom Reduces Chances</strong></td>
<td>0.98</td>
<td>0.47</td>
<td>0.87</td>
<td>0.76</td>
<td>0.85</td>
<td>0.68</td>
<td>0.53</td>
<td>0.11</td>
</tr>
<tr>
<td>of HIV?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HIV Passed on by Mosquito</strong></td>
<td>29.41</td>
<td>29.41</td>
<td>25.31</td>
<td>29.41</td>
<td>28.71</td>
<td>19.74</td>
<td>21.68</td>
<td>21.17</td>
</tr>
<tr>
<td>Bites?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HIV Passed during Delivery?</strong></td>
<td>4,669</td>
<td>4,669</td>
<td>4,563</td>
<td>4,669</td>
<td>4,303</td>
<td>4,216</td>
<td>3,304</td>
<td>3,123</td>
</tr>
<tr>
<td><strong>HIV Is Limited by Having</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only 1 Partner?</td>
<td>40.99</td>
<td>40.99</td>
<td>36.56</td>
<td>40.99</td>
<td>38.99</td>
<td>29.59</td>
<td>28.5</td>
<td>27.94</td>
</tr>
<tr>
<td><strong>Has Common Misconceptions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>about HIV?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Notes
- Dependent variables are equal to 1 if the answer is correct, 0 if wrong. Years of schooling is instrumented using the discontinuity at age 15 in 1980. The sample includes women between age 9 and 20 in 1980. All regressions include controls for urban/rural location, region fixed effects, survey fixed effects, and a linear spline in age in 1980. Standard errors clustered by age in 1980 and region in parentheses.
- * Significant at 10%.
- ** Significant at 5%.
- *** Significant at 1%.
OLS estimates are likely to be upward biased due to unobserved ability as explained earlier. So finding that 2SLS estimates are larger than OLS estimates could be surprising. However, this is not the case in this literature, as these findings are consistent with previous estimates of the causal relationship between education and health (Cutler and Lleras-Muney 2010). Card (2001) finds similar results in his review of the causal impact of education on wages. To understand this pattern, note that the use of supply-side interventions, such as the education reform in Zimbabwe, allows us to estimate the effect of education on health by altering the schooling choices of individuals who, in the absence of the reform, would otherwise have very low levels of schooling. This change in the choice set is independent of the individuals’ ability and therefore creates an exogenous source of variation. If those affected by the reform faced higher-than-average costs of schooling, as it was clearly true in the case for black Zimbabweans in Rhodesia, the 2SLS estimates provide local average treatment effects that are expected to be larger than the average marginal effect of schooling on HIV-related behavior and knowledge. Nonetheless, as discussed by Card (2001), from a policy point of view, the average effect of education on health for the group affected by the reform is the main object of interest—as opposed to the average marginal effect of education for the entire population. This means that our estimates are relevant to similar populations facing larger-than-average barriers to education as observed in several sub-Saharan African countries.\(^1\)

2. How Do Women Acquire Health Knowledge?

It is not immediately clear how the more educated gather more information about HIV. It is unlikely that this information was learned in sex education classes in school in the 1980s. In fact the first reported cases of HIV in Zimbabwe date to the mid-1980s, a time when the group of people affected the most by the reform had already left school.\(^2\) Table 6 examines the relationship between educational status and the use of various media outlets where one could learn about HIV (the reduced-form graph is presented in fig. 8a). Appendix table 3 provides the relevant OLS estimates.

An extra year of schooling significantly raises the possibility of reading newspapers (2SLS magnitudes suggest a 15% increase in the likelihood of reading newspapers). In one of the few studies on the topic from the late

\(^1\) Measurement error could be another possible reason why 2SLS estimates are larger than OLS estimates (Hertz 2003). However, given the institutional settings of Zimbabwe in 1980, the local average treatment effect seems to play a major role.

1980s and early 1990s, Pitts and Jackson (1993) find that newspapers were the most important medium for conveying information about HIV/AIDS. Pertinent to our findings on knowledge and behavior, Pitts and Jackson (1993) note that the focus of these newspaper articles was more on education and awareness of HIV, rather than prevention. We reproduce a graph (fig. 8b) from their study suggesting that the more educated were able to assimilate more information about HIV via such newspaper articles.\(^{21}\)

A United Nations Population Fund country fact sheet on HIV from Zimbabwe in 2009 reports that “historical mapping also found consistent high media coverage since the early 1990s and high coverage of STI treatment since the late 1980s.”\(^{22}\) Thus, we believe the above results provide suggestive evidence that the more educated might know more about HIV via more frequent media use.

We also explore a few other ways by which the more educated might obtain more knowledge about health. As mentioned above, while examining

\(^{21}\) Perhaps surprisingly we do not see any impact of education on watching TV or listening to the radio, even though OLS estimates suggest a positive correlation.

Figure 8. Frequency of reading newspapers (a) and newspaper articles by HIV focus, 1987–91 (b; based on data from Pitts and Jackson [1993], table 1). Note: The fitted lines are local polynomials estimates.
antenatal visits is not possible, we explore other factors such as spousal education, insurance take up, and the quality and frequency of family planning services.

A more recent source of information on HIV are antenatal clinics. Antenatal clinics often provide pregnant mothers information about HIV transmission and its role during the birth and breast-feeding processes. In the DHS, women are asked whether antenatal visits provided them with information about the transmission probabilities of HIV from mother to children, general knowledge about HIV, and information about testing for HIV. We combine these measures into a single index, and the first panel in appendix figure 2 shows that it does appear that more educated women are more likely to receive information from antenatal clinics. However, the graph is only suggestive, as the line for older women changes slope substantially at age 18. Appendix table 7 shows that the point estimates are statistically significant; however, we interpret this with caution as the sample size is quite low. This is because the question on antenatal clinics is only asked of recent mothers, and our relevant sample, who are teenagers in 1980, are usually past childbearing age by 2005 and 2010 (the years when this question is asked in the DHS).

The remaining graphs of appendix figure 2 show that other ways of acquiring knowledge such as medical insurance or access to family planning workers (being informed about side effects of family planning is a proxy for the quality of family planning) do not seem to be different across the cutoff. Hence, while we are quite limited in being able to examine precisely how the reform enabled women to acquire more knowledge, we think access to media outlets might play the most important role. Appendix table 7 contains the regression analog of these graphs.

Finally, if the reform made women below the cutoff marry more educated husbands, then the education or socioeconomic status of the husband could be a channel via which women might acquire more knowledge. In appendix figure 3 we show that the reform does appear to have led to women marrying more educated husbands. However, the reform also directly affected the educational attainment of men. Hence, we think it is difficult to disentangle in this instance whether the effect of more education for women results from their directly acquiring more knowledge or from knowledge via more educated husbands. We are also mindful that with the exception of media access, the other mechanisms might directly affect health behavior, without altering the knowledge channel.

VI. Conclusion
This article examined the relationship between education and knowledge about health and health behaviors in the salient setting of HIV in Africa. By
exploiting an education reform that affected enrollment into secondary school in 1980 in Zimbabwe, we are able to provide some of the first causal estimates on whether education matters for knowledge about health and whether education matters for health behavior with regard to HIV. This evidence provides support to the hypothesis about the allocative efficiency of education (Rosenzweig and Schultz 1989; Rosenzweig 1995).

Our results reflect that the more educated are more likely to know complex and nuanced information about HIV. In particular, having had some secondary school has a large impact on having comprehensive knowledge about HIV. Moreover, the mechanism of greater knowledge appears to be via more frequent use of media outlets where information on HIV might be obtained. Finally, the more educated seem to undertake more preventive actions with regard to HIV by having fewer sexual partners. Perhaps surprisingly, this does not lead to any statistically significant effect of education on HIV status, although the direction of the relationship suggests that the more educated are less likely to be HIV positive.

Understanding whether education leads to lower HIV rates would underscore an important externality of education in developing countries. Hence, as we get better and more data on HIV status and HIV-related behaviors, such empirical exercises should reveal more insight on this important link.

References


