

Length of secondary schooling and risk of HIV infection in Botswana: evidence from a natural experiment



Jan-Walter De Neve, Günther Fink, S V Subramanian, Sikhulile Moyo, Jacob Bor

Summary

Background An estimated 2.1 million individuals are newly infected with HIV every year. Cross-sectional and longitudinal studies have reported conflicting evidence for the association between education and HIV risk, and no randomised trial has identified a causal effect for education on HIV incidence. We aimed to use a policy reform in secondary schooling in Botswana to identify the causal effect of length of schooling on new HIV infection.

Methods Data for HIV biomarkers and demographics were obtained from the nationally representative household 2004 and 2008 Botswana AIDS Impact Surveys (N=7018). In 1996, Botswana reformed the grade structure of secondary school, expanding access to grade ten and increasing educational attainment for affected cohorts. Using exposure to the policy reform as an instrumental variable, we used two-stage least squares to estimate the causal effect of years of schooling on the cumulative probability that an individual contracted HIV up to their age at the time of the survey. We also assessed the cost-effectiveness of secondary schooling as an HIV prevention intervention in comparison to other established interventions.

Findings Each additional year of secondary schooling caused by the policy change led to an absolute reduction in the cumulative risk of HIV infection of 8.1 percentage points ($p=0.008$), relative to a baseline prevalence of 25.5% in the pre-reform 1980 birth cohort. Effects were particularly large in women (11.6 percentage points, $p=0.046$). Results were robust to a wide array of sensitivity analyses. Secondary school was cost effective as an HIV prevention intervention by standard metrics (cost per HIV infection averted was US\$27 753).

Interpretation Additional years of secondary schooling had a large protective effect against HIV risk in Botswana, particularly for women. Increasing progression through secondary school could be a cost-effective HIV prevention measure in HIV-endemic settings, in addition to yielding other societal benefits.

Funding Takemi Program in International Health at the Harvard T.H. Chan School of Public Health, Belgian American Educational Foundation, Fernand Lazard Foundation, Boston University, National Institutes of Health.

Copyright © De Neve et al. Open Access article distributed under the terms of CC BY.

Introduction

HIV continues to be a major global health challenge with an estimated 2.1 million new infections each year.¹ Formal education, particularly of girls, has been hailed as a social vaccine to reduce the spread of HIV.² However, little causal evidence exists for this claim.³ Cross-sectional and longitudinal studies have reported conflicting evidence for the association between education and HIV risk. Early national surveillance surveys found higher rates of HIV among people with more education in several sub-Saharan African countries.⁴⁻⁶ However, other studies have shown a protective association between education and HIV infection, particularly as the epidemic has matured and information on prevention strategies has become more widely available.⁷⁻⁹ Education might reduce HIV risk through increased exposure to information about HIV and prevention methods;^{10,11} improved cognitive skills to make complex decisions;¹² better financial security,¹³⁻¹⁶ reducing participation in transactional sex for women;¹⁷ greater ability to match with lower-risk sex partners;¹⁷⁻²¹ and increased planning for the future and delayed gratification. However,

education might increase the size of an individual's sexual network, prolong the period of premarital sex,²² and increase transactional sex among men.²³ In addition to the implication of this study for HIV prevention, we aimed to contribute to the debate about whether the relationship between education and health more broadly is causal.²⁴⁻²⁷

The challenge in establishing the causal effect of schooling on HIV infection risk is that educational attainment is closely related to factors such as socioeconomic status, psychological traits, and preferences, which are difficult to control for fully in observational studies and which can also affect HIV risk. Thus, bivariate and covariate-adjusted associations between years of schooling and HIV status might not reflect a causal relation.^{7,9,11} Several randomised trials have aimed to identify the effect of schooling on HIV risk, but they have been underpowered to look at HIV incidence and have been paired with other interventions that make it difficult to attribute any effects to schooling.²⁸⁻³⁰

In this study, we exploit variation in educational attainment generated by a policy reform in Botswana in



Lancet Glob Health 2015; 3: e470-77

Published Online
June 29, 2015
[http://dx.doi.org/10.1016/S2214-109X\(15\)00087-X](http://dx.doi.org/10.1016/S2214-109X(15)00087-X)

See [Comment](#) page e428

Department of Global Health and Population (J-W De Neve MD, G Fink PhD), Department of Social and Behavioral Sciences (S V Subramanian PhD), and Center for Population and Development Studies (G Fink, S V Subramanian), Harvard T.H. Chan School of Public Health, Boston, MA, USA; Botswana-Harvard AIDS Institute Partnership, Gaborone, Botswana (S Moyo MSc); and Department of Global Health and Center for Global Health and Development, Boston University School of Public Health, Boston, MA, USA (Jacob Bor ScD)

Correspondence to: Jacob Bor, Department of Global Health, Boston University School of Public Health, Boston, MA 02118, USA
jb@bu.edu

January, 1996, in which the tenth year of education shifted from senior secondary to junior secondary school.³¹ Completion of junior secondary school is required for many vocational programmes and is a common exit point from formal schooling. The reform thus increased the benefits of completing grade ten. Additionally, because there are many more junior secondary schools than senior secondary schools, the reform substantially reduced transport costs and increased access to grade ten. As a result of the policy change, average years of schooling increased by almost 1 year (appendix p 14). The policy change affected specific birth cohorts—ie, those who would have entered secondary school in 1996 or later—and was unlikely to have affected HIV risk through mechanisms other than schooling itself. This reform constitutes a natural experiment to estimate the causal effect of schooling on risk of HIV infection through comparison of birth cohorts exposed to the reform versus those unexposed. Using multiple survey cohorts to disentangle age and cohort effects, we used the resulting variation in exposure to the reform to identify the causal effect of education on the cumulative risk of HIV infection. We also evaluated the cost-effectiveness of extended secondary school education as an HIV prevention measure.

See Online for appendix

Methods

Study population and data source

Botswana has one of the highest rates of HIV in the world, with 25·4% of adults aged 15–49 years infected in 2008.³² We obtained data from the Botswana AIDS Impact Surveys (BAIS) II (2004)³³ and III (2008),³² nationally-representative, cross-sectional, household surveys with HIV biomarker collection. For each survey, about 8300 households were selected; all members aged 10–64 years were eligible to be interviewed. Verbal informed consent was sought from each participant for both the survey interview and HIV biomarker collection.

Household and individual participation rates were, respectively, 92% and 93% for survey year 2004, and 87% and 82% for survey year 2008, yielding a total sample of 29 606 individuals. HIV test participation rates were 61% for survey year 2004, and 67% for survey year 2008.^{32,33} Data for age, sex, and years of schooling were available for 99·7% of respondents with a valid HIV test result.

We excluded respondents younger than 18 years because they would not have had the opportunity to complete junior secondary education, which at the time of the reform ran from age 15 years to 17 years in Botswana. We excluded respondents born before 1975 because previous school reforms led to rapid changes in education for these older cohorts.³¹ We excluded immigrants to Botswana because they would not have been exposed to the schooling intervention if they migrated in adulthood.

Procedures and statistical analysis

The key exposure in our analysis was the total years of schooling at the time of the survey. Our outcome of interest was HIV status at the time of the survey. HIV status reflects a binary stochastic realisation of an underlying probability: the cumulative probability of HIV infection up to a respondent's age at the time of the survey. We defined an indicator—taking the value one if the respondent was born in a cohort exposed to the 1996 education policy reform and zero otherwise. Given that children were historically expected to start primary school at age 7 years, children were expected to enter junior secondary school at age 15 years. Therefore, individuals born in 1981 or later would have entered junior secondary school in 1996 or later, and were classified as exposed. We included in our analyses controls for age in years, year of birth, sex, and district of birth.

As a benchmark for our analysis of policy reform, we assessed the naive association between years of schooling and HIV status in the BAIS II and BAIS III sample. We assessed the crude bivariate relation graphically and then adjusted for covariates in descriptive multivariate ordinary least squares (OLS; linear probability) regression models.³⁴ We estimated several specifications, modelling years of schooling as a continuous covariate; allowing for different slopes for 0–9 years and 10 years to 13 years and more of schooling; and with separate indicators for each additional year of schooling completed. We present linear probability models (ie, as opposed to logistic models) to allow comparison with the two-stage least squares (2SLS) instrumental variable models used to analyse the policy reform.³⁵

We analysed the policy reform in three steps. First, we assessed whether birth cohorts exposed to the reform (reform cohorts) had higher educational attainment than had those not exposed. We estimated the effect of exposure to the reform on total years of schooling completed in multivariate OLS regression models

	BAIS II (2004)		BAIS III (2008)	
	Women (n=1760)	Men (n=1354)	Women (n=2205)	Men (n=1699)
HIV positive	494 (28%)	145 (11%)	610 (27%)	216 (12%)
Age (years)	23 (20–25)	22 (20–25)	25 (21–28)	25 (21–28)
Years of schooling	10 (9–12)	10 (9–12)	10 (9–12)	10 (9–12)
Has at least 10 years of schooling	1108 (62%)	896 (65%)	1601 (73%)	1194 (73%)
Ever had sex	1556 (88%)	1058 (78%)	2048 (93%)	1407 (83%)
Age at first intercourse*(years)	18 (17–19)	18 (16–19)	18 (17–20)	18 (16–19)
Ever married	84 (5%)	13 (1%)	149 (7%)	40 (3%)
Literacy	1462 (83%)	1089 (80%)	2002 (91%)	1434 (86%)

Data are n (%) or median (IQR). Sample includes survey respondents who were citizens of Botswana, at least 18 years old at the time of the surveys, born in or after 1975, and with a valid HIV test result. *Number with age at first intercourse was 1520 for women and 1012 for men in BAIS II (2004), and 1987 for women and 1348 for men in BAIS III (2008). Sample weights used as provided. Data from Botswana AIDS Impact Surveys (BAIS) II (2004; N=15 479)³³ and III (2008; N=14 127).³²

Table 1: Participant characteristics

(first stage). We also assessed the effects of the reform on the probabilities of completing at least 7, 8, 9, 10, 11, 12, and 13 or more years of schooling and show graphically how this distribution changed across birth cohorts. Second, we assessed the intention-to-treat (ITT) effect of being in a reform cohort on HIV status in multivariate linear probability models. Natural experiments that change the probability of an exposure can be analysed like randomised controlled trials with non-compliance.³⁶ Third, we estimated 2SLS regression models using exposure to the reform as an instrumental variable (IV) for total years of schooling and adjusting for covariates. Under plausible assumptions, the treatment effect among so-called compliers is the ratio of the ITT and the difference in the probability of receiving treatment

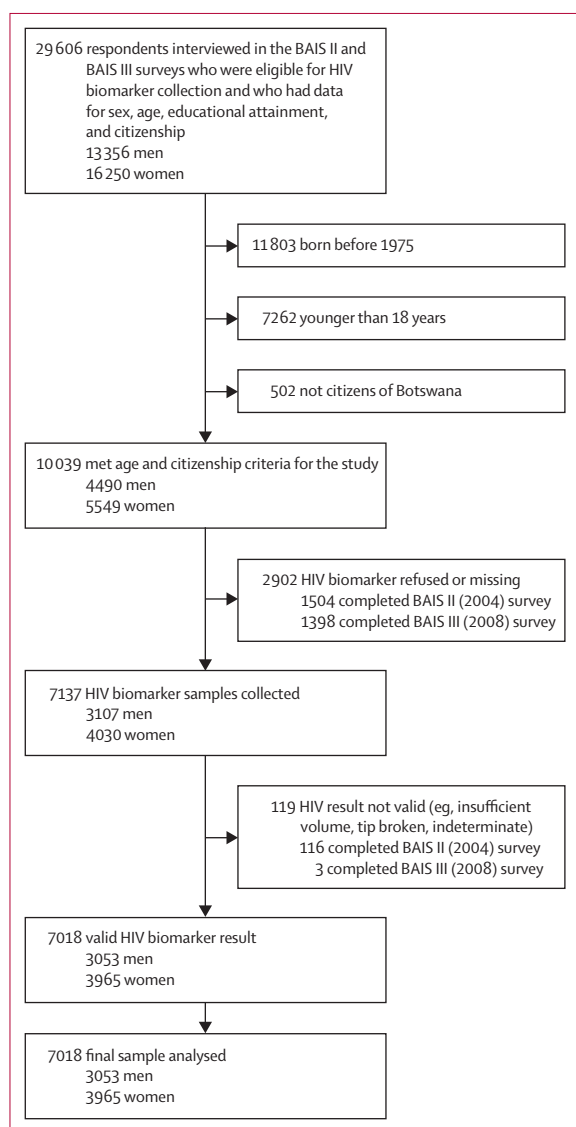


Figure 1: Study participants

Shows the flow of participants through the Botswana AIDS Impact Surveys (BAIS) II (2004) and III (2008).^{32,33}

(ITT/first stage=IV). We interpret our IV estimates as local to the subpopulation who complied with their treatment assignment—ie, individuals who increased their years of schooling because of the reform.³⁷

In all models, we controlled flexibly for age with a full set of single-year age indicators to account for the non-monotonic pattern of HIV infection across ages in Botswana and lower expected years of schooling for individuals at younger ages.³⁸ We also included indicators for district of birth. Finally, we adjusted for a continuous linear term in year of birth to account for continuous trends in HIV infection risk across birth cohorts. Exposure to the reform was modelled as an intercept shift for participants born in or after 1981. We estimated all models first for women and men separately, and then on the pooled sample. When pooling sexes, we included indicators for sex and the interactions of sex with all other covariates; however, we did not interact sex with the main exposure, so that the coefficient of interest reflects a weighted average of effects for men and women.

For our effect estimates to have a causal interpretation, four assumptions must be satisfied (appendix p 9). First, the instrument (reform cohort [Z]) must have had an effect on schooling (E); this is testable and we found large effects. Second, the instrument (Z) must be independent of unobserved confounders such as socioeconomic status (U), conditional on observed covariates (age, birth cohort trend, sex, and district of birth [X]); in our application this implies that people born before and after 1981 were similar, after controlling flexibly for age, district of birth, and a linear trend in HIV risk across birth cohorts. The availability of two survey years enables us to identify these cohort effects, while controlling flexibly for age and period effects. Our models control for period effects implicitly by simultaneously adjusting for age and a continuous term in year of birth. For a diagram encoding the assumptions behind the study design, see appendix, p 9.

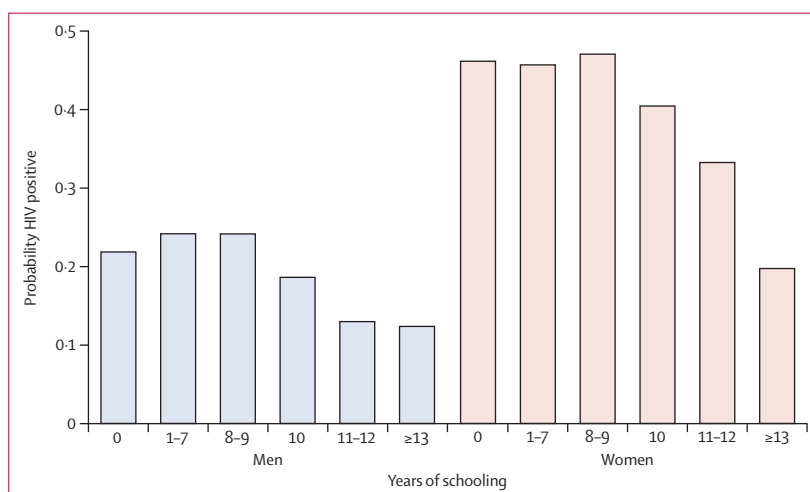


Figure 2: HIV prevalence by years of schooling in Botswana

Sample includes survey respondents who were citizens of Botswana, at least 24 years old at the time of the surveys, born in or after 1975, and had a valid HIV test result.

	Exposure	Outcome	Effect estimate (SE)
Women (n=3965; 32.3% HIV positive pre-reform)			
Model 1: first stage	Reform indicator	Years of schooling	0.64‡ (0.22)
Model 2: intention-to-treat	Reform indicator	HIV positive	-7.4† (3.1)
Model 3: 2SLS (IV)	Years of schooling	HIV positive	-11.6† (5.8)
Men (n=3053; 16.8% HIV positive pre-reform)			
Model 1: first stage	Reform indicator	Years of schooling	1.01‡ (0.32)
Model 2: intention-to-treat	Reform indicator	HIV positive	-5.0* (2.6)
Model 3: 2SLS (IV)	Years of schooling	HIV positive	-5.0* (2.9)
All participants (n=7018; 25.5% HIV positive pre-reform)			
Model 1: first stage	Reform indicator	Years of schooling	0.79‡ (0.19)
Model 2: intention-to-treat	Reform indicator	HIV positive	-6.4‡ (2.1)
Model 3: 2SLS (IV)	Years of schooling	HIV positive	-8.1‡ (3.1)

Coefficient and standard errors for HIV positive outcome are multiplied by 100 and reported on a percentage point scale. Percent HIV-positive pre-reform refers to HIV prevalence in the 1980 cohort. Regression 1 is an ordinary least squares (OLS) model. Regression 2 is an OLS linear probability model. Regression 3 is a two-stage least squares (2SLS) linear probability model, in which exposure to the reform was used as an instrument for years of schooling. All models included the following controls: single-year age indicators, a linear term for year of birth, an indicator for survey wave and indicators for district of birth. Regressions for the subsample with both sexes additionally control for age*sex, districtofbirth*sex, yearofbirth*sex and surveywave*sex interactions. The instrument was defined as=1 if year of birth >1980. *p<0.1. †p<0.05. ‡p<0.01. No weights were used. F statistics in the 2SLS models were 18.0 in the pooled sample, 8.6 in women, 9.5 in men. See appendix pp 17–22 for sensitivity analysis. IV=instrumental variable.

Table 2: Regression results

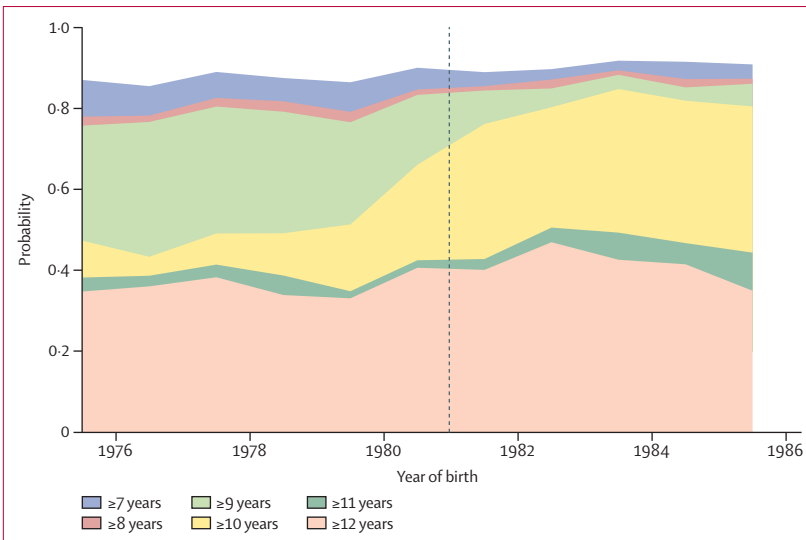


Figure 3: Educational attainment by birth cohort in Botswana
Shows probability that the respondent has attained at least X years of schooling. Sample includes survey respondents who were citizens of Botswana, at least 18 years old at the time of the surveys, born between 1975 and 1985, and had a valid HIV test result. Survey weights used as provided. Individuals born in 1981 or later (dotted line) would have entered junior secondary school in 1996 or later, and were classified as exposed.

To allow for potential non-linearities in underlying cohort trends, we did robustness checks including quadratic terms for year of birth, reducing the window of observation to a narrower set of birth cohorts, and allowing the slope of the trend across birth cohorts to differ before and after 1981. Identification comes from the fact that the policy reform led to a discontinuous change in schooling across cohorts. Our identifying assumption is that no other unobserved factors led to a

discontinuous change in HIV risk for precisely the same cohorts. To generate added confidence in this assumption, we did a placebo test, assessing the effect of the reform on individuals with fewer than 9 years of schooling—a population that was not affected by the reform. We also estimated difference-in-differences models, exploiting the fact that the policy reform was expected to have a larger effect in some districts than others, based on the share of the district populations with exactly 9 years of schooling pre-reform. Third, for our 2SLS estimates to have a causal interpretation, we assume that exposure to the policy reform (Z) affected HIV risk (Y) only through changes in schooling (E; exclusion restriction); this assumption is highly plausible given that the reform was a supply-side intervention that would not have specifically affected the reform cohorts except through their increased access to grade ten. Fourth, to interpret our results as complier causal effects (also known as local average treatment effects), we assume monotonicity—ie, that exposure to the reform (Z) only caused individuals to obtain more schooling or to have no change in schooling and did not lead some individuals to obtain less schooling. Violations of this assumption are possible but unlikely (eg, a person with a very strong preference for small class size might have continued to grade ten pre-reform but dropped out after grade nine after the reform).^{37,39}

In addition to the robustness checks described above, we conducted a range of sensitivity analyses including sampling weights, using alternate functional forms for age, modelling the outcome using a Probit link function, and imputing HIV status for respondents who did not consent to biomarker collection using two different methods. We also assessed whether the policy reform was associated with changes in consent rates and sample size that could bias results.

To compare the cost-effectiveness of secondary schooling versus other proven HIV prevention interventions, we calculated the costs per HIV infection averted and per disability-adjusted life-year (DALY) averted using estimates of the annual per-pupil costs of secondary education published by the UN Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics (appendix pp 2–3).

This study was reviewed by the Harvard T.H. Chan School of Public Health institutional review board and considered exempt from full ethics review because it was based on an anonymised dataset.

Role of the funding source

The funders of the study had no role in study design, data collection, analysis, interpretation of data, writing of the report, or in the decision to submit for publication. J-WDN and JB had full access to all of the data in the study and took responsibility for the decision to submit for publication.

Results

The 2004 and 2008 BAIS surveys included valid HIV biomarker data for 3965 women and 3053 men, for a total of 7018 respondents (table 1, figure 1). The crude association between HIV infection risk and amount of schooling was non-monotonic, peaking for people completing 8–9 years of education and declining sharply thereafter (figure 2). The strong association between schooling and HIV risk at higher school grades persisted in multivariate regression (appendix p 13). Each additional year of schooling after 9 years was associated with a 3.6 percentage point (SE 0.4) lower risk of HIV infection (appendix p 13). By contrast, there was no association between schooling and HIV risk in lower grades (coefficient 0.3 percentage points; SE 0.2). Although suggestive, these associations (like those previously reported in the literature) might be confounded by unobserved characteristics. Because of this, we use the natural experiment to estimate causal effects.

The reform increased average years of schooling completed by 0.792 years (SE 0.188; $p < 0.0001$; table 2; appendix p 14). The fraction of students completing at least 7, 8, or 9 years of schooling rose gradually and continuously across birth cohorts (figure 3; appendix p 14). However, the proportion of students with at least 10 years of schooling was much higher for cohorts born in 1981 or later. Due to grade repetition or late entry into school or both, some respondents born in 1979 and 1980 were also likely affected by the reform. Slight increases in completion of years 11 and 12 were also observed for reform cohorts (figure 3; appendix p 14).

Table 2 presents intention-to-treat results, in which HIV status was regressed directly on the instrument and covariates. Women who were exposed to the reform were 7 percentage points less likely to be HIV positive than those not exposed to the reform ($p = 0.017$); men were 5 percentage points less likely to be HIV positive ($p = 0.052$). The pooled coefficient was 6 percentage points ($p = 0.002$). Observed HIV prevalence closely matched the model predictions (appendix p 10). In the 2SLS IV models, each additional year of schooling induced by the reform reduced infection risk by 8 percentage points ($p = 0.008$) overall: 12 percentage points for women ($p = 0.046$) and 5 percentage points for men ($p = 0.085$; table 2). We were not able to reject the hypothesis that schooling had the same effect for men as for women. To put our estimates into context, table 2 displays the percent of respondents who were HIV positive in the 1980 birth cohort, the last cohort prior to the reform. Assuming that the respondents affected by the reform would have experienced HIV infection rates similar to those experienced by the pre-reform cohort, our full sample estimates imply an 8.1 percentage point reduction in HIV risk from 25.5% to 17.4% from an additional year of secondary schooling.

We did many robustness checks for the 2SLS results (table 3; appendix pp 17–22). In general, our results were

	Specification	N	Exposure	Effect estimate on HIV status (SE)
Model 1	2SLS, sampling weights	7018	Schooling	-7.8* (3.5)
Model 2	2SLS, HIV status imputed	8281	Schooling	-9.1* (4.4)
Model 3	IVProbit, marginal effects	7018	Schooling	-5.2† (1.4)
Model 4	2SLS, quadratic in YOB	7018	Schooling	-7.3* (3.0)
Model 5	2SLS, slope change in YOB	7018	Schooling	-7.9† (3.0)
Model 6	2SLS, diff-in-diff	7018	Schooling	-14.7 (23.1)
Model 7	ITT, education <9 years (placebo)	1175	Reform	-1.9 (5.9)
Model 8	ITT, education ≥9 years (antiplacebo)	5843	Reform	-7.2† (2.3)

Presents robustness checks for the 8.1 percentage point reduction in HIV infection risk per year of schooling induced by the reform in the pooled (both sexes) sample (table 2). For additional robustness checks and sex-specific sensitivity analyses see appendix pp 17–22. Coefficients and SEs are multiplied by 100 and presented on a percentage point scale. Models 1, 2, 4, 5, and 6 are two-stage least squares (2SLS) models. Model 3 is a Probit model using Stata's IVProbit command. Models 7 and 8 are ordinary least squares (OLS) models (intention to treat [ITT]). In all 2SLS models, exposure to the reform was used as an instrument for years of schooling. Model 2 uses two additional covariates, age at first intercourse and ever married, to impute HIV status. The instrument was defined as 1 if year of birth (YOB) >1980. F statistics for 2SLS models were 13.4 (model 1), 17.9 (model 4), 18.5 (model 5), 0.6 (model 6). Sample includes survey respondents who were citizens of Botswana, at least 18 years old at the time of the surveys, born in or after 1975, and had a valid HIV test result. * $p < 0.05$. † $p < 0.01$. Source: Botswana AIDS Impact Surveys II (2004)³³ and III (2008).³²

Table 3: Robustness checks

	Cost-effectiveness ratio (US\$ per infection averted)	Study (year)
Medical male circumcision	\$551; \$1096	Kahn et al (2006); ⁴³ Bärnighausen et al (2012) ⁴¹
Treatment as prevention (CD4 count ≥350 cells per µL)	\$8375	Bärnighausen et al (2012) ⁴¹
Pre-exposure prophylaxis	\$12 500–20 000; \$6000–66 000	Pretorius et al (2010); ⁴⁴ Hallett et al (2011) ⁴²
Secondary school	\$27 753	De Neve et al (2015)*

*This study; other benefits of schooling are not captured in the cost-effectiveness ratio.

Table 4: Cost-effectiveness ratio of secondary school and known HIV prevention interventions

not sensitive to sampling weights, imputation for HIV sampling non-consent, different specifications of the outcome, alternate specifications of the continuous trend across birth cohorts, nor to different modelling strategies for age (table 3; appendix pp 17–22). Our difference-in-differences analysis returned similar effect estimates as our main results, although SEs were larger (table 3). In our placebo test, the effect of the reform on HIV risk was driven entirely by respondents with at least 9 years of schooling, with no effect among respondents with fewer than 9 years (table 3).

UNESCO estimates from 2005 and 2007 place the annual per-pupil cost of secondary education in Botswana at US\$2248.⁴⁰ Since individuals who stayed in school for an additional year had an 8.1 percentage point lower risk of HIV infection, the cost per HIV infection averted

For more on WHO's cost-effectiveness benchmarks see http://www.who.int/choice/costs/CER_thresholds/en/

For the World Bank's 2009 GDP estimates see <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

was \$27753. By WHO's standard cost-effectiveness benchmarks an intervention is very cost-effective if it costs less than the gross domestic product (GDP) per person for each DALY averted. Based on calculations presented in the appendix (appendix pp 2–3), we estimate that an HIV infection at age 20 years would lead to 16·3 lifetime DALYs lost for someone who did not start antiretroviral therapy (ART) and 3·5 lifetime DALYs for someone who started ART, with a lifetime cost of \$12400. These calculations imply cost-effectiveness ratios of \$4387 per DALY with ART and \$1703 per DALY without ART. Each of these ratios is lower than World Bank estimates of Botswana's \$5178 GDP per person in 2009, implying that secondary school is very cost effective as an HIV prevention intervention. Secondary schooling is more expensive than circumcision and treatment as prevention, but of similar cost-effectiveness to the upper range of estimates for pre-exposure prophylaxis (table 4).^{41–44} Importantly, unlike these other interventions, secondary schooling has large benefits beyond the reduction of HIV transmission, which are excluded from the above calculations.

Discussion

We show, using an education policy reform as a natural experiment, that secondary schooling has a large protective effect against risk of HIV infection in Botswana. Effects were particularly large among women and were consistent across a wide array of robustness checks. Our estimates are somewhat larger, but generally consistent with, the strong negative associations we found between secondary schooling and HIV risk in multivariate OLS regression. We interpret our estimates as causal because they are not vulnerable to the types of unobserved factors (eg, psychological traits, unmeasured socioeconomic status) that can confound studies of the association between education and HIV.

The effects of schooling on HIV risk are likely to be heterogeneous, and our effect estimates are local in several important ways. First, our estimates are local to the specific grades affected by the policy change (grades 10–12); these grades might be a crucial exposure period⁴⁵ to determine lifetime HIV risk because this is a period when sexual behaviour patterns and labour market opportunities are formed;⁴⁶ effects of schooling might be qualitatively different in primary school, and indeed in our descriptive analysis, we found no association between primary schooling and HIV risk. Second, the causal effects that we estimate are local to the subpopulation of compliers—ie, those induced to increase schooling because of the reform. This subpopulation consists of people who, in the absence of the reform, would have dropped out after ninth grade—a group likely to be at particularly high risk of HIV. Third, the results are local to an epidemiological context in which HIV is hyperendemic with very high incidence for people aged 20–29 years; effects of this magnitude might not be

observed in lower-prevalence settings. Fourth, the effects are local to the years of risk exposure under study (the 1990s through early 2000s). Previous studies have reported changing associations between education and HIV risk over time, and we caution against generalising to cohorts born earlier who formed sexual behaviour patterns before HIV emerged as an epidemic in Botswana;^{22,47} however, we do suspect that our effect estimates are likely to be informative of current and future benefits of education in a society in which HIV is endemic. In addition to heterogeneity of treatment effect, 2SLS results might be larger than the OLS result because unobserved factors, such as personal charisma, might be positively associated with both educational attainment and HIV risk, thereby leading to downward bias in the OLS coefficient.

Our study has some limitations. First, some survey respondents chose not to provide HIV biomarkers, and migration or mortality could have affected the composition of the study sample. However, neither consent rates nor birth cohort sizes varied systematically with exposure to the reform (appendix p 11) and our results were robust to imputation. Second, some respondents might have acquired HIV before the age when they would have entered grade ten. Infection rates are very low before grade ten. More importantly, since our analysis was done on a risk-difference scale, our approach is robust to the existence of prevalent HIV by grade ten, so long as prevalence was smooth across birth cohorts. Third, we observed people only up to age 32 years. We cannot know whether we are measuring HIV infections truly averted or whether they have been delayed. However, this is a common limitation of prevention studies, and our analysis of cumulative incidence captures much longer follow-up than most randomised controlled trials, which observe incidence over a shorter (eg, 3 year⁴⁸) horizon.

Fourth, as discussed above, our analysis relies on the assumption that, conditional on age, period, district of birth, and a smooth trend in birth cohort, no other cohort-specific effects existed that would have led to a discontinuous change in HIV risk coinciding with the reform. It is worth interrogating this assumption. There are many reasons why HIV risk might change across birth cohorts but the likely candidates (infection rates among sexual partners, access to HIV treatment, changes in prevention programming, etc) are events that are either gradual over time (changes in the epidemic context) or that affect people of many different ages (eg, a national prevention campaign or the introduction of ART): in both cases, these events would result in gradual changes in HIV infection across birth cohorts, which we control for. For example, the scale-up of ART might have reduced infectiousness among respondents' sexual partners; however, we expect both selection of sexual partners and take-up of ART (among those sexual partners) to be smooth across birth cohorts. One example of a potential

Panel: Research in context**Systematic review**

Two systematic reviews synthesised evidence from more than 36 cross-sectional, cohort, and case-control studies.^{3,9} The authors of these reviews concluded that, over time, HIV risk seems to have increased in people with less schooling in sub-Saharan Africa. Studies from before 1996 found either no association between schooling and HIV risk or a higher risk among more educated people. All of the studies reviewed were observational and none used exogenous variation from a natural experiment to identify the causal effect of schooling. We searched PubMed for reports published in English between Jan 1, 2006, and July 6, 2014, with the search terms “education” and “HIV”. We identified three studies assessing the causal effect of education on HIV infection: one randomised trial, which was underpowered,¹⁷ and two natural experiments, of which the first one was underpowered and the second one did not control for age and could be biased.^{10,50}

Interpretation

Our study is among the first to estimate the causal effect of education on HIV infection by using an education policy reform as a natural experiment. Secondary schooling had a large protective effect against risk of HIV infection in Botswana, particularly among women. Increasing progression through secondary school could be a cost-effective HIV prevention measure in HIV-endemic settings, in addition to yielding other societal benefits.

confounder (and violation of our assumption) would be an HIV prevention programme implemented in a specific year, targeted to a specific school grade, and thus only affecting specific birth cohorts. However, Botswana's school-based HIV curriculum was not in place in 1996.⁴⁹ To generate confidence that our results are not confounded by other policy changes, we estimated difference-in-differences models and a placebo check, exploiting the fact that the reform was expected to affect mainly people with at least 9 years of schooling. Indeed, the reform had no effect on HIV risk for people with fewer than 9 years of schooling. And the reform had a larger effect for people born in districts where the reform was expected to have the greatest impact on educational attainment.

Finally, as with all infectious disease interventions, we expect spillover effects on incidence beyond the individuals directly affected by the reform. Given that people have sexual relationships across birth cohorts, these spillovers would be expected to be smooth across birth cohorts and would not bias our estimates. However, by excluding these spillover benefits, we might be underestimating the cost-effectiveness of secondary schooling.

Expanding access to secondary school had a large protective effect against HIV infection in Botswana. Our findings confirm what has been long suspected

(panel)—that secondary schooling is an important structural determinant of HIV infection and that this relation is causal. Further, our estimates indicate that secondary schooling is very cost effective as an HIV prevention intervention, in addition to its other societal benefits. Investment in expanded access to secondary schooling would be an effective HIV preventive measure and should be considered as part of combination HIV prevention strategies in countries with large, generalised HIV epidemics.

Contributors

J-WDN and JB conceived and designed the study. J-WDN did the statistical analysis under the guidance of JB. J-WDN and JB wrote the report. GF and SVS suggested improvements to the statistical analysis. GF, SM, and SVS contributed important revisions to the report. All authors approved the final submitted version of the report.

Declaration of interests

We declare no competing interests.

Acknowledgments

J-WDN was funded by the Takemi Program in International Health at the Harvard T.H. Chan School of Public Health, the Belgian American Educational Foundation, and the Fernand Lazard Foundation. JB received financial support from National Institutes of Health grant (K01MH105320-01A1) and the Peter Paul Career Development Professorship of Boston University. We thank study participants in the BAIS II and III surveys and staff at Statistics Botswana. In particular, we thank Virginia Sebekedi for assistance in understanding how the survey data were collected. The contents are the responsibility of the authors and do not necessarily reflect the views of any of the funders or the US government.

References

- 1 Joint United Nations Programme on HIV/AIDS. Global report: UNAIDS report on the global AIDS epidemic 2013. Geneva, Switzerland, 2013.
- 2 Vandemoortele J, Delamonica E. The ‘education vaccine’ against HIV. *Curr Issues Comp Educ* 2000; 3: 6–13.
- 3 Jukes M, Simmons S, Bundy D. Education and vulnerability: the role of schools in protecting young women and girls from HIV in southern Africa. *AIDS* 2008; 22 (suppl 4): S41–56.
- 4 Dallabetta GA, Miotti PG, Chipangwi JD, et al. High socioeconomic-status is a risk factor for human-immunodeficiency-virus type-1 (HIV-1) infection but not for sexually-transmitted diseases in women in Malawi - implications for HIV-1 Control. *J Infect Dis* 1993; 167: 36–42.
- 5 Kirunga CT, Ntozi JP. Socio-economic determinants of HIV serostatus: a study of Rakai District, Uganda. *Health Transit Rev* 1997; 7 (suppl): 175–88.
- 6 Fortson JG. The gradient in sub-Saharan Africa: socioeconomic status and HIV/AIDS. *Demography* 2008; 45: 303–22.
- 7 Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS* 2007; 21 (suppl): S29–38.
- 8 Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F, Monze M. Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. *AIDS* 2001; 15: 907–16.
- 9 Hargreaves JR, Bonell CP, Boler T, et al. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. *AIDS* 2008; 22: 403–14.
- 10 Agüero J, Bharadwaj P. Do the more educated know more about health? Evidence from schooling and HIV knowledge in Zimbabwe. *Econ Dev Cult Change* 2014; 62: 489–517.
- 11 de Walque D. Does education affect HIV status? Evidence from five African countries. *World Bank Econ Rev* 2009; 23: 209–33.
- 12 Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. *J Health Econ* 2010; 29: 1–28.

- 13 Psacharopoulos G. Returns to investment in education—a global update. *World Dev* 1994; **22**: 1325–43.
- 14 Schultz TW. Capital formation by education. *J Polit Econ* 1960; **68**: 571–83.
- 15 Siphambe HK. Rates of return to education in Botswana. *Econ Educ Rev* 2000; **19**: 291–300.
- 16 Willis RJ. A new approach to the economic theory of fertility behavior. *J Polit Econ* 1973; **81**: S14–64.
- 17 Baird SJ, Garfein RS, McIntosh CT, Ozler B. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *Lancet* 2012; **379**: 1320–29.
- 18 Behrman JR, Rosenzweig MR. Does increasing women's schooling raise the schooling of the next generation? *Am Econ Rev* 2002; **92**: 323–34.
- 19 Bosma H, Appels A, Sturmans F, Grabauskas V, Gostautas A. Educational level of spouses and risk of mortality: the WHO Kaunas-Rotterdam Intervention Study (KRIS). *Int J Epidemiol* 1995; **24**: 119–26.
- 20 Pettifor AE, Rees HV, Kleinschmidt I, et al. Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. *AIDS* 2005; **19**: 1525–34.
- 21 Rosenberg M, Pettifor A, Miller WC, et al. Relationship between school dropout and teen pregnancy among rural South African young women. *Int J Epidemiol* 2015; published online Feb 24. DOI: 10.1093/ije/dyv007
- 22 Case A, Paxson C. HIV risk and adolescent behaviors in Africa. *Am Econ Rev* 2013; **103**: 433–38.
- 23 Kohler HP, Thornton RL. Conditional cash transfers and HIV/AIDS prevention: unconditionally promising? *World Bank Econ Rev* 2011; **26**: 165–90.
- 24 Clark D, Royer H. The effect of education on adult mortality and health: evidence from Britain. *Am Econ Rev* 2013; **103**: 2087–120.
- 25 de Walque D. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *J Health Econ* 2007; **26**: 877–95.
- 26 Lleras-Muney A. The relationship between education and adult mortality in the United States. *Rev Econ Stud* 2005; **72**: 189–221.
- 27 McCrary J, Royer H. The effect of female education on fertility and infant health: evidence from school entry policies using exact date of birth. *Am Econ Rev* 2011; **101**: 158–95.
- 28 Cho H, Hallfors DD, Mbai II, et al. Keeping adolescent orphans in school to prevent human immunodeficiency virus infection: evidence from a randomized controlled trial in Kenya. *J Adol Health* 2011; **48**: 523–26.
- 29 Duflo E, World Bank. Office of the Senior Vice President and Chief Economist Development Economics. Education and HIV/AIDS prevention: evidence from a randomized evaluation in Western Kenya. Washington, DC: World Bank, Development Economics, Office of the Senior Vice President and Chief Economist, 2006.
- 30 Hallfors D, Cho H, Rusakaniko S, Iritani B, Mapfumo J, Halpern C. Supporting adolescent orphan girls to stay in school as HIV risk prevention: evidence from a randomized controlled trial in Zimbabwe. *Am J Public Health* 2011; **101**: 1082–88.
- 31 Botswana. The revised national policy on education. Gaborone: Botswana Government Printer, 1994.
- 32 Central Statistics Office Government of Botswana. 2008 Botswana AIDS Impact Survey III. Gaborone: Government of Botswana, 2008.
- 33 Central Statistics Office Government of Botswana. 2004 Botswana AIDS Impact Survey II. Gaborone: Government of Botswana, 2004.
- 34 Wooldridge JM. Introductory econometrics: a modern approach. East Lansing: Michigan State University, 2013.
- 35 Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 2001; **15**: 69–85.
- 36 Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health* 2000; **21**: 121–45.
- 37 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–55.
- 38 Central Statistics Office, National AIDS Co-ordinating Agency (Botswana). Botswana AIDS impact survey III: statistical report. Gaborone: Central Statistics Office, 2009.
- 39 Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–72.
- 40 UNESCO Institute for Statistics. Education indicators, 2014. <http://data.uis.unesco.org/Index.aspx?queryid=192> (accessed Oct 5, 2014).
- 41 Bärnighausen T, Bloom DE, Humair S. Economics of antiretroviral treatment vs. circumcision for HIV prevention. *Proc Natl Acad Sci USA* 2012; **109**: 21271–76.
- 42 Hallett TB, Baeten JM, Heffron R, et al. Optimal uses of antiretrovirals for prevention in HIV-1 serodiscordant heterosexual couples in South Africa: a modelling study. *PLoS Med* 2011; **8**: e1001123.
- 43 Kahn JG, Marseille E, Auvert B. Cost-effectiveness of male circumcision for HIV prevention in a South African setting. *PLoS Med* 2006; **3**: e517.
- 44 Pretorius C, Stover J, Bollinger L, Bacaer N, Williams B. Evaluating the cost-effectiveness of pre-exposure prophylaxis (PrEP) and its impact on HIV-1 transmission in South Africa. *PLoS One* 2010; **5**: e13646.
- 45 Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health* 2003; **57**: 778–83.
- 46 Alsan MM, Cutler DM. Girls' education and HIV risk: evidence from Uganda. *J Health Econ* 2013; **32**: 863–72.
- 47 Gregson S, Waddell H, Chandiwana S. School education and HIV control in sub-Saharan Africa: from discord to harmony? *J Int Dev* 2001; **13**: 467–85.
- 48 068 HPTN. Effects of cash transfer for the prevention of HIV in young South African women. http://www.hptn.org/research_studies/hptn068.asp (accessed March 27, 2014).
- 49 Tsheko G. HIV/AIDS and Education in Botswana. In: Alexander W, Wiseman RNG, eds. The impact of HIV/AIDS on education worldwide. Bingley: Emerald Group Publications; 2012: 141–59.
- 50 Behrman JA. The effect of increased primary schooling on adult women's HIV status in Malawi and Uganda: universal primary education as a natural experiment. *Soc Sci Med* 2015; **127**: 108–15.