From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 | Public and International Affairs

2011

In Sickness and Wealth:Three Essays on Health Human Capital and HIV in sub-Sahran Africa

Gummerson, Elizabeth Anne

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp01xp68kg22n

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

IN SICKNESS AND WEALTH: THREE ESSAYS ON HEALTH, HUMAN CAPITAL AND HIV IN SUB-SAHARAN AFRICA

Elizabeth Anne Gummerson

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE WOODROW WILSON SCHOOL OF PUBLIC AND INTERNATIONAL AFFAIRS

Adviser: Anne Case

November 2011

© Copyright by Elizabeth Anne Gummerson, 2011.All Rights Reserved

For my Gram, Phyllis Gummerson, who never got to go to university, and my Grandpa, Professor Arleigh Hess, who never left.

Acknowledgements:

Properly acknowledging all the people who helped this piece of work along the way would probably take more paper and ink than the dissertation itself, but I would like to express my particular gratitude to a few people. First and foremost, I owe an incredible debt of gratitude to my advisor, Anne Case, who has been a mentor, friend and guide through this process and who has stuck with me through quite a few rounds of triumph and disaster. Long before I came to graduate school, her work demonstrated to me that scholarship can work towards the lofty aim of serving social justice without conceding rigor or academic honesty. I continue to aspire to her example.

I am also extremely grateful to the members of my committee; to Taryn Dinkelman, who read more terrible drafts than one person ever should with patience and good humor and returned excellent suggestions every time; to Chris Paxson for encouraging me to pursue a policy degree in the first place and then taking time from her busy schedule as dean to be on my committee, and finally, to Doug Massey, who afforded me the thrill of learning something completely new during my last weeks as a student.

I have been so fortunate to end up as a graduate student in the Office of Population Research. The professors, researchers and staff at OPR create a culture where it is truly a joy to work and learn. I am especially grateful to Marta Tienda and Tom Espenshade for taking an interest in me and my work despite it being outside their own areas. Mary-Lou, Joyce, Wayne, Jennifer, Nancy, Judie and Lynn keep OPR running with professionalism, humor and a great deal of patience. Moreover, I've been able to spend my days with a truly inspirational group of fellow students: Danny, LaTonya, Stacie, Dennis, Beth, Laura, Lauren, Diane, Heidi, Eric, Tin-Chi, John, Kim, Alex, Michelle, Lucia, Carlos, Rohan and Dean have all shared ideas, insights, commiseration, coffee, pizza and advice which I have found invaluable.

I am also blessed with the most wonderful group of funny, clever, and wise friends. I couldn't possibly express how much I owe them, so thank you (in no particular order) to: Alice Meuhlhof, who somehow always made me laugh at myself and believe in myself simultaneously; Cally Ardington, who made me feel part of her family every time I visited ZA and answered far

more than her share of anxious emails; Lisa Markman-Pithers, for all the lunches, laughs and general sanity maintenance; Jeff Colgan for the beers, pep talks, endless mocking and occasional statistics insights; Marey Herrfeldt, for the phone teas, care packages to Africa, actually typing out all those episodes of "our show", and for renewing the friendship contract for another 20 years (phew); and to the MPA '06 All Stars (you know who you are) for coming back time and again to visit Princeton and report on "the real world", for never failing to make me laugh and make me think, and for all the love and support through it all- keep it coming.

And of course, I want to say thank you to my incredible and wonderful family members who have fretted me through this and every other academic step since kindergarten. They are the best support team any perpetual student could hope for- funny, kind, unwavering, supportive, and just mad enough to be interesting. I am very lucky to have them. And finally, thanks to Kevin, who makes every day (even the ones when you are writing a dissertation) so much fun.

Abstract:

This dissertation concerns itself with the effects of education and household wealth on health in Africa. The first two chapters focus on the impact of adult education levels on the HIV epidemic in four sub-Saharan countries. The last chapter turns to children's health and examines whether the urban advantage in health persists despite rapid urbanization in South Africa.

Chapter one employs longitudinal HIV data from Mali, Tanzania, Kenya and Zambia to examine whether the positive relationship between educational attainment and HIV prevalence is changing. I find evidence that the relationship between HIV and education has begun to reverse. Although it remains positive at the regional level, its much weaker for the youngest cohort. Furthermore, I find no association between HIV and education at the individual level among the youngest cohort. Secondarily, I test two explanations for such a change- the erosion of educational infrastructure and the adoption of protective knowledge among the educated. I find evidence consistent with the hypothesis that education is becoming protective as the epidemic matures; regions with higher levels of adult education at baseline experience larger drops in HIV prevalence over time.

Chapter two builds on these findings by examining whether the behavioral response to HIV is stronger among the more educated. I find a robust positive association between education and condom use, HIV testing, and age at marriage, with evidence that younger cohorts may be reducing age at marriage. I also find that more educated individuals are increasing their rates of HIV testing and reducing age of marriage more than the less educated.

Finally, I use anthropometric scores from two national surveys from South Africa to examine changes in urban and rural children's health over 15 years. I find that the urban health advantage disappears despite urban children retaining advantages in average household wealth. I then explore several common explanations for this pattern, including the growth of particularly vulnerable urban populations or deepening urban poverty. I find no evidence of deteriorated circumstances for the urban poor, although urban-rural migrants have begun to show a health disadvantage. I find that the differential gains are likely due to large improvements made by very poor rural households.

Dedication iii
Acknowledgements iv
Dissertation Abstract vi

Chapter 1: From Predictive to Protective The Changing Relationship of HIV and Education in Africa 1
Chapter Abstract 2
I. Introduction 3
II. Data and measures 6
III. Is the relationship between HIV and education changing? 8
IV. Examining potential pathways for change 13
Evidence for Educational Erosion 16
Evidence for Protective Knowledge 19
V. Testing the relationships at the individual level 22
VI. Conclusion 26
Tables 1.1-1.5 28
Figures 1.1a-1.6b 33

Chapter 2: Education and HIV Risk Behaviors 39
Chapter Abstract 40
I. Introduction 41
II. Education and HIV risk behaviors 43
III. Data and measures 46
IV. Results 51
Has education become unambiguously associated with more protective behaviors? 53
Are the more highly educated changing their behaviors more? 56
V. Conclusion 62
Tables 2.1-2.9 65
Figures 2.1-2.4 74

Chapter 3: The Urban Health Advantage in South Africa 78
Chapter Abstract 79
I. Introduction 80
II. Setting and empirical Strategy 83
III. Data and measures 85
IV. Results 88
Are the comparatively small gains in urban health a result of expanding vulnerable populations in urban areas? 92
Is deepening poverty depressing average urban health scores? 93
Can the different trajectories of urban and rural children be explained by distributional changes in socioeconomic characteristics? 98
V. Conclusion 102
Tables 3.1-3.7 104
Figures 3.1-3.5 111

References 118

Chapter 1: Predictive to Protective?

The Changing Relationship of HIV and Education in sub-Saharan Africa

Abstract:*

This analysis uses longitudinal HIV prevalence data from Mali, Tanzania, Kenya and Zambia to examine whether the positive relationship between educational attainment and HIV prevalence is changing. My findings support the hypothesis that the relationship between HIV and education has begun to reverse in Africa. Although I find that, at the regional level, the relationship is still positive across age groups, it is much weaker for the youngest generation. Furthermore, when I examine the relationship at the individual level, using region and country fixed effects, I find no positive association between HIV risk and education in the youngest cohort. Secondarily, I test two commonly hypothesized explanations for such a change- the erosion of educational infrastructure in high prevalence areas and the adoption of protective knowledge among the educated. I find evidence consistent with the hypothesis that education is becoming protective as the epidemic matures; regions with higher levels of adult education at baseline experience larger drops in HIV prevalence over the time period. I do not find evidence that educational attainment is eroding over time in higher prevalence areas. In fact, there is a consistent trend towards increased years of schooling for children in both high and low prevalence regions.

*This work was presented previously at Princeton's Office of Population Research Notestein Seminar Series in May 2011

I. Introduction

Mark Twain once famously quipped "Be careful about reading health books. You may die of a misprint." Despite his skepticism, there is a preponderance of evidence that in fact, people who read books (health or otherwise) actually enjoy longer life and better health. There is a robust body of research demonstrating that being well educated is associated with better outcomes across a broad range of health measures. Moreover, scholars have documented this relationship across a variety of contexts and time periods. In part, the inverse association between education and illness may be explained by access to information. People with more education may know more about health risks and consequently exhibit healthier behaviors. Education may also affect health indirectly through its association with wealth. More educated people tend to have more income, and higher incomes can confer a host of health benefits such as access to better care, less physically demanding work, better nutrition and lower levels of psychological stress (Cutler & Lleras-Muney 2006; Ross & Wu 1995). Consequently, having more years of education is generally viewed as protective against an array of disease risks.[1]

One very notable exception to this pattern has been HIV in sub-Saharan Africa. Data collected from Africa in the early decades of the epidemic demonstrated that there was a robust *positive* association between education and HIV infection (Gregson et al. 2001; Hargreaves & Glynn 2002; Fortson 2008). This association likely exists because many of the correlates of education are also risk factors for HIV. For example, areas with high numbers of educated people also tend to be urban, have more developed economies

[1] The health impact of education independent of wealth is a subject of much scholarship. I do not attempt to disentangle them here, but focus on educational attainment. I note, however, that Fortson (2008) uses DHS to test the associations between wealth and HIV risk versus education and HIV risk independently, and finds that education has a significant relationship with HIV rates, but wealth does not.

and transport networks, and larger migrant labor populations- all of which are associated with the spread of HIV. In addition, at the individual level, people who are more educated tend to marry later and may have more sexual partners before marriage, which can potentially increase exposure to HIV risk (Singh & Samara 1996; Gregson et al. 2001; Bongaarts 2007; Fortson 2009).

However, there are reasons to think that the relationship between education and HIV risk may be changing as the epidemic matures. A variety of scholars in the field have hypothesized that the present positive association may ultimately be replaced by the familiar negative relationship as populations respond to the development of the epidemic. (Gregson et al. 2001; Hargreaves & Glynn 2002; Fortson 2008). These hypotheses are largely based on two theoretical pathways through which HIV prevalence rates could directly reverse the relationship between HIV and education. The first, which I characterize as the *'educational erosion pathway'* explains the change via erosion of the educational system in high HIV prevalence areas. In this scenario, the causality runs from HIV to education, where high prevalence causes a widespread social decline, including the deterioration of educational infrastructure. Consequently, high HIV prevalence eventually comes to be associated with low educational attainment in the area.

The second pathway, which I call the *'protective knowledge pathway'* theorizes that more educated people may change their behaviors earlier as knowledge about HIV prevention disseminates- either because they have better access to information or possibly more agency or motivation for change. In this case, the causal pathway runs from education to HIV rates. It predicts that places with high numbers of educated people

will see more pronounced declines in HIV prevalence until high education rates come to be associated with low prevalence, a reverse of the current pattern.

This analysis sets out to investigate two related questions. I first ask whether there is evidence for a reversal of the positive relationship between HIV and education, and I then examine the available evidence for each of the two hypothesized causal pathways described above. I use data from Demographic and Health Surveys from the four sub-Saharan African countries with multiple rounds of HIV testing- Kenya, Tanzania, Mali and Zambia.

I approach the first question, whether the relationship between education and HIV is reversing, by splitting the individuals in the sample into three different age cohorts and looking at the relationship in successively younger cohorts. Because risk of HIV infection is hypothesized to relate both to individual behaviors and local risk environment, I examine this question at both the regional and individual level. My findings support the hypotheses of previous scholars that the relationship between HIV and education has begun to reverse in sub-Saharan Africa. Although I find that, at the regional level, the relationship is still positive across age groups, it is much weaker for the youngest generation. At the individual level, using country by region fixed effects, I find no association between HIV risk and educational attainment for the youngest cohort.

I then use multiple waves of HIV and education data to calculate the change in HIV prevalence over time and examine whether it relates to the initial level of education in the region. I find evidence consistent with the hypothesis that education may be becoming protective in the present state of the epidemic; regions with higher levels of

adult education at baseline experience larger drops in HIV prevalence. Further, this does not appear to be a result of either selective mortality or migration. I do not, however, find evidence that educational attainment is eroding over time in higher prevalence areas. In fact, there is a consistent trend towards increased years of schooling for children in both high and low prevalence regions.

The remaining sections of the paper are organized as follows. The next section introduces the data and describes the demographic characteristics of the sample regions and respondents. Section 3 analyzes evidence for a change in the relationship between education and HIV at the regional level. Section 4 uses regional level data to examine the available evidence for the *'educational erosion pathway'* and the *'protective knowledge pathway'*. Section 5 uses a subsample of the data to test whether the change in the education-HIV relationship identified at the regional level applies to individuals when country by region fixed effects are included, and section 6 concludes.

II. Data and Measures

The data for this analysis is drawn from the Demographic and Health Surveys (DHS). The DHS are large nationally representative household surveys that cover household members' education, nutrition, fertility, and health. More recently, some African DHS have asked a subsample of adult respondents to provide blood samples for HIV testing[2]. The four countries examined here - Kenya, Mali, Tanzania and Zambia- are, to date, the only ones for which two rounds of HIV testing data are available. For

[2]Response rates were high and analysis on a subset of countries has shown that bias from non-response is small (Mishra et al. 2006).

Mali and Tanzania, the two waves of HIV and household data were collected approximately five years apart. In Mali, HIV tests were collected in 2001/02 and 2006. In Tanzania they were collected during 2003/4 and again in 2008. For Kenya and Zambia, the two waves were collected six years apart-in 2003 and 2009 in Kenya and in 2001 and 2007 in Zambia. The analysis here rests upon measures of HIV prevalence and educational attainment calculated at the regional level at the two points in time. Regions are identified within each country by using the geographic regions as specified in the DHS and further dividing them by urban and rural sectors where applicable, resulting in 91 distinct regions. I calculate regional HIV prevalence using information on 39,913 adults aged 15-49 in the HIV tested subsample.[3]

I also use two measures of regional educational attainment, which are calculated using the reported education of the 149,116 adults aged 15-49. The first measure is mean years of completed education among adults in the region. However, because some of the respondents are still quite young and could still be completing education, I also use percent of adults aged 15-49 who have completed primary schooling as an additional measure of regional educational levels. Both of the education measures come from the household questionnaires. In the section testing the educational erosion hypothesis, I examine changes in children's schooling over time. To this end, I use the responses to the household survey to calculate regional attendance rates and mean years of completed education based on 82,304 children who are aged 7-14 in the sample households.

Table 1.1 presents summary statistics of regional HIV prevalence and regional educational outcomes by country and wave. Kenya, Tanzania and Zambia are fairly

[3] DHS surveys provide HIV sample weights to adjust the tested subsample to be representative of the sample as a whole, and I use them here to compute regional HIV prevalence.

typical of the East and Central African region in that they all have relatively high levels of adult education and correspondingly high levels of HIV. Mali is the lone West African country and has a different pattern with lower levels of adult education and lower levels of HIV. Over the course of the two waves, there is a general trend of a slight increase in mean years of education completed and relatively little change in percentage of adults with primary education. There are also modest drops in mean regional HIV prevalence across all four sample countries over time. As mentioned above, the analysis relies on regional aggregations of individual level data, and Table 1.2 contains basic descriptive information on the individuals from which the regional means are constructed. Within both the adult and child samples, the basic age distributions are comparable across both countries and waves. Consistent with the regional means in Table1.1, there is an upward trend overall in educational attainment and Mali has lower measures of education for both adults and children than the other countries.

III. Is the Relationship Between HIV and Education Changing?

I approach the question of whether the relationship between education and HIV is changing by splitting the sample into three different age cohorts and looking at the relationship in successively younger cohorts. I divide the adults into a 'youngest' cohort who are aged 15 to 24 in the first wave, a 'middle' cohort who are aged 25 to 34 in the first wave, and an 'oldest' cohort who are aged 35 to 44 in the first wave. I am then able to follow the cohorts as they age in the second wave, which is either five or six years later, depending upon the country. The youngest cohort is of particular interest because they would have been born between 1976 and 1987 and entered adolescence during the

1990s and early 2000s. In effect, this group is the first that could have begun making decisions about sexual behaviors in a period when the transmission pathways of HIV were widely accepted and the means of prevention were known and publicized.

To begin examining whether the relationship between HIV prevalence and education levels differs by cohort, I estimate a simple model of the relationship between regional HIV prevalence and education for each cohort as below:

$$(1) \qquad HIV_{rcw} = \beta_0 + \beta_1 Ed_{rcw} + \theta X_{rcw} + \varepsilon_{rcw}$$

where the dependent variable HIV_{rcw} is the regional HIV prevalence for adults in region *r*, within age cohort *c* at wave *w*. Independent variable Ed_{rcw} is the regional mean for the educational outcome (either 'years of education' or 'primary school completion rate') for adults in age cohort *c* and region *r* at wave *w*. X is a vector of control variables for the same region, cohort and wave, including year of the survey measurement and indicators identifying each country.

Table 1.3 shows the coefficients of ordinary least squares regressions of regional HIV prevalence on the mean regional education level for each cohort in the first wave. Figures 1.1a and 1.1b show the same regression lines in graphical form. For both measures of education, the relationship is positive and statistically significant across age cohorts. However, the coefficient for the youngest cohort is less than half those for the older two cohorts. For adults between 25 and 34, one additional year in the mean education for adults in the region is associated with a correspondingly higher level of HIV prevalence of about 2.6 percentage points, whereas for 15-24 yr olds, the same

difference in mean years of education is associated with only about 1.2 percentage point greater HIV prevalence. Further, this difference between cohorts is statistically significant (pval<0.01).

A similar pattern exists for primary school completion. Whereas 25-34 year olds see an approximately 0.27 percentage point higher HIV prevalence associated with a one percentage point increase in the share of adults who completed primary school, the younger cohort has a much weaker association, with a 0.1 percentage point greater HIV prevalence associated with the each one percentage point increase in the share of adults who completed primary school.

This weaker relationship between HIV and education for the younger age cohort is consistent with the idea that the strong association between education and HIV is beginning to change. Assuming there is consistency in education levels within regions across cohorts, young people in highly educated areas are potentially still exposed to more risk, in part due to the residual higher prevalence in their communities. However, the attenuation of the education-HIV association suggests that they may be also altering behaviors more at the individual level to protect themselves.

However, protective behavioral changes are not the only possible explanation: the weakening of the association depicted in Figures 1.1a and 1.1b could also be consistent with an association that simply strengthens as people age. As time goes on, and the young cohort experiences a longer period of sexual activity, those in highly educated regions may contract the diseases at higher rates and ultimately the relationship will look exactly as it does for the older cohorts. Because I have two rounds of HIV testing data, I am able

to examine this possibility by looking at the progression of the association for all the cohorts over a 5-6 year period.

Figures 1.2a and 1.2b overlay the regression lines for the same three age cohorts when education and HIV prevalence is measured again in second wave. The solid lines are the same regression lines from Figures 1a and 1b. The dotted line represents model (1) repeated in wave 2 for the youngest cohort, now approximately aged 20-31. The dashed lines similarly represent the relationship for the older cohorts as they are measured in wave 2, either five or six years later.[4] The corresponding coefficients are also presented in Table 1.3. If the relationship between HIV and education strengthens as the cohort ages, I would expect to see the slope of the dotted line begin to rotate up – approaching the slopes observed for the older cohorts -and the OLS coefficient on the education measure to increase from wave to wave. In fact, it does not. HIV prevalence goes up somewhat among the younger group at all education levels, but the slope of the line does not change in any substantive way. The coefficient estimate on 'years of education' which is 0.012 in the first wave decreases to 0.011 in the second wave and the difference between the two is non-significant. The pattern for the middle cohorts is similar, with a small statistically non-significant change in the coefficient on years of education or primary school completion rate. The oldest cohort actually demonstrates the most change in the coefficient between waves. The slope of the regression line rotates considerably downwards as the group ages, and the coefficient on years of education drops from 0.030 to 0.020.

[4] Part of the oldest age group ages out of measurement over the course of the two waves (i.e. they pass 49 yrs of age and are no longer asked the individual questionnaires), As such, the wave 2 measurement is based on only the younger members of this cohort.

This pattern could indicate several things about the oldest age group. It is possible that less educated people in the older group are contracting HIV at a much higher rate than the more educated. But it is also possible that selective mortality by education level is driving the observed change. While I test for selective mortality by education for all three age cohorts below, I am most concerned about it for this age group- who are the ones who are more likely to have contracted HIV early in the epidemic and therefore the most likely to have been infected long enough to be experiencing high levels of AIDS mortality. If there is selective mortality of the highly educated, then the mean education levels of the cohort should decrease over time. Although it is not statistically significant, I do find a slight drop in mean education levels of this cohort between waves.[5] In addition, a small part of this cohort ages out of the survey between wave one and two. Because of these two factors, I treat the estimates from this cohort with caution, and the remaining analyses presented are conducted both including this age cohort and excluding it, to ensure the results are robust to their exclusion. Ultimately, the inclusion or exclusion of this cohort does not change the substantive results of the analysis, but I focus the remaining sections on the younger two cohorts, for whom I have complete information in both waves, and for whom there is no evidence of selective mortality by education.

Collectively, I interpret the differences between cohorts illustrated in this section as early evidence that the relationship between HIV and education at the regional level is beginning to change. Moreover, this change is concentrated among younger people, who both have lower rates of HIV prevalence and a weaker association between regional

[5] In contrast, the younger two cohorts have slightly positive changes, although they are also not statistically significant. A more detailed treatment of the selective mortality test is included in section IV.

education rates and regional HIV prevalence. In an additional analysis, which is not shown, I repeated the above separately for men and women but did not find substantively different results by gender.

IV. Examining Potential Pathways for Change

The observation that the positive relationship between education and HIV is weakening across age cohorts, although interesting in itself, does not account for the causes or pathways for change. For policy makers and scholars interested in influencing the course of the epidemic, identifying the mechanism of change is of deep interest. The literature implies at least two pathways through which this reversal could take place. The first is through a general societal deterioration resulting from high HIV rates. For some time, observers and scholars have warned that the particular characteristics of the HIV epidemic– that it targets working age adults, that it clusters in households, that the rates of infection are higher among the educated – mean that the effects of the disease will go far beyond the afflicted households and result in widespread societal deterioration. In this scenario, schools, hospitals and civil institutions deteriorate as a result of the deaths of a large swath of the educated laboring population. In addition, parents in affected communities, who perceive a high risk of early mortality, may reduce investment in their children's education because they do not expect to live long enough to reap the benefits of the investments. Over time, this process (which I refer to subsequently as *educational erosion*) could plausibly work to reverse the positive relationship between HIV and education. If enough erosion of educational infrastructure and investment were to occur,

high HIV prevalence regions could eventually become associated with lower levels of education rather than higher.

Recent research lends some support to these bleak projections. Fortson (2011) uses Demographic and Health Surveys in sub-Saharan Africa to show that areas with higher HIV prevalence have larger schooling declines among those who were of school going age when the epidemic hit. Case and Paxson (2011) document deterioration of health infrastructure in areas of high HIV prevalence. Behrman et al (2010) show, in a high prevalence area of Malawi, that mothers who perceive a higher risk of HIV reduce investment in children's education. In addition, Akbulut-Yuksel and Turan (2010) use the DHS surveys to document that children in high HIV prevalence areas accumulate comparatively less education than their parents, even within households without HIV infected members. Collectively, this research suggests that HIV is associated with losses in education and health that go beyond the individuals, or even the generation, directly infected.

The second path through which the relationship between education and HIV might reverse is simply through the dissemination of information about the disease and its pathways of infection. Historically, as disease mechanisms became better understood, the more highly educated have adopted protective behaviors earlier and an education gradient in mortality and morbidity has subsequently emerged. For example, in the 1960s, there was little difference in mortality for older American men across education levels, but once the risks of smoking became understood (beginning with the release of the Surgeon General's report on smoking in 1964), better educated men's mortality dropped throughout the 70s and 80s, becoming significantly below that of men with less

education (Cutler & Lleras-Muney 2006; Crimmins & Saito 2001; Preston & Elo 1995; Feldman et al. 1989). Some researchers have suggested that a similar process may take place with HIV if people with more education respond differently as information about the disease disseminates (De Walque 2007; Gregson et al. 2001).

In the case of HIV in Africa, it is plausible that this process is only now taking place. Transmission pathways were imperfectly understood until the late 1980s and early 1990s. Even once transmission pathways were identified in the scientific community, disinformation and political resistance to the new knowledge slowed its impact in sub-Saharan Africa. Influential figures such as the then President of South Africa publicly questioned the scientific consensus, effectively obfuscating the public narrative about how HIV was spread and how it might be prevented. By the 2000s, however, the public skepticism about the pathways to HIV infection had been largely overcome, and there now seems to be widespread acceptance in both the scientific and popular press on the causes of HIV. Now that the knowledge of how HIV/AIDS spreads has become better accepted in sub-Saharan Africa, I might expect to see the more highly educated adopting protective behaviors first, leading the relationship between education and HIV eventually to develop the more familiar inverse association.

Recent work by Case and Paxson (2011) lends support to this theory (which I refer to henceforth as the '*protective knowledge*' pathway). The paper examines behavioral changes among younger women and finds that young women in high prevalence areas respond by decreasing teen sexual activity outside of marriage and by entering marriage earlier. More importantly for this analysis, they find that the behavioral effect on early marriage is magnified for highly educated girls. These findings suggest

both that younger cohorts of Africans may be changing their behaviors in response to the epidemic, and that there may be some differential in response by education level. If the more highly educated adopt protective behaviors earlier and/or at higher rates, there may ultimately be a reversal in the positive correlation between HIV and education.

If the changes observed in these data reflect behavioral reactions among the educated resulting from the dissemination of prevention knowledge, then it's a hopeful development in the efforts to curtail the epidemic. Conversely, if they are due to the deterioration of educational infrastructure in previously highly educated areas, then it is hardly good news. Although their implications differ, there is no reason that these two mechanisms are exhaustive or mutually exclusive. Both could be happening simultaneously. The remaining sections of the paper begin to explore the evidence for the possible pathways for this change, looking first at the 'educational erosion' pathway and secondly at the 'protective knowledge' pathway.

Evidence for Educational Erosion

As noted above, much recent work on HIV in sub-Saharan Africa has documented an association between high prevalence and deterioration in human capital. To my knowledge, this data offers the first opportunity to examine both the changes in HIV and education longitudinally in several nationally representative datasets. If high HIV prevalence is causing a general deterioration in educational infrastructure, I would expect to see education levels to fall over time in high prevalence areas. (Note that high prevalence areas are generally more educated to begin with, so the overall levels of both adults and children will be higher there.)

In order to assess the possibility that HIV prevalence negatively impacts provision or access to education, I compare changes in educational levels over time across regions with varying levels of prevalence. Specifically, I look at children aged 7-14 in both waves, and test whether the education levels of children in high HIV prevalence areas fall (or rise less) than those in lower prevalence areas. I concentrate on children for two reasons; first because the educational attainment of adults is largely completed and unlikely to change much over the observed time period; and second, because of the potential bi-directionality of the HIV/education relationship among adults. School age children, in contrast, are far less likely to have contracted HIV themselves, and so losses in their educational attainment speak more to generalized erosion of schooling in communities than to the direct consequences of infection. Once again, I use two different measures of regional education levels. The first is a measure of share of children aged 7-14 who are attending school. However, the corrosive effect of HIV might not be seen simply in attendance rates, but also in children's overall completion or progress through school. As such, I also use average years of completed education for the 7-14 age group. This could be a problematic outcome variable if there were large differences in age distribution across the samples. For example, if one country or region has many more very young children, then the average number of years of schooling completed might be much lower, but there wouldn't necessarily be reason for concern. However, the age distributions, as shown in Table 1.2, are almost identical across countries and waves. For both educational outcomes, I estimate the least squares regression for the model,

$$\Delta Ed_{rw2-rw1} = \beta_0 + \beta_1 HIV_{rw1} + \theta X_{rw1} + \varepsilon_{rw1} \quad (2)$$

in which $\Delta Ed_{rw2-rw1}$ is the change in the regional mean for the educational outcome (either 'average years of education completed' or 'school attendance rate') for children in region *r* between wave 1 and wave 2. Independent variable HIV_{rw1} is the *adult* regional HIV prevalence for region *r* in wave 1, and X_{rw1} is the same vector of regional control variables from model (1). If high HIV prevalence is leading to a general erosion of educational opportunity, I would expect to see a drop in children's educational attainment between the waves (or perhaps smaller rises) in places with high adult HIV prevalence at baseline. Figures 1.3a and 1.3b show the regional changes in children's educational outcomes in relation to the adult HIV prevalence in wave 1. Table 1.4 shows the regression coefficients which correspond to the regression lines depicted in the figures. As you can see in the figures, the general educational trend is positive. Educational outcomes for children in most areas changed for the better over the period covered by the two waves. I do not find any statistically significant evidence that the positive trend in educational attainment is slowed or reversed in regions with higher HIV prevalence. Although the coefficients on the OLS regressions are negative, they are not significant either in aggregate or separated by gender.

In summary, contrary to the previous work cited, I don't find evidence for educational erosion in high HIV areas. I interpret these results with caution, however, since the sample is limited to 91 regions. It is possible that, as more longitudinal data becomes available, there will be stronger evidence either for or against the educational erosion pathway.

Evidence for Protective Knowledge

To examine the second pathway through which the relationship between education and HIV might change, I ask whether areas with higher education levels experience greater decreases in HIV prevalence. If education has become protective as a result of dissemination of knowledge, I might expect to see more highly educated people adopt protective behaviors first and, consequently, see HIV prevalence fall or rise more slowly in more highly educated areas. To examine this possibility, I look at the change in the adult HIV prevalence between the waves in relation to the mean adult education levels in the region at baseline. To do this, I use least squares regression to estimate the equation,

$$\Delta HIV_{rcw2-rcw1} = \beta_0 + \beta_1 Ed_{rcw1} + \theta X_{rcw1} + \varepsilon_{rcw1} \quad (3)$$

where $\Delta HIV_{rcw2-rcw1}$ is the change in HIV prevalence between wave 1 and wave 2 for adults in region *r*, within age cohort *c* . Ed_{rcw1} is the regional mean for the educational outcome (either 'years of education' or 'primary school completion rate') for adults in age cohort *c* and region *r* at wave 1. X is the vector of regional control variables from model (1) including survey year and country indicators. Table 1.5 reports coefficient β_1 for model (3) for both education measures. Results are presented for the population as a whole and independently for the younger and middle age cohorts. There is a statistically significant association between the education level at baseline and the trajectory of HIV prevalence for both measures of education. One additional year of mean adult education is associated with -0.7 percentage point difference in the change in regional HIV

prevalence between the two waves. In other words, regions with higher adult education experience larger falls (or smaller rises) in HIV prevalence (as shown graphically in Figures 4a and 4b) and this pattern appears consistent across age cohorts and in both high and low prevalence countries.[6] The finding is also consistent for both men and women when model (3) is run separately by gender.

While this pattern is supportive of the theory that education has begun to become protective, it is equally consistent with selective mortality by education level. If, as we know, highly educated people also experienced higher HIV rates early in the epidemic, then it is possible that there is differential mortality by education. That is, those more educated people that first contracted the disease are dying at a higher rate, bringing the prevalence down among the educated via mortality rather than via averted cases. Unfortunately, I cannot distinguish HIV mortality from other types of mortality with this data, so it is difficult to examine this directly. However, I can test for differential mortality by education by looking at whether education levels drop among the same age group over time. Since respondents don't normally lose years of education once they have accumulated them, the average education levels of a cohort should only be able to rise or stay the same as time goes on. If average education drops, then it is very likely that more educated people are dying at a higher rate. As mentioned above, I might expect to find this effect most among the older cohorts since the youngest are less likely to have begun experiencing high mortality in the relatively short time periods covered by the two waves.

[6] The coefficient on regional mean adult years of education and mean primary completion for Mali (the lone low prevalence country) are not statistically significant on their own as there are only 16 regions in Mali (and therefore very low statistical power) but the pattern is the same. i.e. the coefficients are negative and only slightly smaller than those for the high prevalence countries.

For each of three cohorts, I calculate the mean years of education and primary school completion at baseline, and then again in the second wave, and test to see if there has been a fall in mean education levels. I find that, for all three cohorts, there is no statistically significant change in years of completed education or primary school completion rate over the period in question. For the youngest cohort, both the mean years of education and primary school completion increase slightly, by .39 years and .06 percentage points respectively. However, the change is not statistically significant. The middle cohort sees virtually no change in education. For the oldest cohort, the mean regional education values decrease by -.19 years of education and -.037 percentage points for primary completion (although, again, this decrease is not statistically significant). I take this as evidence that differential mortality by education is unlikely to be driving the observed changes in the education-HIV relationship. However, I also calculate model (3) for just the younger two cohorts, to account for the possibility that there is differential mortality by education in the oldest cohort which is not captured adequately by the test here. The findings are robust to the exclusion of the oldest group and the correlation between high adult education levels and reduction in regional HIV prevalence is statistically significant and of similar magnitude when only the younger cohorts are analyzed.

Another potential explanation for the patterns observed here is selective migration by HIV status. Prior research has shown that there is often a positive relationship between HIV status and migration experience (Anglewicz 2007) and that HIV positive individuals commonly return home to rural areas when they become sick (Clark et al. 2007). It is possible that the patterns described thus far are driven by people who are

infected with HIV leaving places of high average education (such as metropolitan areas) to return home. If this were the case, the falls in prevalence in these areas have less to do with high average adult education levels than with high baseline HIV rates causing high out-migration. However, because individuals in the survey report how long they have been living in their current place of residence, its possible to identify people who migrated during the period covered by these two waves. In order to exclude the hypothesis that my results are driven by selective migration, I restricted the sample to non-migrants and re-ran the analysis to see if the relationships described above held in the absence of migration. I find that they do, which suggests the falls in regional prevalence are unlikely to be solely a result of selective migration away from areas with high average adult education.

V. Testing the Relationships at the Individual Level

As with most communicable diseases, the relationship between individual behaviors and HIV risk is partly due to individual choices but also related to the environment in which the individual resides. Two individuals in different regions who chose to engage in unprotected sex are making the same behavioral decision, but the one who lives in an area with high prevalence is at a higher risk of contracting HIV from a partner than the one who lives in an area with low prevalence. As noted previously, this could partly explain why the relationship between HIV and education weakens for the youngest cohort but does not disappear entirely in the regional level analysis. Because HIV risk has both a regional and an individual behavior component, as a final step, I move from comparing averages across regions and compare individuals within the same

region. I ask, *within the same country and region* do more educated individuals have higher risk than the less educated, and does that change over time?

While using individual level data to test the relationship within regions is logical step, I am somewhat limited by the structure of the available data. Because the initial HIV surveys in Mali and Zambia were not linkable to the household survey, I cannot link individuals to their HIV results. Therefore, for those two countries, I cannot test variation in HIV prevalence as it relates to an individual's characteristics, such as education, age or gender. I am limited to the adult samples from Tanzania and Kenya for this part of the analysis. In total, I am able to link HIV test data to complete individual data for 24,943 individuals in the youngest and middle cohorts.

Using individual questionnaire responses, I begin by testing the underlying premise that adults in the sample with more education actually are more knowledgeable about HIV/AIDS prevention. I use the same two education measures at the individual level (years of schooling completed and a binary measure of whether the individual completed primary school) and I find that both education measures are positively and statistically significantly associated with HIV knowledge.[7] More educated people are more likely to know that a healthy looking person can be infected with HIV and that HIV risk can be minimized by regular condom use, abstinence, and/or minimizing concurrent sexual partners. Education is also negatively associated with incorrect beliefs that HIV can be transmitted through mosquito bites or sharing food with an infected person. This

[7] As only 4% of the sample has more than 12 years of education, I top code years of education at 12 yrs, but results are robust to using 16 years of education as well.

association between education and HIV/AIDS knowledge is consistent across cohorts and across both waves.[8]

I next test whether having more years of education translates to lower prevalence among the educated. As a first step, I compare the probability of testing HIV positive at each year of educational attainment for the two cohorts in each wave. Figure 1.5 shows this relationship for both cohorts in the first and second wave. Several things are evident from this simple cross sectional relationship. The first is that the youngest cohort is at lower risk across all education levels in both waves. The second is that there is no positive relationship between HIV and education among the younger cohort at the individual level. Furthermore, for individuals at the lower end of the educational scale, the prevalence of HIV increases considerably between waves. It appears that for individuals in the youngest cohort, not only has the relationship between education and HIV ceased to be positive, it may actually be becoming negative. I test this relationship more directly by splitting the sample into adults with primary education or more, and those with less than a primary education. I then run a difference of differences model by estimating,

$$(4) \quad HIV_{icw} = \beta_0 + \beta_1 Primary_{icw} + \beta_2 W2 + \beta_3 Primary_{icw} * W2 + \theta X_{icw} + \varepsilon_{icw}$$

where HIV_{icw} is the probability that individual *i* from cohort *c* in wave *w* tests positive for HIV, and $Primary_{icw}$ is a binary indicator of whether that individual reported having

[8] This analysis is not shown here, but is available on request. I run probit regressions of an individual's probability of correctly answering questions about HIV on the education measures. I control for the individuals age, age squared, urban residence, country, region, gender and the year of measurement. All results reported are statistically significant at 5% level or higher.

completed primary school or more. $Wave2$ is an indicator for whether the observation is during the second wave of the survey, and X_{icw} is a vector of controls including age, gender, urban residence and country by region fixed effects. The coefficient β_3 represents the difference-in-difference estimator. In other words, β_3 is an estimate of how much the change in HIV prevalence between waves for people with primary education differs from the change in HIV for people with less than a primary education. I find that for the youngest cohort, β_3 is statistically significant and equal to -0.016 (pval<.05). In other words, the probability of being HIV positive for an individual with less than a primary education increases by 1.6 percentage points more than it does for someone with a primary education or more. [9] For the middle cohort, the coefficient estimate of β_3 is -0.009 and is non-significant.

Finally, I estimate the above model for each year of completed education rather than simply splitting the respondents by completion of primary school. Figures 6a and 6b show the estimates and confidence intervals for the change in probability of being HIV positive at each year of completed education, first for the youngest cohort (6a) and then for the middle cohort (6b). While most of the estimates are non-significant, a pattern is detectible for the youngest cohort, with lower years of education being associated with positive changes and higher years of education being associated with little change. For the middle cohort, there is not a clearly discernable pattern. I interpret this as evidence that, at the individual level as well as the regional level, the relationship between

[9] Estimates are based on a sample of 14,323 HIV tested individuals in the youngest cohort, and 10,620 HIV tested individuals in the middle cohort.

education and HIV has begun to change for the youngest generation.[10] Whereas for the older generations, education was a risk factor for HIV infection, it now appears to be comparatively protective for the young, at least at the primary school level and above.

VI. Conclusion

The analysis above takes advantage of some of the first available nationally representative and longitudinal evidence on both HIV prevalence and educational trends. My findings support the hypotheses of previous scholars that the relationship between HIV and education has begun to reverse in sub-Saharan Africa. Although I find that, at the regional level, the relationship is still positive across age groups, it is much weaker for the youngest generation. For the youngest cohort in Tanzania and Kenya, I also find that, within the same country and region, individuals with more education do not have higher risk of HIV infection.

In my examination of potential pathways for this change, I find support at both the regional and individual level for the idea that education is becoming more protective, perhaps as knowledge about the disease disseminates. Regions with the highest levels of adult education saw the largest drops in HIV prevalence over time and younger people with a primary education or higher experienced less increase in HIV prevalence than those in the same age cohort with less education. I do not, however, find evidence of general erosion in education in high prevalence regions. Although the evidence on the pathways is not conclusive, it does support a more hopeful interpretation of the changes in the education-HIV relationship. Rather than being evidence of general societal

[10] In an additional analysis, I performed the same differences-in-differences estimation at the regional level (comparing regional changes in HIV prevalence across waves for primary educated and those with less than a primary education) and results were equivalent to the individual level analysis.

deterioration as a result of AIDS, the changes may reflect some success for the educational campaigns of the last two decades.

Table 1.1: Regional HIV and Educational Outcomes by Country and Wave

		Kenya (2003)	Kenya (2009)	Mali (2001)	Mali (2006)	Tanzania (2003)	Tanzania (2008)	Zambia (2001)	Zambia (2007)	Total (Wave 1)	Total (Wave 2)
HIV Rates	Mean	0.076	0.063	0.015	0.012	0.081	0.069	0.173	0.154	0.087	0.075
	Standard Error	0.010	0.015	0.003	0.002	0.010	0.009	0.018	0.014	0.008	0.007
	25th Percentile	0.041	0.033	0.004	0.007	0.041	0.031	0.102	0.117	0.035	0.021
	75th Percentile	0.114	0.079	0.021	0.017	0.107	0.097	0.224	0.199	0.117	0.116
Adult Years of Completed Education	Mean	7.28	8.14	2.15	2.25	6.08	6.08	6.90	7.24	5.75	5.97
	Standard Error	0.66	0.59	0.37	0.36	0.18	0.19	0.39	0.38	0.25	0.25
	25th Percentile	6.45	7.24	0.86	1.04	4.86	5.09	5.39	5.57	4.68	5.01
	75th Percentile	9.53	9.67	3.00	3.25	7.13	7.24	8.32	8.59	7.29	7.56
Percent Adults with Primary Education	Mean	0.57	0.65	0.19	0.19	0.69	0.68	0.61	0.64	0.57	0.58
	Standard Error	0.06	0.06	0.02	0.02	0.03	0.03	0.05	0.04	0.03	0.03
	25th Percentile	0.44	0.52	0.06	0.07	0.55	0.57	0.43	0.43	0.43	0.43
	75th Percentile	0.76	0.81	0.28	0.29	0.80	0.79	0.79	0.79	0.77	0.79
	Number of Regions	15	15	16	16	42	42	18	18	91	91

All regional means calculated based on men and women aged 15-49, HIV prevalence calculated using DHS supplied HIV weights, Educational means calculated using DHS sample weights

Table 1.2: Sample Demographics by Country and Wave

	Kenya 2003	Kenya 2009	Mali 2001	Mali 2006	Tanzania 2003	Tanzania 2008	Zambia 2001	Zambia 2007	Total
Adults Aged 15-49									
Mean Age	27.90	28.33	28.96	28.50	28.45	28.09	27.54	28.11	28.30
	(0.07)	(0.07)	(0.06)	(0.06)	(0.08)	(0.08)	(0.07)	(0.08)	(0.02)
Mean Years of Education	7.45	7.85	1.80	2.16	5.77	5.73	6.32	7.08	4.99
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
Proportion w/ Primary Education	0.58	0.61	0.15	0.18	0.64	0.65	0.55	0.62	0.44
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	17,464	16,937	25,925	29,442	13,742	14,167	16,469	14,970	149,116
Percent of Total Sample	12	11	17	20	9	10	11	10	100
Children Aged 7-14									
Mean Age	10.49	10.41	10.26	10.22	10.34	10.40	10.33	10.44	10.34
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)
Mean Years of Education	2.42	2.97	1.02	1.47	2.46	2.06	2.11	2.50	1.96
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Proportion Attending School	0.72	0.83	0.35	0.42	0.76	0.74	0.67	0.76	0.60
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
N	8,313	8,608	15,645	17,087	7,369	7,656	8,951	8,675	82,304
Percent of Total Sample	10	10	19	21	9	9	11	11	100

Standard Errors in Parentheses, Means calculated using DHS sample weights

Table 1.3: OLS of Regional HIV Prevalence on Regional Education by Cohort and Wave

	Regional HIV Prevalence Youngest Cohort		HIV Prevalence Middle Cohort		HIV Prevalence Oldest Cohort	
	Wave 1 (15-24)	*Wave 2 (20-30)*	*Wave 1 (25-34)*	*Wave 2 (30-40)*	*Wave 1 (35-44)*	*Wave 2 (40-49)*
Completed Years Education for Youngest Cohort in Region	0.012 **	0.011 **				
	(0.003)	(0.003)				
Completed Years Education for Middle Cohort in Region			0.026 ***	0.021 **		
			(0.004)	(0.005)		
Completed Years Education for Oldest Cohort in Region					0.030 ***	0.020 **
					(0.006)	(0.007)
Percent with Primary Education in the Region among Youngest Cohort	0.103 **	0.126 ***				
	(0.035)	(0.034)				
Percent with Primary Education in the Region among Middle Cohort			0.270 ***	0.259 ***		
			(0.037)	(0.052)		
Percent with Primary Education in the Region among Oldest Cohort					0.313 ***	0.232 ***
					(0.056)	(0.084)
N	91	91	91	91	89	89

Based on adults aged 15-44 in first wave, regional HIV prevalence calculated using DHS provided HIV weights, regional education means calculated using provided sample weights, robust standard errors are in parentheses and additional controls are included for year of measurement and country.

Table 1.4: OLS of Change in Regional Educational Outcomes of Children 7-14 on Regional HIV Prevalence of Adults

	Change in Percent Attending School in Region		
	All Children	Girls	Boys
HIV Prevalence among 15-49 Year Olds in Region in Wave 1	-0.194	-0.161	-0.185
	(0.114)	(0.161)	(0.142)
Number of Regions	91	91	91
	Change in Average Years of Completed Education in Region		
	All Children	Girls	Boys
HIV Prevalence among 15-49 Year Olds in Region in Wave 1	-0.907	-0.152	-1.550
	(0.584)	(0.817)	(1.270)
Number of Regions	91	91	91

Educational outcomes of children aged 7-14 taken from DHS household questionnaire and calculated using provided sample weights, Regional HIV prevalence calculated based on adults 15-49 and using HIV survey weights, robust standard errors in parentheses and an additional control for year of measurement and duration of epidemic is included but not shown; * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 1.5: Change in Regional HIV Prevalence on Regional Education Levels

	Regional Change in HIV Prevalence: *All Cohorts* (15-44 in Wave1)	Regional Change in HIV Prevalence: *Youngest Cohort* (15-24 in Wave1)	Regional Change in HIV Prevalence: *Middle Cohort* (25-34 in Wave1)
Completed Years Education of Adults 15-44 in First Wave	-0.007 ***		
	(0.002)		
Completed Years Education of Adults 15-24 in First Wave		-0.006 *	
		(0.003)	
Completed Years Education of Adults 25-34 in First Wave			-0.010 *
			(0.005)
Percent of 15-44 yr olds with Primary Education in First Wave	-0.059 **		
	(0.023)		
Percent of 15-24 yr olds with Primary Education in First Wave		-0.054 *	
		(0.031)	
Percent of 25-34 yr olds with Primary Education in First Wave			-0.100
			(0.053)
N	91	91	91

Based on adults aged 15-49, regional HIV prevalence calculated using DHS provided HIV weights, education means calculated using provided sample weights, robust standard errors in parentheses and an additional controls include year of measurement and country indicators. * p<0.05, ** p<0.01, *** p<0.001

Figure 1.1a: Relationship between Regional HIV Prevalence and Mean Years of Education by Cohort in Wave One

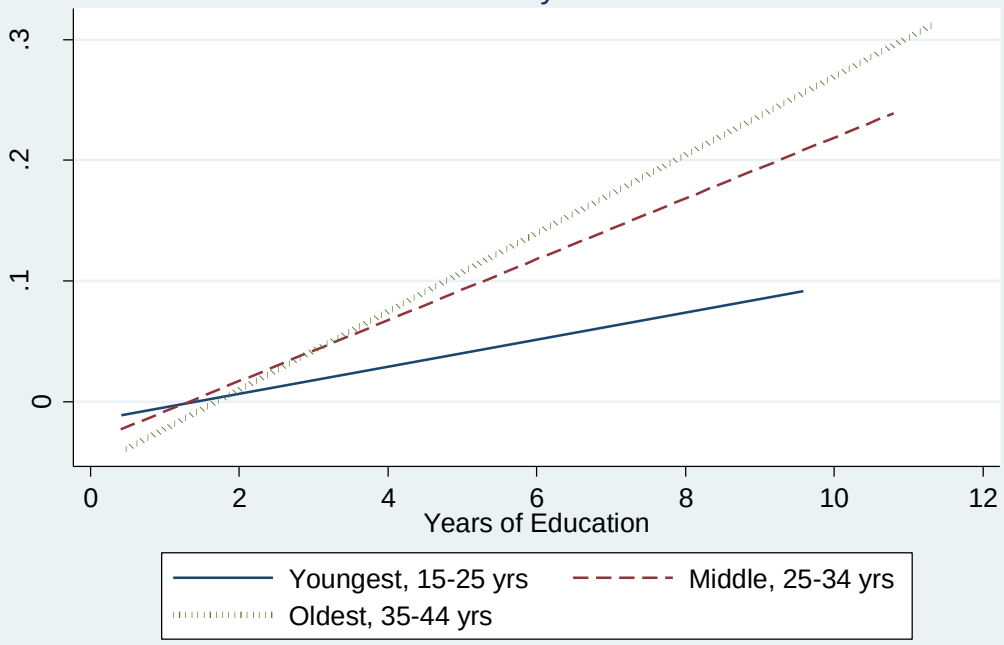


Figure 1.1b: Relationship between Regional HIV Prevalence and Primary Education Completion Rate by Cohort in Wave One

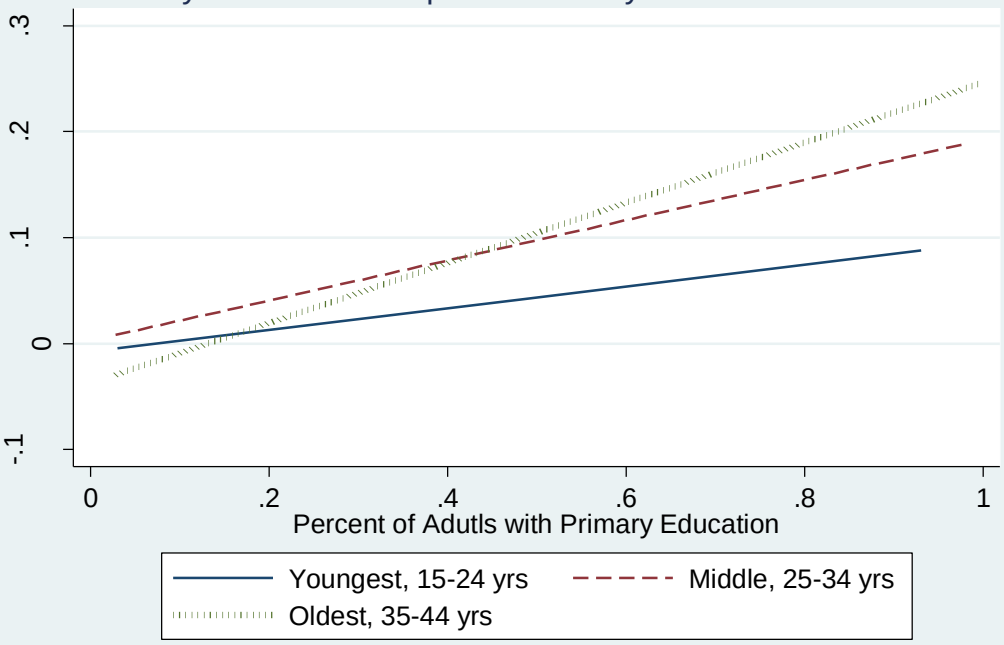


Figure 1.2a: Relationship between Regional HIV Prevalence and Years of Education by Cohort in Both Waves

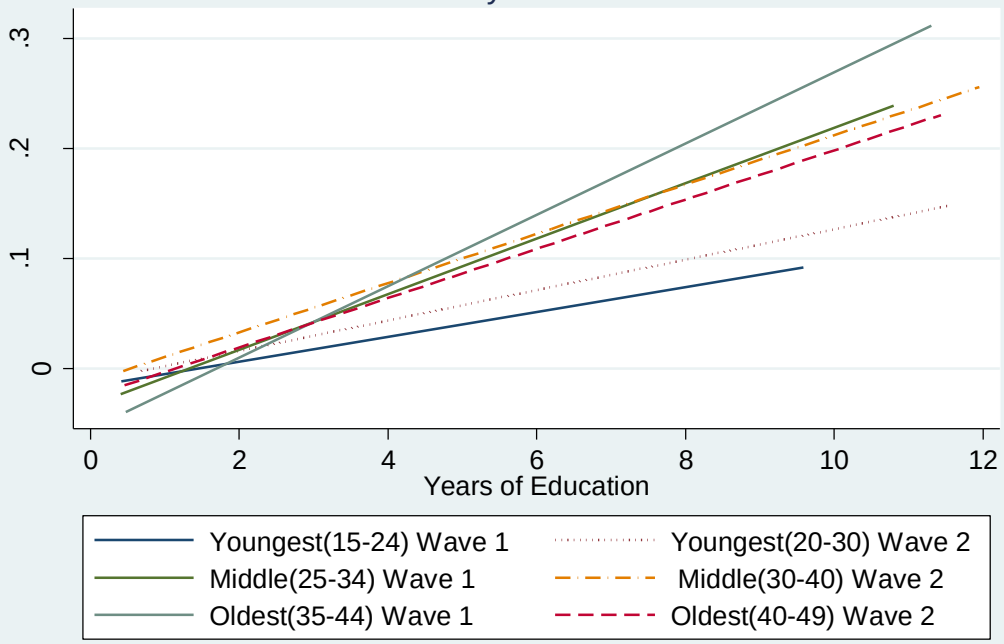


Figure 1.2b: Relationship between HIV Prevalence and Primary Education by Cohort in Both Waves

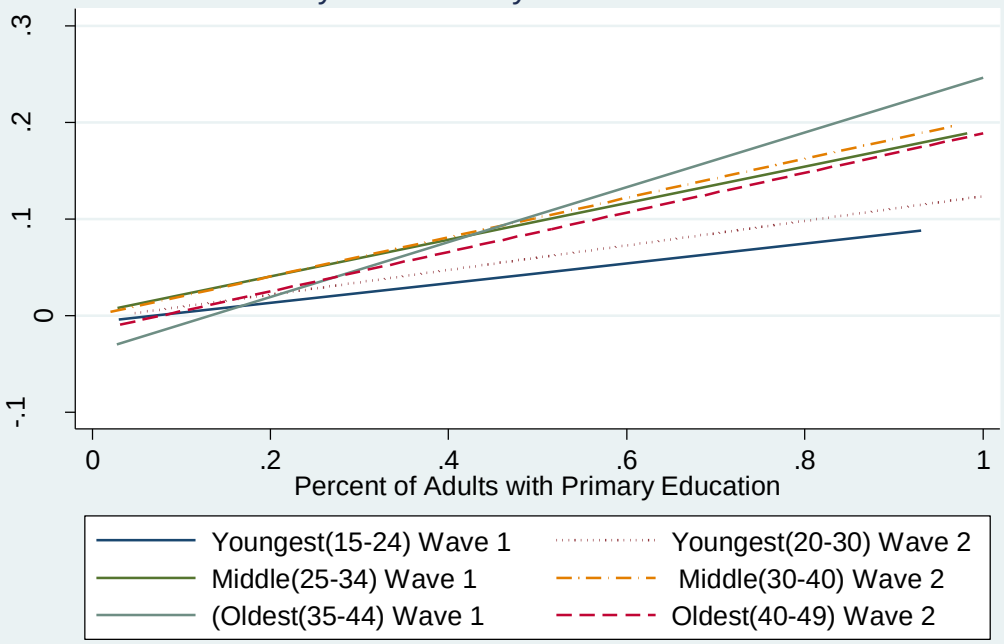


Figure 1.3a: Regional Changes in Children's School Attendance Rate by Adult HIV Prevalence

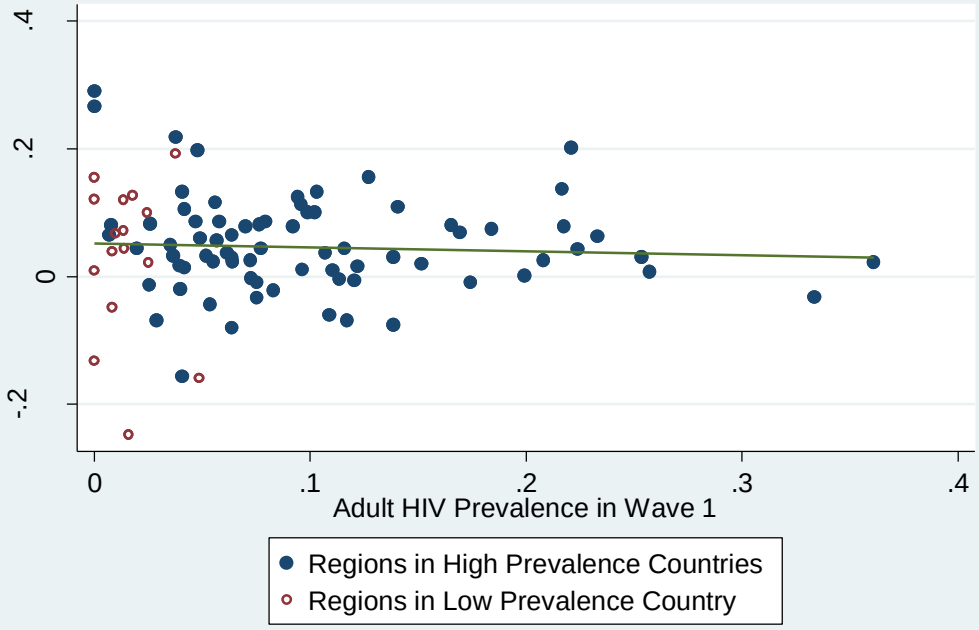


Figure 1.3b: Regional Change in Children's Years of Completed Education by Adult HIV Prevalence

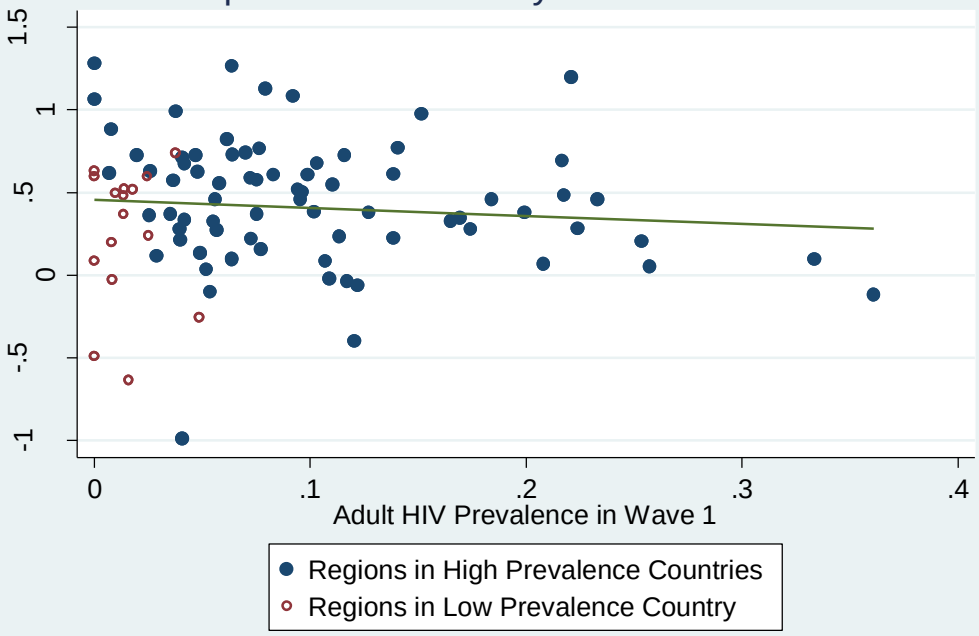


Figure 1.4a: Years of Completed Education and Change in HIV Prevalence

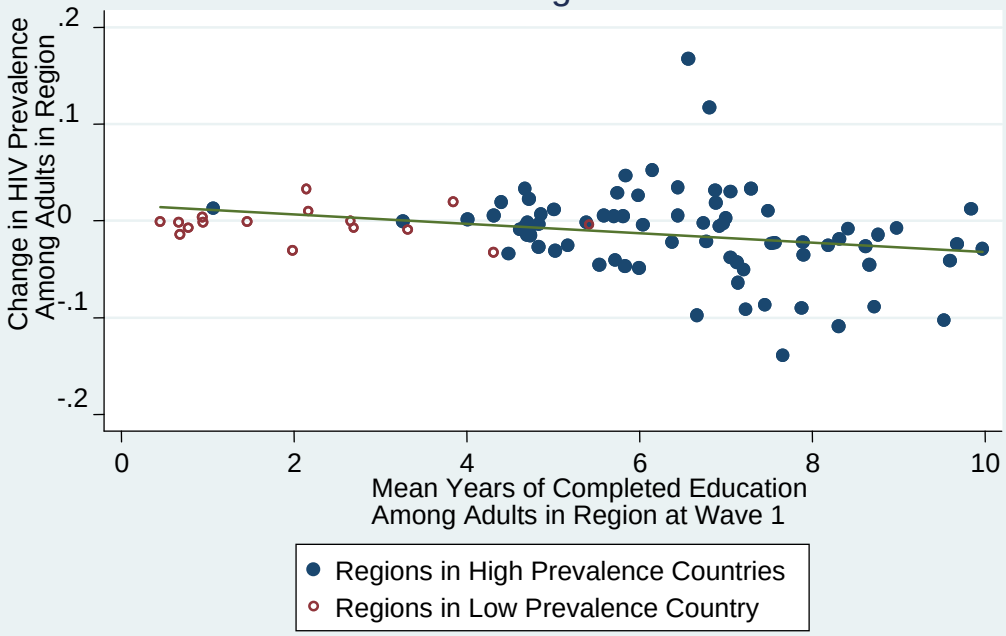
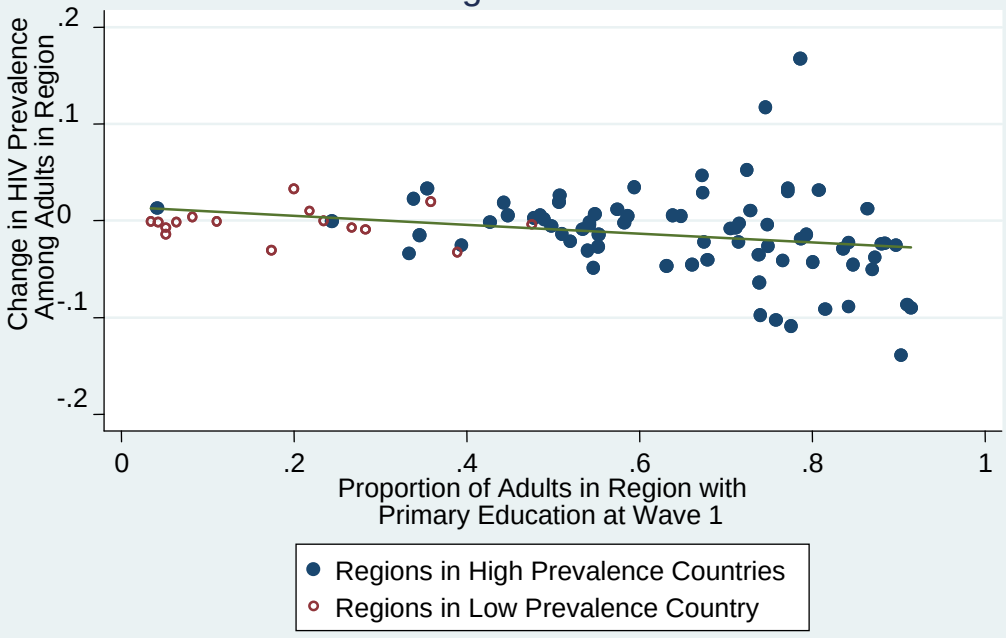


Figure 1.4b: Primary Education Rate and Change in HIV Prevalence



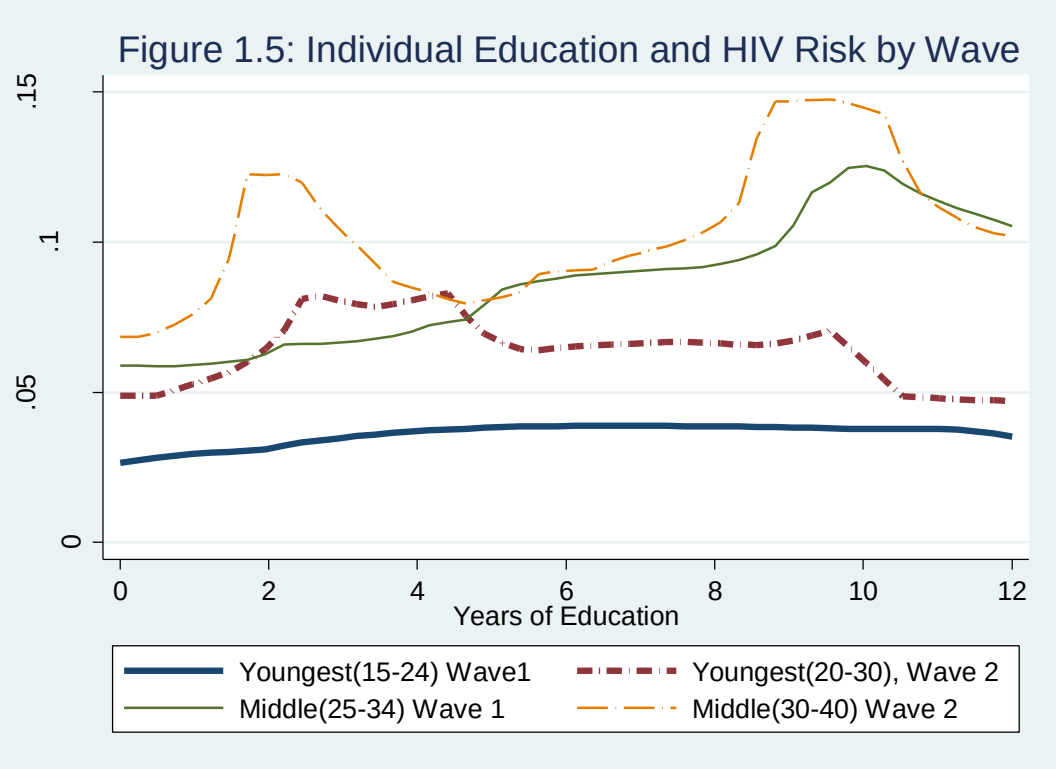
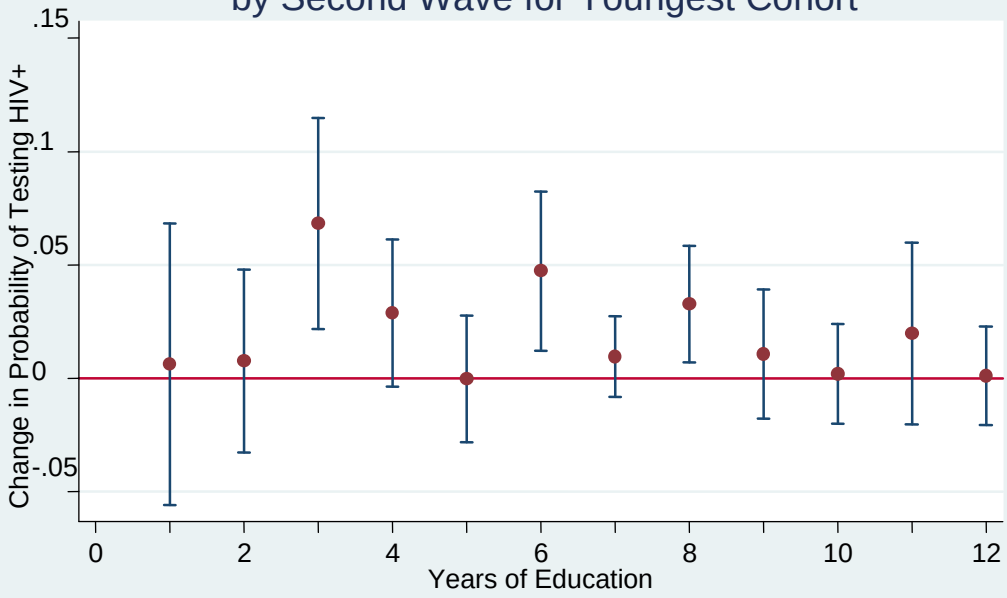
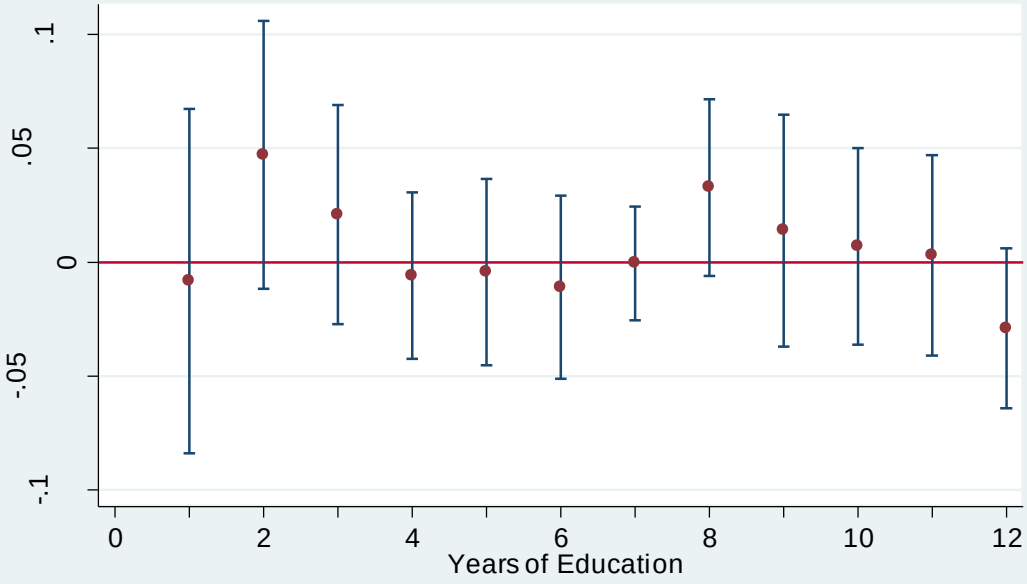

Figure 1.5: Individual Education and HIV Risk by Wave
.15
.1
.05
0
0
2
4
6
8
10
12
Years of Education
Youngest(15-24) Wave1
Youngest(20-30), Wave 2
Middle(25-34) Wave 1
Middle(30-40) Wave 2

Figure 1.6a: Change in Probability of testing HIV+ by Second Wave for Youngest Cohort



based on adults 15-24 in first wave, controlling for age, gender urban residence and region fixed effects

Figure 1.6b: Change in Probability of testing HIV+ by Second Wave for Middle Cohort



based on adults 25-34 in first wave, controlling for age, gender urban residence and region fixed effects

Chapter 2: Education and HIV Risk Behaviors

Abstract:

I use DHS data from four sub-Saharan African countries to examine whether there is evidence that the behavioral response to the HIV epidemic is stronger among the more highly educated. I find that there is a robust positive association between education and condom use and HIV testing, and age at marriage, with evidence that younger cohorts may be reducing age at marriage. I also find evidence that more educated individuals are increasing their rates of HIV testing and reducing age of marriage more than the less educated. These findings are robust to controlling for HIV status in the two countries where HIV tests were linkable to individual data

I. Introduction

The most recent UNAIDS annual report takes an optimistic tone that has, to date, been relatively rare in the literature on HIV/AIDS. The report opens with the observation that "on the cusp of the fourth decade of the AIDS epidemic, the world has turned the corner-it has halted and begun to reverse the spread of HIV" (UNAIDS 2010). In Africa, in particular, many of the most hard-hit countries have finally begun to see reductions in incidence. While the halt of the epidemic's growth is worthy of celebration, some scholars have predicted that the changes in the forward march of the epidemic are likely to be accompanied by a shift in who bears the brunt of new infection (Glick and Sahn 2007b; Hargreaves et al. 2008). Although HIV was initially associated with higher socioeconomic status in much of Africa (Hargreaves and Glynn 2002; Fortson 2008), it could ultimately become associated with poverty if the wealthier and more educated get better information, change behaviors more dramatically, or have better access to health care resources. In other words, as information and resources for prevention and treatment become widely available, they may likely accrue to the more privileged first, thus altering the shape of the epidemic.

A few recent papers have suggested that such a change is already underway. In a recent meta-analysis of research conducted in the late 1990s, Hargreaves et al (2008) find that, in contrast to earlier findings, HIV prevalence fell more consistently among more educated groups than less educated groups. Similarly, in an analysis of longitudinal DHS HIV prevalence data, I find that areas with higher average education see larger drops in HIV prevalence on average, and that younger people with more education are at lower

risk of contracting HIV over time than those with less than a primary education (Gummerson, 2011). These observed changes in the relationship between education and HIV risk are particularly interesting because prior evidence on risk behavior by education level suggested that wealthier and more highly educated people did not necessarily have safer sexual practices (De Walque 2009; Fortson 2008; Glick and Sahn 2007a).

This analysis examines whether the previously documented relationships between HIV risk behaviors and education level have changed in a way that is consistent with the observed shifts in HIV prevalence. Specifically, I use DHS data from four sub-Saharan African countries to measure the association between HIV related risk behaviors and education levels. I then ask whether more educated people appear to be changing their behaviors more over time than comparatively less educated people within the same region. I examine four specific behaviors, age of marriage (or entry in stable partnership), condom use at last sex, HIV testing, and number of partners in the last year. I find that protective behaviors have largely increased over time for both men and women. I also find that more educated people are more likely to have used a condom at last sex, are more likely to have had an HIV test, and that there is no longer a clear relationship between education level and number of partners. Education is associated with later entry into marriage, but this association is considerably weaker for the youngest cohort.

Finally, I find that, not only is education associated with protective behaviors, more educated people appear to be adopting some protective behaviors at a higher rate than less educated. Among both men and women, those with more years of education

increase their HIV testing more over time. More educated women also appear to be entering marriage earlier than their predecessors and more educated men report greater decreases in the number of sexual partners over time. In a subsample of HIV tested respondents, these findings are largely robust to the inclusion of HIV status. Together, these findings are consistent with the notion that reductions in HIV incidence can be at least partly attributed to behavior change. Furthermore, the more educated are changing their behaviors more, potentially signaling a shift in the future burden of the disease towards more marginalized sectors of society.

The next section reviews the literature and theory on education on HIV risk behaviors. Section 3 introduces the data and measures. Section 4 tests first whether protective behaviors have generally increased over time in the sample countries and then examines if the prior documented associations between risk behaviors and education levels are found in this data. Finally, it tests whether behavior change between waves is stronger for people with more education. Section 5 concludes.

II. Education and HIV Risk Behaviors

Early in the epidemic, scholars noted that in sub-Saharan Africa the relationship between HIV and socioeconomic status differed from many infectious diseases in that it was more strongly associated with wealth than poverty (UNAIDS 1998; Gregson, Waddell, and Chandiwana 2001). Countries with higher literacy rates were also those with the most severe epidemics, and surveys demonstrated that markers of socioeconomic advantage such as education level, formal sector employment and urban

residence were all positively associated with HIV infection (Barongo et al. 1992; Fylkesnes et al. 1997; Boerma et al. 1999). This was attributed to a variety of structural and behavioral factors. Places with higher levels of adult literacy also had undergone more urbanization and higher levels of migration, both of which facilitated the spread of HIV (J. Caldwell and P. Caldwell 1993). Furthermore, African women with more education were shown to enter marriage later, have more sexual partners before marriage and favor pill-based contraceptive methods, all of which increased the risk of contracting HIV (Blanc and Way 1998; Bongaarts 2007).

At the same time, some hypothesized that as the epidemic matured, the burden of new infections might shift down the socioeconomic gradient (Over and Piot 1993; Gregson, Waddell, and Chandiwana 2001; Fortson 2008). These predictions were based on economic and sociological theory suggesting that more educated individuals could be more motivated to adopt protective behaviors. Oster (2009) argues that people with more wealth and longer life expectancy (setting aside mortality risk from HIV) may be more motivated to change their behavior because the lifetime costs of an HIV infection are higher for them both in terms of life years lost and forgone income. In addition, people with more education may have better access to protective information. For instance, if HIV prevention campaigns are school based or rely largely on written materials, more educated and literate people are at an advantage. Wealthier and more educated people also likely have more exposure to formal health care and may receive more prevention messages through media like television and radio (Glick and Sahn 2007b). Finally, some scholars suggest that more educated people not only receive more prevention

information, but are better able to employ that information to protect themselves (De Walque 2007; Glick and Sahn 2007b).

While some of these proposed mechanisms linking socioeconomic status and behavior work through education, others are more direct consequences of wealth. Because the two are highly correlated, disentangling the effect of being wealthy from that of being highly educated is problematic. Further, people who accumulate wealth or go far in school may have other inherent or special characteristics that could explain their success in school and the marketplace as well as their reactions to HIV risk. By and large, the research focuses on education as a proxy for socioeconomic advantage as a whole.[11] This analysis follows in that vein, concentrating on educational attainment, while acknowledging that it is likely related to other characteristics which are salient in determining HIV related risk behaviors.

While the theory suggests that the more educated should adopt more protective behaviors, the empirical evidence for such a change is mixed. Several recent reviews of Demographic and Health Surveys (DHS) from sub-Saharan Africa find that having more years of completed education is associated with increased condom use and increased probability of HIV testing, but also with having more sexual partners, a higher probability of premarital sex and lower levels of abstinence (Fortson 2008; de Walque 2009; Glick and Sahn 2007a). However, given that more recent DHS surveys demonstrate a strong association between regional education levels and decreasing prevalence (Gummerson 2011), it seems opportune to re-visit this question with the most

[11] One notable exception is Fortson (2008) which uses DHS data to analyze the effect of education and a wealth index separately and finds education is a more consistent predictor of behavior and infection rates.

recent data and examine whether the behaviors of the more educated have begun to change differentially, as predicted.

III. Data and Measures

The data for this analysis is drawn from the Demographic and Health Surveys (DHS). The DHS are large nationally representative household surveys that cover household members' education, nutrition, fertility, and health. More recently, some African DHS have asked a subsample of adult respondents to provide blood samples for HIV testing[12]. The four countries examined here - Kenya, Mali, Tanzania and Zambia- are, to date, the only ones for which two rounds of HIV testing data are available. I use two waves of cross sectional data drawn largely from the men and women's questionnaires and the HIV questionnaire. For two of the countries, Tanzania (2003 and 2008) and Mali (2001 and 2006), the data were collected approximately five years apart. In Kenya (2003 and 2009) and Zambia (2001 and 2007) the interval is six years. I base my analysis on 104,538 adults ages 15-49. Table 2.1 includes sample means for the four countries covering basic demographics as well as HIV prevalence and education rates. It is of note that the majority of the sample is female. While all eligible women were asked the women's questionnaire, a subsample of 1/2 or 1/3 of the households were chosen for the men's questionnaire. Tanzania and Zambia (2007) are the exceptions in that all men

[12]Response rates were high and analysis on a subset of countries has shown that bias from non-response is small (Mishra et al. 2006).

and women were administered the individual questionnaires.[13] I report behavioral results for men and women separately. Similarly, HIV tests were administered in a representative subsample of households rather than all households, with the exception of Tanzania- where all respondents were offered an HIV test.

The analysis here rests upon measures of HIV risk related behaviors, educational attainment and HIV prevalence. The educational and HIV means reported in Table 2.1 illustrate that the four countries fall into two types. Kenya, Tanzania and Zambia are all relatively high prevalence countries in East Africa which also have high levels of adult education. More than half of the adults in these three countries have at least a primary education and HIV prevalence, although falling, remains above 5 percent. Mali, the lone West African country, has a different profile. It is a comparatively low prevalence country (under 2 percent) with much lower levels of adult education and higher rates of marriage.

I include two measures of educational attainment in 2.1. The first measure is years of completed education. However, because some of the respondents are still quite young and could still be completing education, I also include primary school completion rate. To ensure validity, I conducted this analysis using 'completion of primary schooling' rather than 'years of completed schooling' as the measure educational attainment. Results were robust to the choice of measure, and only the 'years of schooling' results are presented for reasons of brevity. In addition to the characteristics listed in 2.1, I also make use of the DHS defined sub-national regions within each country to compare behaviors over time within distinct regions.

[13] The Tanzanian survey administered in 2003 was the AIDS Indicator Survey (AIS), which is not the standard DHS, but for the subjects covered in this analysis they are equivalent, and I do not distinguish between them.

My primary outcome measures are four self-reported risk behaviors- condom use, number of partners, HIV testing, and age at first marriage. I discuss each in more detail below:

Condom Use- The measure of condom use here is whether the respondent used a condom with their partner the last time they had sexual intercourse. Measured in this way, rates of condom use are remarkably low in the sample. In part, this is likely because a large percentage of the sexually active respondents (82%) are in married or stable relationships and may consider themselves to be in low-risk relationships and/or may be eschewing condoms for reasons of desired fertility. For this reason, I report the results on condom use on two separate groups within the sample. The first is respondents who report themselves to be sexually active during the last 12 months and *not* in a stable relationship. If there is increased regular condom use, I might expect to find it more among this group (who may perceive themselves at more risk). I also report the results for condom use among sexually active respondents who report themselves in a married relationship.

Number of Partners- I evaluate the number of partners that the respondent reports in the last 12 months (among those who report sexual activity). A higher number of partners is associated with higher risk of HIV infection. More than any of the other outcome variables, reporting on partnership is highly divergent by gender. Men report more partners on average than women, both in and out of marriage. Approximately 22 %

of sexually active men report more than one partner in the last year, whereas only about 3% of sexually active women do.[14,15]

HIV Testing- I also examine whether the respondent reports ever having been tested for HIV test prior to the interview. This measure is potentially ambiguous to interpret as a risk-related behavior. Having had an HIV test could indicate that respondents consider themselves at higher risk because they know they have engaged in risky behavior. Alternately, it could signal that a respondent is particularly risk-averse or that the individual has simply had exposure to a health care system where HIV tests are becoming more and more routine. I follow prior literature in interpreting testing as a protective behavior, since individuals who know their status are better able to both seek treatment and alter risk behaviors.

Age at Marriage- Finally, I examine age at first marriage. Both the male and female individual questionnaires ask about age at first marriage for all respondents who report ever having married. Definitions of marriage may differ across place and culture and the questions on marriage in the DHS are worded to include people who are 'living together as if married'. As such, marriage is defined largely by the respondent, and I interpret the respondents 'age at first marriage' as indicating first entry into a stable long-term union. Given the active practice of polygamy in many parts of Africa, I do not

[14] While the results reported here treat number of partners as a linear variable, the findings are robust to specifying it as a categorical variable, dividing respondents into categories of one partner, two partners, and three or more partners.

[15] Several analysis have found evidence that suggests that women tend to under-report the number of sexual partners compared to men- see Gersovitz et al (1998) deWalque (2006)and Glick and Sahn (2007)- However, the degree of discrepancy appears to be highly consistent across multiple countries and waves of the DHS. Nevertheless, I analyze men and women's reports separately, in part to account for this.

assume marriage to mean monogamous union, although the majority of respondents who report themselves married in this data also report their unions to be monogamous.

In the context of HIV risk, the effect of early marriage for women is debated. Clark (2004; 2006) finds that girls who enter marriage young are at higher risk of HIV. In contrast, Bongaarts (2007) finds that later marriage is associated with an increase the number of premarital partners, which can increase exposure to HIV risk. While the empirical evidence for the effect of marriage on HIV remains mixed, there is little question that entry into stable monogamous partnership is widely *perceived* as a protective behavior by the populations examined here. In a review of data from Africa, Gersovitz (2005) finds that entry into long term monogamous relationships is most the commonly cited strategy for avoiding HIV infection, and Reniers (2008) demonstrates that Africans use marital strategies as a means of managing HIV risk. As such, I treat earlier entry into union as evidence of protective behavior change.

Finally, it is worth noting that although the DHS questionnaires ask directly about age of sexual debut, extramarital partnerships and concurrency, I do not make use of those measures of risk behavior. I omit them based on evidence from prior reviews that the data on these sensitive questions suffers from a high degree of social desirability bias in reporting and that the bias is not consistent across ages or waves of the surveys. (Zaba 2004; Gersovitz 2005; Glick and Sahn 2007a).

IV. Results

Are protective behaviors increasing overall?

Tables 2.2 and 2.3 show the means of the four behaviors by country and wave for women and men, respectively. In general, the figures show increasing protective behaviors across both genders. Condom use goes up overall, although in all but two cases, less than half of respondents report using condoms at last sex. Condom use at last sex is, predictably, much higher among single and sexually active people than married people. Mali, the lone low HIV prevalence country, is the exception in that condom use is much lower in all categories and does not increase for single men or women between waves. Although the increases are larger for single respondents, condom use largely goes up between waves among married people as well, with greater increases for men than women. Further, in all the countries, men are more likely to report having used a condom at last sex than women, regardless of marital status. The male respondents also report much more extramarital sex than women, (results not shown) so this pattern may reflect men differentiating and using condoms in their extramarital or casual sexual relations but not within marriage or more serious dating relationships, as has been suggested by some scholars (De Walque and Klein 2011). In addition to increasing condom use, respondents also report decreases in the number of sexual partners. The pattern here is analogous to condom use, in that behavior change appears greater among men (who report far more partners to begin with) and is confined to the high prevalence countries.

As noted above, these measures include only individuals who report any sexual activity, and therefore they exclude the possibility that people are choosing abstinence at a higher rate. I found that women in all four countries reported constant abstinence rates

across waves. Rates of abstinence among women in Mali were lower (16%) than they were in the high prevalence countries (23-30%). For men, the abstinence rates are similar across the four countries (26-30%) and were constant everywhere but Mali. In Mali, the percentage of men who report zero sexual partners in the last year rises from 25 percent to 31 percent between waves.

By far the most dramatic changes in Tables 2.2 and 2.3 are in the proportion of respondents who report having been tested for HIV. In the three high prevalence countries it goes up by ten percentage points or more for men and by between 27 and 44 percentage points for women. This rapid rise over a short period may reflect an increase in supply of HIV testing in high prevalence areas- either through expansion of voluntary testing and counseling services and/or inclusion of opt-out testing as part of routine medical care. In an analysis of earlier waves of the DHS, Glick and Sahn (2007b) uncovered evidence signifying a large unmet demand for HIV testing, so the rapid rises in testing may reflect a change in supply as much as a change in demand for testing. Regardless, the comparatively large increases are the most striking change in behavior over the 5/6 years between waves. Mali, again, is an exception with a small rise in testing for women and none for men.

I would expect the last behavioral measure, age of marriage, to show much less change over the period covered by these data. While the other three measures can vary over time among the whole sample, only those who had not already married can potentially contribute to changes in this measure. As such, for the majority of the sample,

this is unlikely to vary between waves.[16] Indeed, the mean comparison shows only small change across the two waves. However, the means of age at first marriage are calculated among those who had ever married. If there were significant increases in the fraction of respondents who never married at all, even small changes in mean age at first marriage could be misleading. Fortunately, the fraction of people who ever marry does not change between waves in any of the countries for either gender. I do find small increases in the age of first marriage for women in Tanzania, Mali and Zambia. Men marry at older ages than women in all the countries and there is a slight increase in age at marriage for men in Tanzania and a slight decrease for men in Mali.

Altogether, there are hopeful trends towards increased protective behavior among both men and women. Condom use appears to be increasing, particularly among the single respondents, people are reporting fewer partners, and HIV testing has gone up considerably.

Has education become unambiguously associated with more protective behaviors?

As a first step towards evaluating the possibility of differential behavior change by education, I examine whether the relationships between education and risk behaviors in these data align with the conclusions of previous work. Figures 2.1-2.4 plot the mean risk behaviors at each year of education, with the size of each circle corresponding to the number of respondents at each education level for all 8 DHS survey rounds. In addition to the point estimates, which are for all ages in the sample, Figures 2.1-2.4 also include the same estimates calculated separately for three age groups. For clarity of presentation,

[16] An analysis (not presented) confirms that the changes in age at marriage are concentrated among those under 30.

these are presented as smoothed lines overlaid on the point estimates.[17] The relationships here largely confirm the previous findings on condom use and HIV testing. There is a strong positive gradient in HIV testing, with more educated being more likely to have ever had a test. The youngest age cohort has lower levels of HIV testing across education levels (Figure 2.1) likely reflecting an age effect. In addition, more educated people are far more likely to have used a condom at last sex in all age groups, although overall levels of use are higher among the younger age cohorts. This association is particularly strong for single people, although it is also detectable among the married (Figures 2.2a and 2.2b).

Age at marriage, similarly, confirms previous work with more educated people marrying at later ages (Figure 2.3). However, the youngest cohort has a weaker association between education and age at marriage than the older cohorts. Because a significant number of the younger cohort have not yet married, its plausible that those who are delaying marriage (and therefore not in the sample of 'ever married') are also those who are pursuing more education. As these people get married at later ages, the relationship between education and age at marriage could shift to resemble that of the older two cohorts. As a test of this theory, I construct a synthetic cohort of individuals who are 15-25 in the first wave and 20-31 in the second wave.[18] I then estimate the relationship for the cohort in the first wave and again in the second to see if it has changed as they aged. I find no strengthening of the relationship between age at marriage

[17] The figures here combine men and women. When the figures are drawn separately by gender, the 'age at marriage', 'condom use' and 'number of partners' mean values are lower for women, but the relationship between the measures and education is similar for both genders. HIV testing is actually slightly higher for women at all education levels than it is for men, possibly due to antenatal testing, but again, the relationship between testing and education is similar for men and women.

[18] For the two countries with 5 years between waves, the cohort is 15-25 in the first wave and 20-30 in the second. For the two with 6 years between waves, the cohort is 21-31 in the second wave.

and education between the earlier and later wave, indicating that the observed difference probably isn't fully attributable to age effect (results not shown).

The final measure, number of partners in the last 12 months, deviates somewhat from the other three outcomes in that it does not exhibit an obvious relationship with years of education. Although people who report zero years of education clearly report fewer partners, the relationship is highly variable for people reporting one year of education or more. When the estimates are plotted separately for men and women, the overall levels of reported partners are much lower for women, but the relationship with years of education is similarly inconsistent for both genders. (figure not shown) This contrasts with the earlier works cited which found that more educated men and women both tend to have more partners. The ambiguousness in this more recent data may potentially reflect changes underway in the partnership patterns of the educated.

Collectively, these graphs suggest that the protective behaviors that were previously positively associated with education -condom use and HIV testing- still are. Of the two relatively risky behaviors associated with education- later marriage and having more partners- there is evidence of a potential change in the number of partners, but age at marriage is still quite strongly positively associated with education levels. However, the younger cohort displays a much weaker relationship between education and age at marriage, which is consistent with highly educated young people entering stable unions earlier than their predecessors did.

Are the more highly educated changing their behaviors more?

The evidence that education is positively associated with most protective behaviors and that protective behavior appears to be increasing in high prevalence areas is promising, in that it suggests constructive behavior change. It does not, however, fully explain why the relationship between education and HIV risk is changing at the regional level. In order to explain the observed changes between HIV risk and education, the educated would not only have to increase their protective behaviors, they would have to do so at a relatively higher rate than those with less education.[19]

Tables 2.4 and 2.5 begin to examine this question by listing the changes in means of each of the four behaviors by total years of completed education. By reading the columns downwards, it is possible to see whether the mean values are changing differentially by education level. In other words, do people at the higher end of the educational scale change their behaviors more than those at the lower end of the scale? For example, among women with zero years of education, the mean number who report having used a condom at last sex goes up 4 percentage points between waves. Similarly, for women with 6 years of completed education, the mean percent who used a condom at last sex goes up by 12 percentage points between wave one and two.

For women (Table 2.4) the evidence on condom use looks mixed. Although we know more educated women are more likely to have used a condom at last sex (Figure 2.2a & 2.2b), the increases between waves do not appear to have a clearly increasing pattern as years of education increase. Similarly, the number of partners women report

[19] An alternate explanation is selective mortality by education level, but evidence from DHS does not support that interpretation. See Gummerson (2011) and Case and Paxson (2011)

appears to be largely going down across education levels, but its not clear that it goes down more for the more highly educated.

In contrast, testing and marital behavior do have clear gradients. The change in percent that have ever had an HIV test gets steadily larger as you read down the column, indicating that while all women are getting more tests, the increases are more dramatic among those with more years of completed education. Interestingly, age at first marriage also has a detectible pattern. As seen in Figure 2.3, the association between education and age at first marriage is positive overall- less educated women marry much earlier. However, the changes in means suggest that more educated women are decreasing their age at marriage, while those at the lower end of the educational spectrum appear to be increasing theirs. As noted earlier, both very early marriage and late marriage have been associated in the literature with higher HIV risk. It may be that both lower and higher educated women are adjusting marriage behavior towards some perceived ideal age of marriage.

Table 2.5 reports the same results for the male sample. Similarly to women, men appear to be increasing rates of HIV testing differentially by education. Men with more years of education have considerably larger changes in the percentage that have tested between waves. This is striking because although increases in testing among women could be partly explained by the expansion of antenatal testing, the increases among men are not as easily attributable to any analogous routine medical care. In addition, men's age at first marriage appears to be decreasing somewhat at both high and low education levels, and coming down more among the more highly educated. Noting that men marry at older ages than women on average (Tables 2.2/2.3), it may be that reducing the age of

entry in unions is a more unequivocally protective for males. The remaining columns suggest that condom use is going up and the number of partners is going down across education levels, but that these changes are not larger for men with more education.

Although the patterns in Tables 2.4 and 2.5 suggest that more educated people are adopting some protective behaviors more, it is possible that the pattern is due simply to regional variations in policy or resources. Wealthier urban regions, which initially had both higher average education and higher rates of HIV, may also be the sites of more aggressive educational and testing campaigns. If this were the case, the associations observed between protective behaviors and education would have less to do with individual education and more to do with regional environments. In order to examine this possibility, I estimate regressions of education on HIV using country by region fixed effects to test if individuals living in the same region differ in their behaviors according to education level. Specifically, I run the model below for each of the four measures.[20]

$$(1) \quad Y_{iw} = \theta + \beta_0 Ed_{iw} + \beta_1 Ed_{iw}^2 + \beta_2 Wave2_{iw} + \beta_3 Ed_{iw} * Wave2_{iw} + \beta_4 X_{iw} + \varepsilon_{iw}$$

where *Y* is the behavioral outcome measure [age at marriage, number of partners, HIV testing, or condom use] for individual *i*, at wave *w*. *Ed* and Ed^2 are measure of the years of completed education and its square as reported by the individual; *Wave2* is a indicator of whether the observation comes from the later wave; *X* is a vector of individual controls including age, age squared and age cubed, an indicator for urban residence, marital status

[20] The results reported here are from ordinary least squares regressions, for the two binary outcomes – condom use, and HIV testing- I also ran the models as probit. In addition, for the outcome 'number of partners' I ran an ordinal probit using the categorical version of the variable. In all cases, the results were robust to choice of model and variable form.

and full set of indicators for country by region. Finally, *Ed*Wave2* is an interaction term between the years of education and *Wave2*. The coefficient on the interaction term *Ed*Wave2* is the key parameter as it reflects whether the association between years of education and the outcome behavior changes significantly between waves. The results for all four outcomes are reported separately by gender in Tables 2.6 and 2.7.

The coefficients in Tables 2.6 and 2.7 on both the education and wave variables confirm the patterns observed in the previous tables and figures. Even when country by region fixed effects are included, people with more education report more use of condoms, more testing, and higher ages of marriage relative to people with less education. However, the education variables are not jointly significantly predictive of the number of partners in the last year for either gender. The coefficients on the wave variable also confirm that protective behaviors are largely increasing over time for both genders.

More noteworthy are the coefficients on the interaction term, which indicate that, even within the same regions, people with more years of education increased their testing at a higher rate. For women, the effect of each additional year of education on the probability of having had an HIV test is 2.4 percentage points greater, on average, in the second wave. For men, the coefficient is smaller (0.014) but it is also statistically significant (p-value<0.001). For men, the coefficient on the interaction term is also negative for the number of partners, suggesting that more educated men are reducing their number of partnerships more between waves. When age of marriage is the outcome, the coefficient on the interaction term is negative for both genders, but it is statistically significant only for women. Although the general trend for women is an increase in the

average age of marriage, it appears that for more educated women the increase is smaller. This is consistent with the theory that young and highly educated women may not be delaying marriage as much as in previous data.

Thus far, the evidence confirms that more educated people both have more protective behaviors overall and are altering at least some behaviors at a higher rate. This would seem to support the theoretical arguments that forecasted people with higher socioeconomic status would change behaviors more in response to the availability of HIV prevention knowledge and resources. One potential challenge to this interpretation of the data would be that it does not reflect people of high SES changing behaviors, but rather people with HIV changing their behaviors as a direct response to infection. Because more educated people were more likely to be infected early on, large scale behavior change by people who are infected could appear to be behavior change by education level. In order to test whether the trends observed hold when HIV status is accounted for, I link HIV tests results to the individual records and test the following model:

$$(2) \quad Y_{iw} = \theta + \beta_0 Ed_{iw} + \beta_1 Ed_{iw}^2 + \beta_2 Wave2 + \beta_3 Ed * Wave2_{iw} + \beta_4 HIV_{iw} + \beta_5 X_{iw} + \varepsilon_{iw}$$

where *HIV* is an indicator equal to one if individual *i* tested positive for HIV in wave *w*. All other terms and controls are equivalent to model (1) above. Unfortunately, because HIV tests were not linkable to individuals in the first wave in Mali and Zambia, I am only able to test this model in both waves in Tanzania and Kenya.[21] In the reduced sample I

[21] In an additional check of internal validity, I ran the preceding analyses using only the smaller HIV test-linked sample as well. The results were not substantively different.

include 37,410 adults aged 15-49 who were tested for HIV.[22] Tables 2.8 and 2.9 report the results of model (2) on each indicator. HIV positive women are more likely to have used a condom at last sex both in and out of marriage and married HIV positive men are more likely to have used a condom, but single men are not.

More salient for this analysis is whether the inclusion of this variable changes the significance of the coefficients on education variables or the interaction between education and wave. In the case of the combined education variables, including a control for HIV status does not change the basic findings from the previous analysis. Education levels are positively correlated with most protective behaviors, with the exception of the number of partners for men. In addition, the coefficients on the *Ed*Wave2* interaction term largely remain statistically significant and substantively similar to model (1). Among women, the association between having more education and ever having had an HIV test is strengthening over time and more educated women have lesser changes in age at marriage between waves. For men, the more educated also increase their rates of testing more between waves.[23] The interaction coefficient for age at marriage for men is negative and statistically significant at the 10% level.

The one outcome variable that does not appear to hold to the patterns observed in model one is the number of partners reported. When HIV status is included in the model, it no longer appears that more educated men are reducing the number of partnerships more than less educated men. In addition, the coefficient on the interaction term for

[22] As noted above, all eligible adults were tested in Tanzania, whereas in Kenya, a representative subsample was tested.

[23] Although the respondents in this sample were tested for HIV as part of their participation in the survey, the survey question about whether they were ever tested for HIV specifically refers to testing *before* their participation in the interview. Therefore, respondents' answers about having had HIV tests should not be influenced by the tests they received as part of their participation in the survey.

women becomes slightly positive, suggesting that more educated women are changing behavior differentially, but are actually moving towards higher numbers of partners compared to less educated women. However, given the high variability in the relationship between education and number of partners and the fact that very few women report more than one partner, this should be interpreted with caution.

In summary, although I am only able to perform this check on a subsample of the data, the evidence from the subsample suggests that the behavior changes observed are robust to the inclusion of HIV status. For the models of HIV testing and age at marriage, both the finding that individuals are changing their behaviors and that more educated are changing their behaviors more stand. The model for number of partners does change with the inclusion of HIV status- although I interpret this result with care. Overall, I construe this as evidence that it is unlikely that the observed behavior changes are solely driven by HIV positive individuals.

V. Conclusion

The results of this analysis are largely supportive of prior theoretical predictions that people at the higher end of the socioeconomic spectrum would be likely to change their behaviors first as the HIV epidemic in Africa matured. While earlier work found that more educated people demonstrated a mixture of more protective and more risky behaviors, the associations in this more recent data suggest that where education was associated with more protective behavior (condom use, HIV testing) it is only becoming more so. Where education was associated with higher risk behavior (multiple partnerships and delayed marriage) there is evidence of change. Education is not

associated with multiple partnerships in the most recent data, and there is some evidence that younger people with more education are not delaying marriage as long as was seen in previous generations. Further, the educated appear to be changing their testing and marital behavior at a higher rate, potentially increasing the behavioral gradient, and, consequently, the health gradient.

It bears emphasizing that the behavior changes described here are by no means universal. Even in this sample of four countries, the changes are concentrated in the high prevalence countries. Subsequent work should focus on identifying causal pathways if policy makers hope to spur behavior change more widely. Hopefully, future research will shed light on the sticky question of what specifically about the more educated makes them more likely to change behaviors. If it is simply access to information, then the extension of protective behaviors will depend largely on developing innovative means of spreading prevention and treatment information beyond the literate and schooled sectors of society. If education actually enhances one's ability or propensity to adopt protective behaviors, then expanding educational access is part of a robust anti-HIV agenda. If the motivation and ability to change behaviors comes primarily from factors associated with education, such as wealth, personal characteristics or increased access to formal health care, then policy makers likely need to concern themselves more broadly with sources of societal inequity.

Even in the absence of full understanding of causality, the patterns described in this analysis provide significant information. The associations uncovered describe a change in the risk profile which has been seen in the past with other diseases. As

information and resources for prevention and treatment become available, they accrue first to the more privileged in society. In the case of HIV, where the wealthier and more educated originally had higher rates of infection, this pattern may be particularly efficient at reducing future prevalence. And although decreased risk among the wealthy does not necessarily predict an increase in absolute risk among the poor, it does suggest that the next frontier for HIV education and treatment in Africa may be among the more marginalized sectors of society.

Table 2.1: Sample Means By Country and Wave

	Kenya		Mali		Tanzania		Zambia	
	(2003)	(2009)	(2001)	(2006)	(2003)	(2008)	(2001)	(2007)
Proportion Female	0.71	0.72	0.81	0.80	0.55	0.56	0.80	0.54
Mean Age	27.88	28.41	28.84	28.53	28.01	28.42	27.66	28.10
Mean Years of Completed Education	7.28	8.02	1.49	1.78	5.76	5.75	6.19	6.93
Proportion With Primary Education	0.55	0.63	0.12	0.14	0.66	0.64	0.53	0.60
Proportion Rural	0.76	0.75	0.70	0.66	0.70	0.75	0.60	0.58
Proportion Married	0.65	0.64	0.82	0.83	0.68	0.69	0.72	0.67
HIV Prevalence	0.076	0.063	0.015	0.012	0.081	0.069	0.173	0.154
Number of Observations	11,548	11,707	15,837	18,278	12,522	11,858	9,639	13,149

Based on all adults ages 15-49. Means calculated using DHS provided sample weights. In all countries except Tanzania, HIV prevalence is calculated based on representative subsample and DHS provided HIV sample weights. Tanzania tested all respondents rather than a subsample but prevalence is also calculated using DHS provided HIV sample weights.

Table 2.2: Female Risk Behaviors by Wave and Country

	Tanzania		Kenya		Mali		Zambia	
	Wave1	**Wave 2**	**Wave1**	**Wave 2**	**Wave1**	**Wave 2**	**Wave1**	**Wave 2**
Condom Use-Single	0.34	0.39	0.23	0.34	0.15	0.14	0.29	0.35
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
	1,096	*1,039*	*1,030*	*1,110*	*753*	*343*	*1,175*	*1,111*
Condom Use- Married	0.05	0.05	0.02	0.03	0.01	0.02	0.09	0.07
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	4,198	*4,118*	*4,646*	*4,875*	*10,051*	*11,682*	*4,547*	*4,210*
Number of Partners	1.07	1.04	1.03	1.02	1.02	1.01	1.03	1.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	5,286	*5,159*	*5,680*	*5,990*	*10,763*	*12,035*	*5,728*	*5,322*
Ever tested	0.14	0.41	0.15	0.59	0.05	0.07	0.09	0.41
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
	6,846	*6,647*	*8,049*	*8,326*	*11,329*	*14,472*	*7,589*	*7,116*
Age married	18.14	18.49	18.88	18.95	16.42	16.56	17.49	17.91
	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)
	5,175	*5,099*	*5,728*	*5,904*	*11,075*	*12,835*	*5,802*	*5,205*

Based on adults ages 15-49. Standard errors in parentheses,number of observations in italics, all means calculated using DHS provided sample weights; mean condom use and number of partners calculated based on people who reported more than zero partners in last 12 months, age of first marriage based on respondents who report ever having been married.

Table 2.3: Male Risk Behaviors by Wave and Country

	Tanzania		Kenya		Mali		Zambia	
	Wave 1	**Wave 2**	**Wave 1**	**Wave 2**	**Wave 1**	**Wave 2**	**Wave 1**	**Wave 2**
Condom Use-Single	0.46	0.50	0.47	0.61	0.31	0.33	0.40	0.47
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)
	1,231	*1,039*	*740*	*720*	*515*	*346*	*492*	*1,297*
Condom Use- Married	0.08	0.09	0.05	0.09	0.03	0.07	0.11	0.15
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
	2,928	*2,712*	*1,626*	*1,609*	*1,709*	*2,200*	*1,085*	*3,144*
Number of Partners	1.42	1.33	1.25	1.20	1.35	1.37	1.41	1.29
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)
	4,163	*3,751*	*2,366*	*2,328*	*2,180*	*2,540*	*1,572*	*4,437*
Ever tested	0.15	0.29	0.17	0.42	0.09	0.09	0.13	0.23
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	5,651	*5,184*	*3,318*	*3,254*	*2,879*	*3,380*	*1,951*	*5,971*
Age married	22.92	23.50	23.84	23.95	24.83	23.93	22.42	22.68
	(0.08)	(0.09)	(0.10)	(0.11)	(0.12)	(0.11)	(0.11)	(0.07)
	3,328	*3,009*	*1,770*	*1,761*	*1,770*	*2,395*	*1,191*	*3,461*

Based on adults ages 15-49. Standard errors in parentheses,number of observations in italics, all means calculated using DHS provided sample weights; mean condom use and number of partners calculated based on people who reported more than zero partners in last 12 months, age of first marriage based on respondents who report ever having been married.

Table 2.4: Changes in Women's Behavioral Outcomes by Year of Education

Years of Completed Education	**Condom Use- Single**		**Condom Use- Married**	**Number of Partners**		**Ever Tested for HIV**		**Age First Married**	
0	0.04	*	0.00	-0.01		0.08	*	0.26	*
1	-0.04		-0.03	-0.01	*	0.22	*	0.23	
2	0.00	*	-0.01	-0.01		0.24	*	0.26	
3	0.06		-0.01	-0.02	*	0.22	*	0.10	
4	-0.03		-0.01	-0.03	*	0.25	*	0.33	*
5	0.09	*	-0.02	0.00		0.23	*	-0.11	
6	0.12	*	-0.01	-0.01		0.27	*	-0.21	
7	0.06	*	0.01	-0.02	*	0.32	*	0.16	*
8	0.03		0.01	-0.01		0.40	*	-0.41	*
9	0.07	*	0.01	0.00		0.31	*	-0.38	*
10	0.16	*	0.02	0.02		0.31	*	-0.06	
11	0.04		0.03	0.02		0.25	*	-0.69	*
12	0.00	*	-0.01	-0.01	*	0.34	*	-0.22	
More than 12 Years	0.03		0.02	-0.01		0.31		-0.22	*

Differences between means calculated using DHS sample weights

* Indicates change between waves statistically significant with a p-value<=0.05

Table 2.5: Changes in Men's Behavioral Outcomes by Year of Education

Years of Completed Education	**Condom Use-Single**		**Condom Use-Married**		**Number of Partners**		**Ever Tested for HIV**		**Age First Married**	
0	0.04		0.02		0.00		0.04	*	-0.13	*
1	0.13		0.02		-0.19		0.10	*	0.64	*
2	0.12	*	0.03		-0.06	*	0.06		0.18	
3	0.07		0.05		-0.07		0.08	*	-0.85	
4	0.10	*	0.05	*	-0.07		0.07	*	-0.58	*
5	0.05		0.07	*	-0.09	*	0.08	*	0.12	
6	0.04		0.05	*	-0.09		0.09	*	-0.04	*
7	0.00		0.03		-0.07		0.14	*	0.25	
8	0.09	*	0.06	*	0.02		0.13	*	-0.20	
9	0.01		0.03	*	-0.06		0.12	*	-0.06	
10	-0.03		0.10		-0.01	*	0.13	*	-0.43	*
11	0.06		0.09		-0.04		0.15	*	-1.35	*
12	0.11	*	0.03		-0.14	*	0.20	*	-0.01	
More than 12 years	-0.02		0.05		-0.06		0.17		-0.30	

Differences between means calculated using DHS sample weights

* Indicates change between waves statistically significant with a pvalue<=0.05

Table 2.6: Regressions of Women's Behaviors on Education and Wave

	Condom Use-Single	Condom Use-Married	Number of Partners	Ever Tested for HIV	Age First Married
Years of Education	0.020 ***	0.002 *	0.001	0.001	0.015
	(0.004)	(0.001)	(0.002)	(0.001)	(0.015)
Years of Education Squared	0.000	0.000 *	0.000	0.000	0.026 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Wave 2	0.074 ***	0.003	-0.008 *	0.120 ***	0.186 ***
	(0.020)	(0.002)	(0.003)	(0.004)	(0.047)
Years of Ed*Wave2	-0.004	0.000	0.000	0.024 ***	-0.044 ***
	(0.003)	(0.001)	(0.000)	(0.001)	(0.008)
P-value for Ftest of joint significance of Years of Education and its square	0.000	0.000	0.235	0.000	0.008
N	7386	44464	51908	65382	51640

Based on adults aged 15-49 in both waves, robust standard errors are clustered by region and country and reported in parentheses;
Additional controls are included for age, age2, age3, urban residence, current marital status and country by region fixed effects.
Models for 'condom use' and 'number of partners' are based on respondents who report themselves sexually active in last 12 months
Model for 'age at first marriage' based on respondents who report ever having been married
* p<0.05, **p<0.01, ***p<0.001

Table 2.7: Regressions of Men's Behaviors on Education and Wave

	Condom Use-Single	Condom Use-Married	Number of Partners	Ever Tested for HIV	Age First Married
Years of Education	0.021 ***	0.007 ***	0.008	-0.003	-0.100 ***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.03)
Years of Education Squared	0.000	0.000 *	0.000	0.001 ***	0.021 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wave 2	0.057 *	0.017 *	-0.038	0.039 ***	-0.018
	(0.03)	(0.01)	(0.02)	(0.01)	(0.14)
Years of Ed*Wave2	-0.001	0.001	-0.007 *	0.014 ***	-0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
P-value for Ftest of joint significance of Years of Education and its square	0.000	0.001	0.263	0.000	0.008
N	6281	14739	21003	29157	16136

Based on adults aged 15-49 in both waves, robust standard errors are clustered by region and country and reported in parentheses; Additional controls are included for age, age2, age3, urban residence, current marital status and country by region fixed effects.

Models for 'condom use' and 'number of partners' are based on respondents who report themselves sexually active in last 12 months

Model for 'age at first marriage' based on respondents who report ever having been married

* p<0.05, **p<0.01, ***p<0.001

Table 2.8: Regressions of Behaviors on Education and Wave:HIV Tested Women

	Condom Use-Single	Condom Use-Married	Number of Partners	Ever Tested for HIV	Age First Married
Years of Education	0.028 ***	0.001	-0.001	0.005 *	-0.009
	(0.01)	(0.00)	(0.00)	(0.00)	(0.03)
Years of Education Squared	0.000	0.000	0.000 *	0.000 *	0.030 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wave 2	0.137 ***	0.008	-0.047 ***	0.282 ***	0.444 **
	(0.04)	(0.01)	(0.01)	(0.01)	(0.14)
Years of Ed*Wave2	-0.010 +	0.000	0.004 ***	0.010 ***	-0.059 **
	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)
HIV Positive	0.087 *	0.048 ***	0.021	0.022	0.065
	(0.03)	(0.01)	(0.02)	(0.02)	(0.16)
P-value for Ftest of joint significance of Years of Education and its square	0.000	0.052	0.003	0.000	0.000
N	2680	10666	13356	17679	12891

Based on adults aged 15-49 in both waves, robust standard errors are clustered by region and country and reported in parentheses; Additional controls are included for age, age2, age3, urban residence, current marital status andcountry by region fixed effects.

Models for 'condom use' and 'number of partners' are based on respondents who report themselves sexually active in last 12 months

Model for 'age at first marriage' based on respondents who report ever having been married

+p<0.10, * p<0.05, **p<0.01, ***p<0.001

Table 2.9: Regressions of Behaviors on Education and Wave:HIV Tested Men

	Condom Use-Single	Condom Use-Married	Number of Partners	Ever Tested for HIV	Age First Married
Years of Education	0.021 **	0.002	-0.005	-0.002	-0.109 *
	(0.01)	(0.00)	(0.01)	(0.00)	(0.05)
Years of Education Squared	0.000	0.000	0.000	0.001 ***	0.024 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wave 2	0.086 *	0.035 *	-0.116 *	0.072 ***	0.435 +
	(0.04)	(0.02)	(0.05)	(0.02)	(0.24)
Years of Ed*Wave2	-0.001	-0.002	0.003	0.016 ***	-0.051 +
	(0.01)	(0.00)	(0.01)	(0.00)	(0.03)
HIV Positive	0.044	0.050 **	-0.026	0.020	-0.458 +
	(0.05)	(0.02)	(0.04)	(0.02)	(0.25)
P-value for Ftest of joint significance of Years of Education and its square	0.000	0.085	0.207	0.000	0.000
N	3170	6631	9806	13867	7355

Based on adults aged 15-49 in both waves, robust standard errors are clustered by region and country and reported in parentheses; Additional controls are included for age, age2, age3, urban residence, current marital status and countryby region fixed effects.

Models for 'condom use' and 'number of partners' are based on respondents who report themselves sexually active in last 12 months

Model for 'age at first marriage' based on respondents who report ever having been married

+p<0.10, * p<0.05, **p<0.01, ***p<0.001

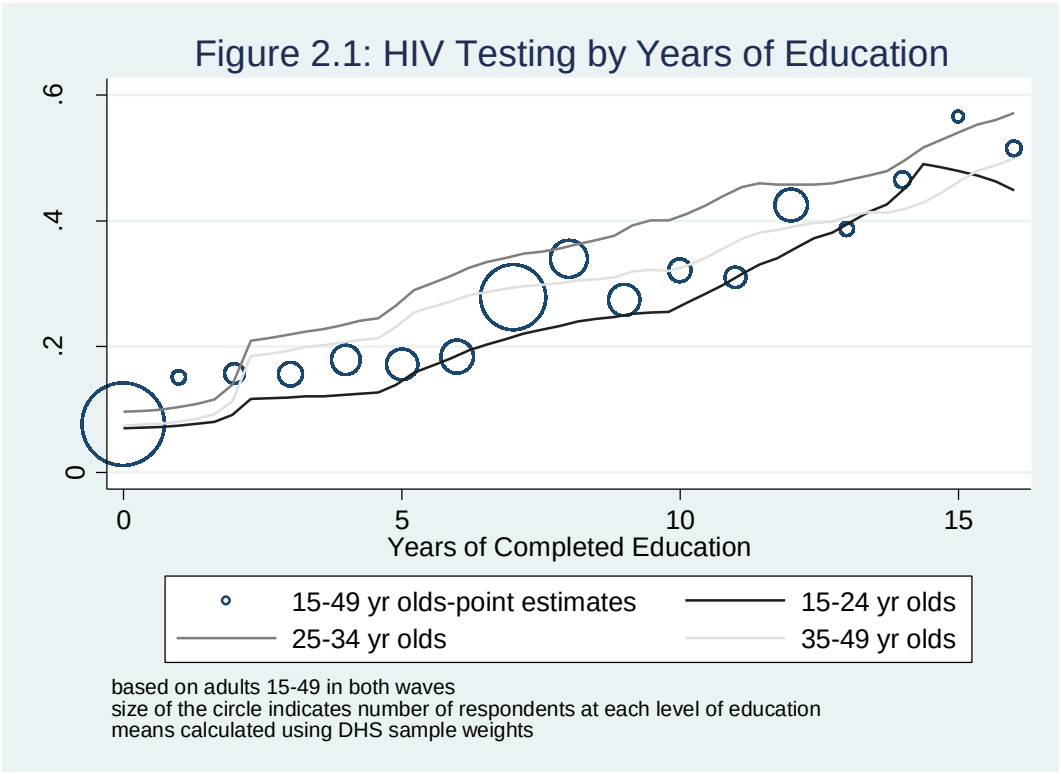
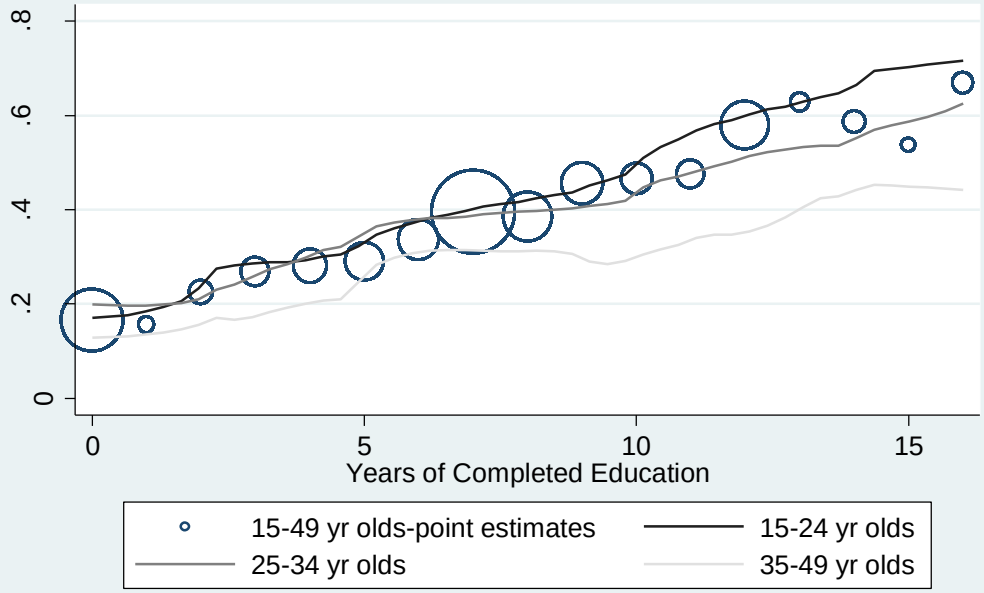
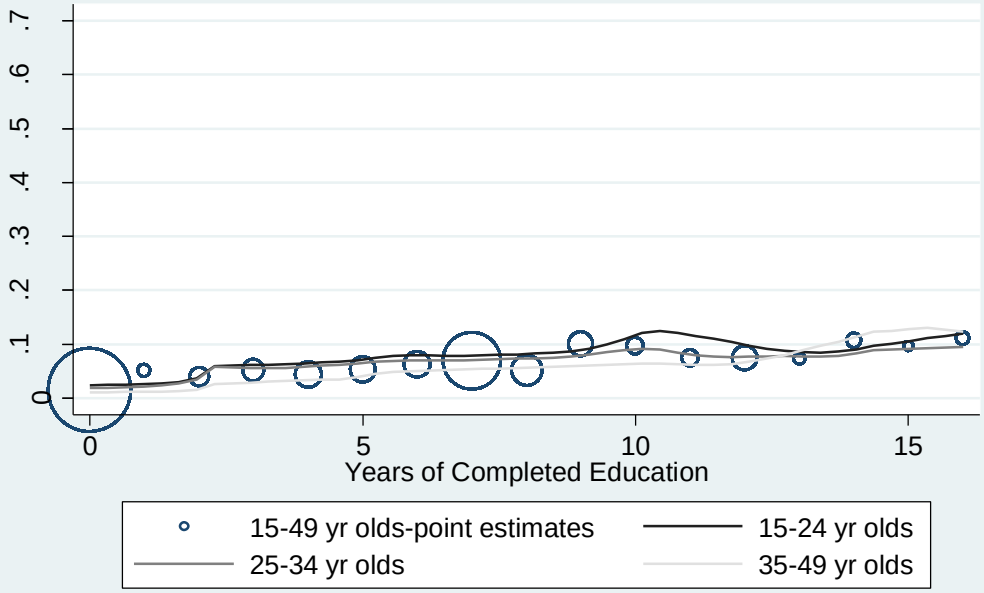

Figure 2.1: HIV Testing by Years of Education
0
.2
.4
.6
0
5
10
15
Years of Completed Education
15-49 yr olds-point estimates
15-24 yr olds
25-34 yr olds
35-49 yr olds
based on adults 15-49 in both waves
size of the circle indicates number of respondents at each level of education
means calculated using DHS sample weights

Figure 2.2a: Condom Use by Years of Education, Single



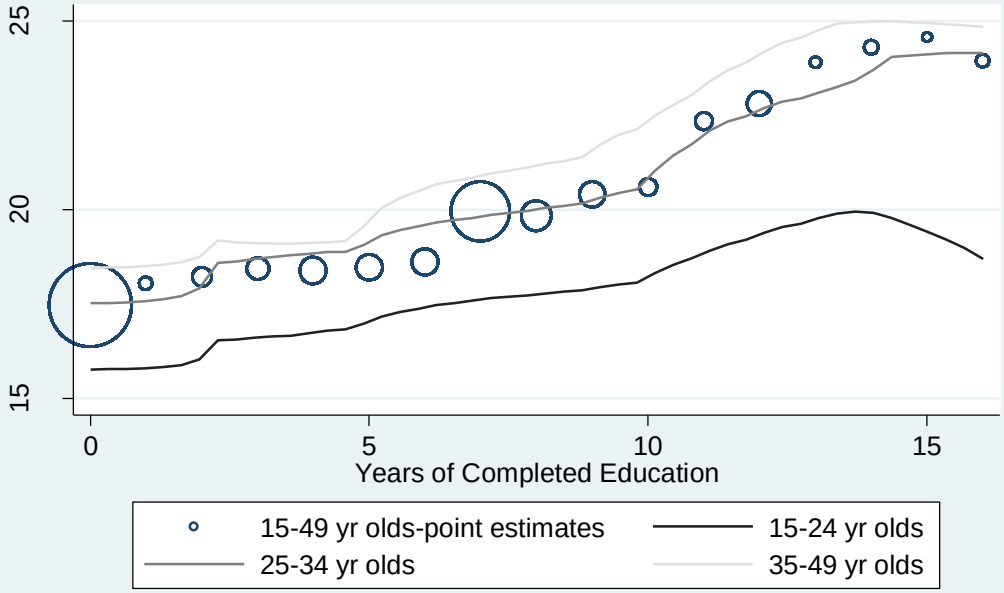
based on adults 15-49 who report themselves unmarried and sexually active in both waves
size of the circle indicates number of respondents at each level of education
means calculated using DHS sample weights

Figure 2.2b: Condom Use by Years of Education, Married

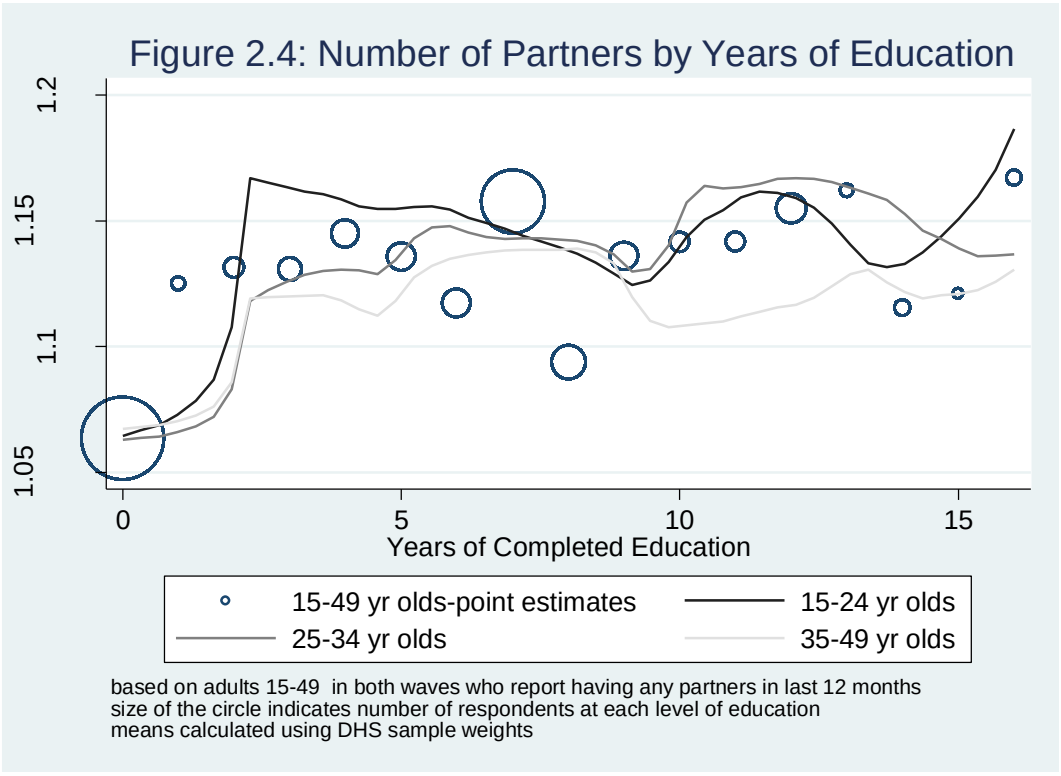


based on adults 15-49 who report themselves married and sexually active in both waves
size of the circle indicates number of respondents at each level of education
means calculated using DHS sample weights

Figure 2.3: Age at First Marriage by Years of Education



based on adults 15-49 in both waves who report themselves ever having married
size of the circle indicates number of respondents at each level of education
means calculated using DHS sample weights


Figure 2.4: Number of Partners by Years of Education
1.05
1.1
1.15
1.2
0
5
10
15
Years of Completed Education
15-49 yr olds-point estimates
15-24 yr olds
25-34 yr olds
35-49 yr olds
based on adults 15-49 in both waves who report having any partners in last 12 months
size of the circle indicates number of respondents at each level of education
means calculated using DHS sample weights

Chapter 3: The Urban Health Advantage in South Africa

Abstract*:

I use anthropometric scores from two nationally representative surveys in South Africa to examine the comparative changes in urban and rural children's health over the 15 years post-Apartheid. I find that the urban advantage in children's health disappears despite urban children retaining considerable advantages in average household socioeconomic status. I then explore several explanations for this pattern common in the urbanization literature, including the growth of particularly vulnerable urban populations, such as migrants and residents of informal areas, and/or the deepening of household poverty in the poorest urban areas. I do not find evidence of deteriorated circumstances for the urban poor, but I do find evidence that urban-rural migrants have begun to show a health disadvantage. However it appears to be related to the relative poverty of urban-rural migrants. I then follow the Dinardo Fortin Lemeiux decomposition method to explore relative shifts in the distribution household wealth and parental education in urban and rural households. I find that the differential gains are likely due to large improvements made by very poor rural households rather than deterioration in urban wellbeing.

*Portions of this work have been presented at the Population Association of America Annual Meeting in April 2010 and at the European Population Conference in September 2011

I. Introduction

Across the world, children are now more likely than ever before to be living in a city. In the developing world in particular, scholars have documented unprecedented urbanization rates in recent decades (Montgomery 2008; Cohen 2006 ; Tienda et. al. 2006). As adults move en masse to urban areas, either in search of opportunity or fleeing harsh circumstance, the number of children who will grow up in urban areas increases correspondingly. Moreover, the picture of a 'typical urban child' in the developing world may be shifting just as quickly. In recent history, the 'typical urban child' has been a child with a myriad of advantages over her rural counterparts. She has enjoyed a comparatively richer home, more educated parents, better access to schooling and health care, better household sanitation, less threat of food scarcity, and other fruits of the urban physical and social infrastructure. As a result, urban children have traditionally exhibited lower mortality rates and greater height and weight relative to rural counterparts (Brockerhoff 1994; Cleland, Bicego and Fegan 1992; Van de Poel, Odonnel and Doorslaer 2007).

However, in recent years, scholars have begun calling attention to a competing vision of the 'typical urban child'; this child has either been born into or moved to a city which is expanding faster than the infrastructure can reasonably support. She may live in a slum area without access to adequate sanitation, education or health care. She may experience violence, extreme poverty, and increased exposure to communicable disease as a result of crowded living quarters and poor household infrastructure. Rather than

enjoying better health than her rural counterparts, she faces equivalent or greater threat of disease and malnutrition.

Urban children in the developing world undoubtedly embody both of these extremes and the spectrum in between, but scholars of urbanization have been warning that the long established phenomenon of the 'urban health advantage' may be disappearing as increasing numbers of children face the grimmer urban scenario above. Recent studies in developing country contexts have found that although the average health of urban residents is still higher than the average of their rural counterparts, such averages obscure the disadvantages suffered by the poorest urban dwellers, who may have health outcomes equal to or worse than rural residents (Fotso 2007; Montgomery et. al. 2003; Van De Poel, O'Donnel & Doorslaer 2007).

Among the chief hypothesized causes of the disappearance of the urban health advantage is the growth of urban slums resulting from the rapid expansion of urban centers. When governments are unable to expand urban health infrastructure to keep pace with population growth, overall health can decline for urban residents, particularly for recent migrants and those living in already underserved informal slums and townships (Matthews et al 2010; Montgomery 2009; Cohen 2006; Fotso 2007). Moreover, newly arriving urban migrants appear to enjoy less of a health advantage than in the past (Cohen 2006). While these phenomena are at work across the globe, they can be particularly exaggerated in Africa, where civil conflict and agricultural collapse have led to especially high levels of migration coupled with strained state resources (Cohen 2006; Tienda et. al.

2006). In recent decades, countries in Africa have gone from having 10% and 30% of their population in urban areas in the 1970s, to 30 to 60% in 2005 (Zlotnik 2006).

If the health effects of urban residence are truly reversing, the scenario is particularly grim where it applies to children. A growing body of evidence points to the singular importance of early life health for children. The nutritional and disease environments that children experience in their early years have long term consequences for their adult health and human capital (Case and Paxson 2009, 2010). If urban environments are becoming more of a health threat than a health boon, then policy makers should be concerned with the health of the growing urban population for both the immediate and long-term wellbeing of their populace.

This paper examines South Africa over the 15 years following the end of Apartheid as a case study of the changing urban health advantage in Africa. I first examine the comparative health of urban and rural children in 1993 and then again in 2008- finding a distinct urban advantage in average child height and weight scores in 1993 but no such advantage in 2008. However, contrary to the concerns of urbanization scholars, the closing of the rural-urban health gap appears to result primarily from gains in health made by rural children than from declines in health among urban children. Interestingly, these gains have been made despite urban children retaining considerable advantages over rural children in household income, maternal education and other socioeconomic endowments typically associated with better health.

The literature highlights several potential explanations for this apparent paradox, including the growth of vulnerable urban populations (such as migrants and residents of

informal areas), and the deepening of poverty and corresponding decline in health for the urban poor. I examine the relative health of urban children in informal areas and migrant children and find that children living in informal urban areas in 2008 are not significantly different in terms of average health scores than children living in formal urban areas. I also find that while children of recent migrants are less healthy than non-migrants, they do not account for the difference in urban and rural health gains. I find no evidence of dramatic increases in poverty in urban areas, but some evidence that change in the distribution of socioeconomic endowments in *rural* areas has contributed to the large improvement in rural health. There is evidence that poor rural households gained more in income over the period in question than poor urban households, which has implications for the urban/rural health gap.

The remainder of the paper is organized as follows. Sections 2 and 3 introduce the empirical strategy and data and measures, respectively. Section 4 presents the results of the cross sectional comparison, and then evaluates empirical support for the expansion or deterioration of vulnerable urban populations, deepening of urban poverty, and distributional shifts of household socioeconomic characteristics in rural areas. Section 5 concludes.

II. Setting and Empirical Strategy

South Africa makes a particularly interesting case to examine the resilience of urban advantage, as it has both a relatively high urbanization rate and a history as a major migrant destination in southern Africa. According to the United Nations projections, 71%

of South Africans will be living in urban areas by 2030 (United Nations 2009). Citizens of neighboring African countries have traditionally been drawn to South African cities as sites of relative economic opportunity; but it is also true for South Africans themselves, who migrate internally at very high rates (Posel 2006, 2009). Migrants, however, look different than they traditionally have. The typical migrant in South Africa is now younger and increasingly likely to be female, less likely to report themselves as moving for work, and less likely to report an expectation of returning home (Posel 2009). In short, South Africa, like many of other developing nations, is experiencing rapid urbanization driven by migration of a new character and pace. This makes it an attractive setting to examine the trends in urban children's health, with particular attention to children of migrants and children living in informal areas.

The empirical framework is a simple multiple cross sectional comparison in two parts. I use two nationally representative household surveys taken 15 years apart, in 1993 and 2008, and compare urban and rural children's outcomes in both. In part one of the analysis, urban and rural children in 1993 are compared across four health outcomes based on direct anthropometric measurements of the children *height for age z-scores* (HAZ) , *weight for age z-scores* (WAZ), *stunting,* and *wasting*.[24] I then use the same four measures and examine urban and rural children in the 2008 sample to observe whether the urban health advantage still exists. In the second part of the paper, I use regression analysis to test the determinants of the urban advantage in both surveys. Finally, I follow the Dinardo Fortin Lemieux nonparametric decomposition method to examine the

[24] The reference population used here is the World Health Organization growth standard. However, I have performed the same analysis the CDC reference population and found that the conclusions are robust to the use of different standard populations.

relative impacts of simultaneous changes in *distribution* of multiple household socioeconomic endowments on the differences in urban and rural children's health gains.

III. Data and Measures

The analysis takes advantage of two nationally representative cross sectional household surveys from South Africa conducted 15 years apart. The first, the Project for Statistics on Living Standards and Development (PSLSD) was conducted in 1993 at the close of the Apartheid government. This survey measured health and economic indicators in 8,809 households across South Africa, and included adult and child modules administered to 43, 687 people. Most importantly for this comparison, all resident children under the age of six were physically weighed and measured, providing direct information on their relative health status for age.

The second cross section of data comes from a new nationally representative data set from South Africa-The National Income Dynamics Study (NIDS). This comprehensive survey of over 7000 households combines individual level and household level questionnaires along with direct anthropometric measurements of adults and children. It also contains specific information on residence in urban formal and informal settlements. Collectively, this information allows direct evaluation of the health, socioeconomic profile, and place of residence for the 31,170 adults and children surveyed.

Together, these surveys provide two comparable snapshots of South African households taken 15 years apart. Further, the fifteen years covered by these two surveys have been ones of enormous social and economic change in South Africa. Although the end of Apartheid meant that racial inequality was no longer dictated by law, the legacy of racial discrimination is still very much evident. As such, it would be disingenuous to compare all South African children and households as similar, and this analysis restricts itself only to the majority racial group- Black South Africans. It is further restricted to analyzing health outcomes for children under the age of six due to the fact that the earlier survey only took measurements for children who were under school age. Ultimately, I have full data for 4332 children and their households in 1993, and I compare them to 2893 children and their households in 2008.

Measures:

The measurement of children's health is satisfied here through the direct measurement of the height in centimeters and weight in kilograms of all children under 6 years of age in the survey households. Those raw anthropometric measurements were then converted to standardized Z-scores for both height and weight using the World Health Organization standard measures. The two measures traditionally reflect slightly different aspects of children's health. Low height for age reflects cumulative nutritional disadvantage, whereas low weight for age often reflects acute nutritional deprivation. I additionally examine two measures of severe malnutrition- whether children are stunted,

i.e. more than two standard deviations below mean height for age, and whether children show wasting (more than two standard deviations below average weight for age).

Table 3.1 shows the weighted means of these measures for urban and rural children in both samples as well as means for maternal characteristics. Maternal measures are significant because older and more educated mothers have been consistently shown to have healthier children (Caldwell 1979; Cleland and Van Ginneken 1988; Hobcraft 1993). Table 3.1 also lists sample means of household characteristics that are known to be relevant for children's health, such as household wealth variables and education and age of the household adults. Household wealth is particularly important for children's health because wealthier households may afford better nutrition and have better access to preventative and curative health care for the children. Household income and expenditure is directly measured in both surveys and the income measure reported includes wage, grant and remittance income received by all household members in the last month.[25]

Finally, the analysis pays special attention to migrant children and children living in informal urban areas. I chose to define 'migrant child' as a child under six who lives with a mother who has migrated in the last 5 years. I use recent maternal migration rather than child migration because of the growing body of evidence that maternal health prior to the child's birth is a strong determinant of child health in the ensuing years (Abu Saad and Fraser 2010). Even if a child was born in the new location, the stresses on the mother during migration might have effects on child health. As such, any child whose mother's

[25] For ease of comparison, the Rand amounts in the 1993 sample have been converted into the equivalent of 2008 Rand using the CPI as reported by Statistics South Africa at http://www.statssa.gov.za/keyindicators/cpi.asp

migrated from a different district council in the last 5 years is categorized as a 'migrant'.[26] Moreover, since mothers and children very often migrate together, there is very high correlation between recent maternal migration and child migration.[27]

The identification of children living in informal areas was done by constructing variable using information on informal housing and urban residence Although the 2008 survey directly included a category for 'urban informal resident', the 1993 survey did not. To ensure comparability across surveys, I constructed a comparable variable in both surveys. I categorized children who were identified as living in "informal dwelling" or "shack" and resident in urban areas as 'living in informal area'. Using informal housing as a proxy for living in an informal urban area is not ideal. However, in a comparison of the constructed variable and the 'residence in an informal area' identifier in the 2008 data, there were only 19 children who were not identified similarly in both. This lends some confidence that, although imperfect, it is a reasonable supposition that children living in informal housing in urban areas are also living in informal urban areas.

IV. Results

Urban Health Post-Apartheid and 15 Years Later

Several points are evident from the mean anthropometric scores reported in Table 3.1. First, although Black South African children still lag behind their global

[26] Using the more restrictive definition and only counting children who physically migrated themselves as 'migrant' does not change results.
[27] In the 1993 sample, 77% of children who had undergone migration lived with mothers who migrated, and 93% of the child migrants in the 2008 sample also had mothers who migrated in the last 5 years.

counterparts, time has been good to South African children on the whole. In the 1993 sample, Black South African children are approximately one and a half standard deviations below the global standardized population in height for age, and just shy of one standard deviation below mean weight for age. In addition, 36% and 21% of the under 6 population qualify as stunted and wasted, respectively. By 2008, the mean height for age has improved by 0.55 standard deviations and the mean weight for age has improved by 0.67 standard deviations, bringing South African children within one standard deviation of the global mean on both scores. In addition there has been a 14 percentage point reduction in stunting and a 13 percentage point reduction in wasting.

Moreover, Table 3.1 illustrates that there were statistically significant differences in the health measures of urban and rural children in 1993 which had largely disappeared by 2008. In the sample taken immediately post Apartheid, urban children are, on average, more than a quarter of a standard deviation taller than rural children. They are also about 7 percentage points less likely to be stunted. The average difference in weight for age is not as extreme, but urban children are still 0.13 standard deviations heavier than their rural counterparts. There are quite high levels of wasting, with about 1 in 5 Black South African children qualifying as wasted by the WHO global standard, but the levels of wasting are not appreciatively different between urban and rural populations. In contrast, in 2008, there are no statistically significant height and weight differences between urban and rural children, nor are there any differences in levels of wasting and stunting. It seems that both urban and rural children have made marked improvements in nutrition, with rural children making much larger strides, allowing them to catch up with urban kids.

Figures 3.1 and 3.2 elaborate on the means comparisons in the regressions by mapping the distributions of height and weight-for-age separately by gender. The charts in Figure 3.1 depict the kernel density distributions of height and weight z-scores for urban versus rural boys in both survey years. The vertical lines are placed at the cut-off points for stunting and wasting, such that the proportion of the distribution to the left of the lines represent children who are either stunted or wasted by the WHO standard. In 1993, urban boys have a strong height advantage as the urban distribution is consistently to the right of the rural distribution. The weight distributions cross one another, but the bulk of the urban distribution lies higher than the rural. Further, in both cases, one sided Komolgorov-Smirnov tests of the urban distribution being higher were statistically significant with p-values<0.05. In contrast, the 2008 means and distributions of the rural and urban populations were not statistically significantly different. For girls (Figure 3.2) the overall pattern is the same, with statistically significantly higher distributions in 1993 and no discernable difference in 2008. It does appear that girls experienced a clearer weight advantage in 1993 whereas boys have a clearer height advantage. However, the overall patterns are similar enough that I combine the results for boys and girls for the remainder of this analysis.

There are a fair number of good reasons to expect urban children to have better health than rural children. The means in Table 3.1 demonstrate that in both the samples rural children are at a statistically significant disadvantage compared to urban children along most correlates of child health. Urban households earn more than 300 Rand more per capita in both surveys, have more assets, fewer children, and have more highly educated adults, on average, when compared to rural households with similarly aged

children. In addition, in both samples the mothers of rural children are, on average, younger and less educated than those of urban children. Further, for those household characteristics believed to be critical to children's health, the changes in rural households from 1993 to 2008 have been smaller than those in urban households. Where urban households gained 137 Rand per capita in monthly income, rural households gained only 78 Rand per person (in 2008 Rand). Rural mothers are only 0.28 years older, on average, compared to their urban counterparts who are 0.7 years older. Rural mothers have gained slightly more education than urban mothers (0.2 years) but still have more than a year and a half less educational attainment in 2008. It is, therefore, somewhat surprising that rural children have made larger gains in health outcomes than urban children over the period in question.[28] The remaining sections examine several potential explanations for this seemingly paradoxical pattern. First, I examine whether the growth of vulnerable (and potentially less healthy) populations- such as recent migrants and informal urban dwellers- can account for the smaller gains in urban health. I then evaluate whether the health of the poorest urban children has declined over time, bringing urban averages down despite gains elsewhere in the income distribution. Finally, I use a nonparametric decomposition to examine whether shifts in the distribution of socioeconomic characteristics can explain the pattern.

[28] One potential explanation is that returns to increases in household socioeconomic status are non-linear. In other words, the same or smaller gains in household wealth make more of a difference to those who are worse off to begin with. In a separate analysis (not shown) I find that, for black households, health returns to income and maternal education are largely the same in the bottom three household income quartiles, and begin to taper off in the top income quintile. The vast majority of children in the sample live in households in the bottom three quartiles, so this is unlikely to be a full explanation of the patterns here, however, I include a square term on the log of household income in all regressions to adjust for this non-linearity.

Are the comparatively small gains in urban health a result of expanding vulnerable populations in urban areas?

To begin to test whether the growth in numbers or deterioration in health of vulnerable urban populations can explain the comparatively small gains in urban children's health, I first test whether these populations are actually growing in urban areas, and second, whether they have worse health outcomes on average. Table 3.2 contains anthropometric scores for children residing in urban areas divided by formal housing status in Panel A, and migrant status in Panel B. Panel A demonstrates that although urban children in informal housing areas were at a height and weight disadvantage in 1993, they had caught up to their counterparts by 2008. Furthermore, the overall percentage of children living in informal housing has actually decreased between the waves. Although the samples are quite small, this evidence does not conform with arguments that the growth of the child population in informal areas or the deteriorating health of those children is holding down average urban health.

In contrast, the measurements of urban migrant children in Panel B are broadly supportive of the existing scholarship on potentially changing migration patterns. First, the percentage of migrant children in urban areas increases markedly between 1993 and 2008. Second, while there are no statistically significant differences in health outcomes between migrants and non-migrants in the 1993 sample, by 2008, migrant children are shorter and lighter than other urban children. Migrant children have both become a larger proportion of the urban population and have become less healthy than their peers over the 15 year period. However, migrant children may not be disadvantaged by their specific

experiences of migration, but simply because they are disproportionately living in poorer households. In fact, in an additional analysis (not shown) I find that children who share a household with migrants but who are *not* migrants themselves have similar height and weight disadvantage, suggesting that household characteristics are really the critical determinant of health. However, I include migrant status separately in the regressions in Tables 3.4 through 3.7 in order to account for the comparatively poorer health of migrant children in 2008.

Is deepening urban poverty depressing average urban health scores?

So far, any evidence that urban migrants and children living in informal housing are particularly disadvantaged seems mostly to do with the fact that they are poorer than other urban children. This doesn't negate the concerns advanced by scholars of urbanization. It simply highlights the underlying concern that what is really driving the disappearance of urban health advantage is increasing urban poverty (with migrants and informal residents being two types of urban poor). The argument that deteriorating circumstances for the urban poor is depressing average urban health could still be valid even if the focus on migration and informal areas is erroneous.

However, an analysis of the data does not bear this out. Table 3.3 shows that children in the lowest household income quintile in urban areas have actually increased in height for age and weight for age by a similar magnitude as the urban population as a whole. In 1993, children in urban households in the lowest income quintile are 1.6 standard deviations below the WHO standard of mean height and 1.1 below mean weight

for age. By 2008, they improve by 0.56 standard deviations in height for age and 0.79 standard deviations on weight for age. The same is true of the percentages of children stunted and wasted in the lowest income quintile. The poorest urban children improve in stunting by 18 percentage points and in wasting by 15 percentage points. Rather than the poorest pulling down the average improvements in height and weight, it appears that the height and weight of children in the lowest income quartiles has improved as much or more than that of wealthier children.

Further, it also doesn't appear that the number of children living in urban poverty is ballooning. In the 1993 sample, 20% of urban 0-6 year olds live in households in the lowest national income quintile. By 2008, 22% of the urban children are living in households in the lowest income quintile, so there is almost no change in the percentages living in relative poverty. In contrast, the percentage of rural 0-6 yr olds living in households in the lowest per capita income quintile falls from 52% to 38% between '93 and '08 respectively.

Figure 3.3 shows the changes in the kernel density distributions of log household income for households with children under six in both urban and rural areas. The figure showing urban income distributions doesn't lend a great deal of support to the theory that incomes are changing dramatically for the poor urban households. Rather, there is a backward shift in the middle of the distribution, which, although surprising, corresponds with previous work done on income levels in South Africa post Apartheid (Leibbrandt, Levinsohn and McCrary 2010). Conversely, rural income does appear to change considerably, with a large tail of very poor households seen in 1993 disappearing by

2008. If the children in these very poor households were the 'unhealthiest' (which seems believable), then the pattern we see of large health improvement in rural areas and less in urban areas would make a great deal of sense, even if overall mean urban income remains above mean rural income.

An important policy change took place during this period, which is likely related to the changes in Figure 3.3. During the post Apartheid period, the South African government carried out an aggressive expansion of social welfare grants. In particular, they introduced a grant particularly targeted to poor children, the Child Support Grant. In 1998, impoverished caregivers of young children started receiving a regular monthly grant of 110 Rand per child under 7.[29] While social welfare grants expanded overall, uptake of child support grants is extremely high, with 69% of rural households and 58% of urban households in the sample receiving a grant for at least one child. (Table 3.1) In a separate analysis, I explore whether child grant income has a positive association with children's anthropometric scores. I am unable to find that the child grant income has an effect distinct from other types of income, but I note that the expansion of the grant likely played an important role in the reduction of the proportion of children living in extreme poverty.[30]

Thus far, there is evidence that there were changes both in urban migrant health and in rural income distribution. I further explore the relationship between urban

[29] Since then the child support grant has been increased steadily (it is now 250 Rand) and expanded to include children up to 15 years of age.

[30] In addition, I performed the regression analyses in Tables 3.4-3.7 including child grant income separately from other types of income, as well as testing model including an indicator for households receiving child grant income. Neither model showed an impact of the child grant on anthropometric scores or the urban coefficient.

residence, socioeconomic status, and migrant status in Tables 3.4 through 3.7. I use least squares regression to predict height for age and weight for age, and probit models to predict probability of stunting and wasting. The generalized model for all the tables is

$$H_{ti=}\ \beta_0 + \beta_1 Urban_{ti} + \beta_2 Migrant_{ti} + \beta_3 X_{ti} + \beta_4 Z_{ti} + \varepsilon_{ti}$$

where H_{ti} is the health outcome score [height for age, weight for age, stunting or wasting] of individual *i* in year *t*, and β_1 is the coefficient for a dummy variable set to one if the child lives in an urban area, such that β_1 represents the average increase or decrease in health score an individual can expect from urban residence. Similarly, β_2 is the coefficient on a binary variable equal to one of the child is a migrant. X_{ti} is a vector of socioeconomic variables comprised of household and maternal characteristics, including log per capita income, log per capita income squared, number of assets held by the household, whether the household has piped water, years of education of the child's mother and age of the child's mother at the child's birth. Z_{it} is a vector of demographic variables including the child's gender, age, age squared, and province of residence. In all tables, the first column contains the model run solely on the demographic variables and urban residence. The second column adds in child's migrant status to test both whether it is a significant predictor of health independently and whether its inclusion in the model changes the size or significance of the coefficient on urban residence. The final two columns repeat the analysis adding in the vector of controls for socioeconomic status.

The most notable message of these tables is the power of the socioeconomic variables to explain the urban health advantage. Across the four outcomes, wherever there is a statistically significant urban health advantage, controlling for SES through

household wealth and maternal characteristics renders the coefficient on urban residence substantively and statistically non-significant. The coefficient on migrant status is not statistically significant and has no substantive effect on the size or significance of the urban coefficients in 1993. Similarly, the coefficient on migrant status is statistically insignificant for all the 2008 models. In the case of stunting, (Table 3.6) its inclusion renders the urban coefficient statistically significant when SES variables are not controlled for, but both urban and migrant coefficients become non-significant when they are. Its notable that the inclusion of an indicator for migrant status in the model in 2008 increases the coefficient on urban residence for height and weight compared to the models without it. Similarly, urban coefficients for stunting and wasting become more negative when migrant status is included. Although they are not significant, this pattern is in keeping with the theory that less healthy migrant children may be represent a downward pressure on the urban health means.

As additional step, I interacted migrant and informal status with the household socioeconomic variables to see if the protective effects of household income and maternal characteristics were different for migrants or informal urban residents, but I found no evidence that socioeconomic status works differently for migrant children or children in informal areas (results not shown). While these tables are largely descriptive, they suggest several things. To some extent they support the concerns of urbanization scholars that the health of urban migrant children has become worse. But a far the stronger finding is that, in South Africa, the urban advantage is entirely explained by socioeconomic status. In both sample years, household income and assets, and maternal age and education are consistently significant predictors of children's health (both jointly and

independently) and their inclusion in the models fully explains the differences between urban and rural children seen in the early sample.

Can the different trajectories of urban and rural children be explained by distributional changes in socioeconomic characteristics?

Given the relative primacy of socioeconomic predictors, the central puzzle of this analysis remains. If health is all about socioeconomic characteristics like household wealth and parental characteristics- then how come rural children, who make much smaller gains in those areas on average, have much bigger gains in health? Here I return to one of the central insights of the urbanization literature, that population averages are likely obscuring important changes in the distributions. Figure 3.3 lends credence to the idea that different levels of income growth at the bottom of the wealth spectrum could driving the differential changes between urban and rural areas, even if it can't be seen in the overall means. Nevertheless, income is not the sole determinant of socioeconomic status, and it is likely that household income works jointly with other factors such as parental education, local infrastructure and other determinants of household wellbeing to affect child health. The means in Table 3.1 indicate that although rural households had smaller gains in income and assets compared to urban households, they did see large increases in access to piped water over the period in question. In addition, both urban and rural households enjoyed expansion of social grant programs, with more rural households receiving grants. While the models in Tables 3.4-3.7 include controls for piped water and household income (including grant income), they are limited in what they can tell us

about the effect of simultaneous shifts in the distribution of multiple socioeconomic variables between surveys.

The final part of this analysis attempts to investigate the role that the changes in distribution of multiple endowments play in the findings by applying the nonparametric approach initially developed by Dinardo, Fortin and Lemiuex (1996) to examine differences in wage distributions. The most common application of this type of decomposition is looking at wage returns to endowments between different groups. For example, one might ask "Would blacks and whites have the same wage distribution if the two groups had the same education profile?" To test this, the DFL method creates a counterfactual where the education distribution is the same and then predicts the wage structure based upon the counterfactual distribution. If the wage functions still differ when endowments are equally distributed, then race probably has an effect on wages which is independent of the differences in education between black and white populations. For the purposes of this analysis, I am interested in comparing children in the 1993 and 2008 groups and asking how the distributions of children's height and weight in 2008 would differ if the whole vector of household and parental endowments (rather than just income) in 2008 had been distributed the same as they were in 1993.

To answer these types of questions, the DFL method takes the density distribution of one group and compares it to a hypothetical counterfactual density distribution created by weighting the data of the second group to match the endowment profile of first. For example, the density function of children's height can be written as:

$$f(y|T=t) = \int g(y|x, T=t)h(x|T=t)dx$$

where T is a random variable denoting the year from which an individual child in the pooled dataset is drawn, $g(y|x, T=t)$ is the density of individual height evaluated at *y* given the observable endowments *x* and survey year *t*, and $h(x|T=t)$ is the density of observable endowments evaluated at *x*, given that the survey year is *t*. In order to draw an analogy to a traditional parametric regression, it is helpful to think of $f(y|T=t)$ as equivalent to the dependent variable (height or weight for age) and to think of $h(x|T=t)$ as analogous to the observable endowments (i.e. income, maternal age and education, child's age etc) and to think of $g(y|x, T=t)$ as a function which "converts" the observable attributes to height and is therefore analogous to the 'returns to endowments'.[31] The creation of the counterfactual rests upon using the data from 2008 and reweighting it such that that the density of attributes from period two $h(x|T=t_{2008})$ matches that of period one $h(x|T=t_{1993})$ and then applying the returns function from period two $g(y|x, T=t_{2008})$. The resulting distribution of height for age can be thought of as the counterfactual of "what children's height would have looked in 2008 like if their households and parental characteristics were distributed in the same way they had been in 1993?"

For the purposes of this analysis, there are two advantages to this intellectual experiment. The first is that it takes into consideration the entire distribution of

[31] This analogy and DFL explanation is heavily borrowed and adapted from Leibbrandt, Levinshon and McCrary's excellent paper on incomes in South Africa (2010) For more detailed explanation of the methods and weighting schemes used to implement DFL decomposition see Johnston and Dinardo (1996) and Fortin Lemeiux and Firpo (2010)

endowments and produces a corresponding distribution of heights and weights, rather than relying solely on population means. Further, it gives concrete form to the changes in height/weight that you might expect from changing the entire vector of health producing endowments. Rather than just asking 'what would the distribution of children's height look like if the rural households in 2008 had the same distribution of income as they did in 1993?", I am able to extend the method to look at what the distribution of rural children's height might look like if *all* the observed socioeconomic variables were jointly set to their 1993 distributions.

I ask this separately of urban and rural children, and the resulting DFL kernel densities are shown in Figure 4 for height for age and Figure 5 for weight for age. The left hand panel has the urban height density in 2008, overlaid with the density that would have been if the endowments were distributed similarly to 1993. The vector of endowments used here are household income and income squared, maternal education, maternal age at birth, piped water supply, child's age and age squared, migrant status and gender. If urban children's health gains were somehow depressed by downward distributional shifts in socioeconomic endowments, then I would expect that holding the entire distribution of endowments steady at their 1993 levels would give rise to a higher distribution of height for age scores than we actually see in 2008 for urban residents. In fact, the curves are largely coincident for urban children. In other words, substituting the 1993 endowments for the 2008 endowments gives rise to the actual 2008 height distribution. This suggests that distributional shifts in these endowments are unlikely to have depressed urban health.

In contrast, the DFL graphs for rural children for height and weight show a counterfactual that is below that of the actual 2008 distribution, suggesting that in these cases, the changes in distribution of endowments led to increases in children's health. Combined with the descriptive evidence above, I interpret this as evidence that the disappearance of the urban health advantage cannot be attributed to changes in levels or distribution of urban poverty. Happily, it appears due primarily to the socioeconomic changes among poor rural households. More specifically, the disproportionate improvements in rural average health are likely due to the gains in SES enjoyed by the poorest rural households.

V. Conclusion

Ultimately, this analysis finds that the urban health advantage has disappeared in South Africa, but rather than being a result of deteriorating conditions in urban areas, it appears to be due to real improvements in the economic circumstances of those who were the worst off at the end of Apartheid in 1993 (i.e. the rural poor). Whether these optimistic findings are generalizable to other rapidly urbanizing countries is uncertain. However, one potential interpretation of the patterns documented here is that the expansion of grant programs specifically targeted to poor children has gone a long way towards decreasing health inequalities among children. I cannot demonstrate empirically that child centered social grants were a policy 'magic bullet'. Rather, they appear to be a component in the improvement of socioeconomic conditions for the very poor which, in turn, reduced health inequality.

In addition, I find evidence of a decline in the relative health of urban migrant children. [32] The relative poverty of urban migrants may be the most salient factor in the decline, but placed in the context of growing urbanization, this is a trend which should attract the attention of researchers and policy makers alike. Also, noticeably absent from this analysis is any discussion of community characteristics and health infrastructure, largely due to the difficulty of getting good national-level measures of these. Although recent work by scholars such as Van De Poel O'Donnell and Doorslaer (2009) has confirmed the primacy of household characteristics in determining children's health, community sanitation and health infrastructure continues to play an important role and may be especially relevant in impoverished areas. This analysis largely validates conventional wisdom that improvements in household SES lead to improvements in children's health. However, it does not negate the possibility that public health projects or other community interventions could be an alternate path to reducing health inequality. Evaluating the impact of such interventions as compared to simple expansion of income support for the poor seems a fruitful direction for future work on urbanization and children's health.

[32] To date, evidence of the effect of internal migration on children's health in Africa is quite limited. One notable exception is Brokerhoff (1995), which finds higher child mortality among the children of urban migrant women in a study of 15 developing countries, although the findings for Africa are less pronounced than those for other regions

Table 3.1: Characteristics of Sample Children and Households in 1993 and 2008

	1993				2008			
	Rural	**Urban**	**Combined**	*Urban Difference*	**Rural**	**Urban**	**Combined**	*Urban Difference*
Child characteristics (children aged 0-6)								
Mean Height for Age	-1.56	-1.30	-1.49	**0.26** *	-0.92	-0.97	-0.94	**-0.05**
Mean Weight for Age	-0.94	-0.81	-0.90	**0.13** *	-0.22	-0.25	-0.23	**-0.03**
Proportion Stunted	0.38	0.31	0.36	**-0.07** *	0.22	0.22	0.22	**0.00**
Proportion Wasted	0.21	0.21	0.21	**0.00**	0.07	0.08	0.08	**0.01**
Mean Child Age	2.42	2.46	2.44	**0.04**	3.00	2.93	2.97	**-0.07**
Proportion Male	0.50	0.53	0.51	**0.02**	0.53	0.50	0.52	**-0.03**
Proportion Migrants	0.04	0.10	0.06	**0.06**	0.21	0.46	0.30	**0.25**
Mother's characteristics								
Mean Maternal Years of Education	6.71	8.25	7.15	**1.54** *	8.62	9.87	9.04	**1.25** *
Mean Maternal Age at Child's Birth	25.99	26.68	26.18	**0.69** *	26.27	27.38	26.65	**1.10** *
Number of observations (children 0-6)	*3100*	*1232*	*4332*		*1939*	*954*	*2893*	
Household characteristics:								
Mean Per Capita Household Income	337.75	652.02	434.78	**314.27** *	416.16	789.44	554.58	**373.29** *
Mean Per Capita Household Expenditure	396.66	657.63	477.24	**260.97** *	300.24	623.22	405.18	**322.98** *
Mean Number of Assets†	2.45	3.97	2.92	**1.51** *	5.01	6.89	5.71	**1.87** *
Proportion of Households with Piped Water	0.42	0.99	0.60	**0.56** *	0.70	0.99	0.79	**0.28** *
Mean Number of Adults	3.55	3.70	3.60	**0.15**	2.99	2.71	2.88	**-0.28** *
Mean Number of Children	3.46	2.68	3.22	**-0.79** *	3.15	2.37	2.86	**-0.78** *
Mean Years Education of Household Adults	5.99	7.52	6.51	**1.52** *	6.83	9.10	7.67	**2.27** *
Mean Age of Household Adults	33.86	33.36	33.69	**-0.50**	38.34	34.82	37.03	**-3.52** *
Proportion Receiving Social Grants††	0.33	0.25	0.31	**-0.08** *	0.86	0.69	0.8	**-0.17** *
Proportion Receiving Grants for Children††	n/a	n/a	n/a		0.69	0.58	0.65	**-0.11** *
Number of Observations (Households with children 0-6)	*1865*	*833*	*2698*		*1203*	*709*	*1912*	

* indicates difference of means is statisitcally significant with a p-value <0.05. Means calculated using sample weights.

†All households were asked about a list of commonly held household assets such as a stove refrigerator, television etc. In the 1993 sample, households were asked a list of 11 assets, and in the 2008 sample, they were asked a list of 22 assets. As such, the asset counts are not directly comparable across years, although they are still informative about the difference in wealth between rural and urban households within the survey year.

†† Government social assistance in the 93 period included old age pension, disability grant, poor relief grant, unemployment insurance and government workmans compensation; in 2008 government grants include old age pension, child support grant, foster care grant, disability grant and care dependency grant; of those grants, child grants, foster care grants and care dependency grants can be categorized as 'social grants for children'.

Table 3.2: Anthropometric Means for Vulnerable Groups in 1993 and 2008

Panel A: Children in Urban Informal Housing

	1993			2008		
	Informal Housing	**Formal Housing**	*Difference (Informal-Formal)*	**Informal Housing**	**Formal Housing**	*Difference (Informal-Formal)*
Mean Height for Age	-1.60	-1.10	**-0.50** *	-1.00	-0.96	**-0.04**
Mean Weight for Age	-1.10	-0.67	**-0.43** *	-0.41	-0.31	**-0.10**
Proportion Stunted	0.42	0.26	**0.16** *	0.23	0.22	**0.01**
Proportion Wasted	0.25	0.18	**0.07** *	0.12	0.08	**0.04**
Number of Observations (Percent of urban sample)	396 *(32%)*	831 *(68%)*		185 *(20%)*	743 *(80%)*	
Percent of total sample	*9%*	*19%*		*7%*	*26%*	

Panel B: Urban Migrant Children

	1993			2008		
	Urban Migrant Children	**Urban Non-Migrant Children**	*Difference (Migrant-Non*	**Urban Migrant Children**	**Urban Non-Migrant Children**	*Difference (Migrant-Non*
Mean Height for Age	-1.30	-1.30	**0.00**	-1.10	-0.84	**-0.26** *
Mean Weight for Age	-0.64	-0.83	**0.19**	-0.46	-0.07	**-0.39** *
Proportion Stunted	0.29	0.31	**-0.02**	0.26	0.19	**0.07**
Proportion Wasted	0.18	0.21	**-0.03**	0.10	0.07	**0.03**
Number of Observations (Percent of urban sample)	123 *(10%)*	1108 *(90%)*		439 *(46%)*	515 *(54%)*	
Percent of total sample	*3%*	*26%*		*17%*	*19%*	

Means calculated using provided sample weights;* indicates difference is statistically significant with a p-value<0.05

Table 3.3: Anthropometric Scores by Income Quintile

	Urban Means 1993					Urban Means 2008				
Household Income Quintile	***1st***	*2nd*	*3rd*	*4th*	*5th*	*1st*	*2nd*	*3rd*	*4th*	*5th*
Mean Height for Age	**-1.62**	-1.44	-1.25	-0.83	-0.44	**-1.16**	-0.94	-1.02	-0.76	-0.78
Mean Weight for Age	**-1.10**	-0.91	-0.71	-0.51	-0.14	**-0.31**	-0.36	-0.29	-0.07	-0.05
Proportion Stunted	**0.40**	0.34	0.29	0.20	0.22	**0.26**	0.25	0.24	0.14	0.18
Proportion Wasted	**0.25**	0.22	0.19	0.18	0.06	**0.08**	0.12	0.09	0.04	0.08
Number of Observations	**210**	301	346	152	18	**130**	140	150	116	38
	Rural Means 1993					**Rural Means 2008**				
Household Income Quintile	***1st***	*2nd*	*3rd*	*4th*	*5th*	*1st*	*2nd*	*3rd*	*4th*	*5th*
Mean Height for Age	**-1.73**	-1.44	-1.22	-1.50	-1.71	**-0.97**	-0.89	-0.95	-0.74	-1.02
Mean Weight for Age	**-1.09**	-0.89	-0.52	-0.79	-1.43	**-0.36**	-0.18	-0.03	-0.07	-0.07
Proportion Stunted	**0.43**	0.36	0.24	0.27	0.43	**0.25**	0.20	0.20	0.20	0.17
Proportion Wasted	**0.24**	0.20	0.15	0.10	0.43	**0.09**	0.05	0.06	0.10	0.04
Number of Observations	**1313**	816	318	90	7	**532**	501	248	83	23

All means calculated using provided sample weights

Table 3.4: OLS models of Determinants of Height for Age Z-scores of Black Children 0-6

	Panel A: 1993 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	0.23 ***	0.24 ***	-0.01	-0.01
	(0.07)	(0.07)	(0.08)	(0.08)
Migrant Status		0.02		0.04
		(0.11)		(0.11)
Controls for SES	No	No	Yes	Yes
Observations	4329	4329	4329	4329
	Panel B: 2008 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	0.11	0.16	0.03	0.12
	(0.11)	(0.11)	(0.13)	(0.11)
Migrant Status		-0.10		-0.19
		(0.12)		(0.12)
Controls for SES	No	No	Yes	Yes
Observations	2893	2893	2893	2893

*** p<0.01, ** p<0.05, * p<0.1; Additional controls include province indicators, child's age, age squared and gender; Vector of controls for socioeconomic status includes household income, household income squared, maternal years of completed education, maternal age at child's birth, household asset count, and whether the household had piped water supply. All SES variables are significant predictors of child health individually and jointly when statistical significance is measured at p-value<0.05, with the exception of piped water, which was not significant independently. All regressions are weighted using sample weights, and standard errors (in parentheses) are robust and clustered at household level.

Table 3.5: OLS models of Determinants of Weight for Age Z-scores of Black Children 0-6

	Panel A: 1993 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	0.25 ***	0.24 ***	0.04	0.03
	(0.06)	(0.06)	(0.07)	(0.07)
Migrant Status		0.11		0.13
		(0.11)		(0.10)
Controls for SES	No	No	Yes	Yes
Observations	4331	4331	4331	4331
	Panel B: 2008 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	0.03	0.10	-0.08	-0.03
	(0.09)	(0.09)	(0.10)	(0.10)
Migrant Status		-0.16		-0.17
		(0.10)		(0.10)
Controls for SES	No	No	Yes	Yes
Observations	2859	2859	2859	2859

*** p<0.01, ** p<0.05, * p<0.1; Additional controls include province indicators, child's age, age squared and gender; Vector of controls for socioeconomic status includes household income, household income squared, maternal years of completed education, maternal age at child's birth, household asset count, and whether the household had piped water supply. All SES variables are significant predictors of child health individually and jointly when statistical significance is measured at p-value<0.05, with the exception of piped water, which was not significant independently. All regressions are weighted using sample weights, and standard errors (in parentheses) are robust and clustered at household level.

Table 3.6: Probit models of Determinants of Stunting among Black Children 0-6

	Panel A: 1993 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	-0.14 **	-0.14 *	0.07	0.07
	(0.05)	(0.06)	(0.07)	(0.07)
Migrant Status		-0.08		-0.09
		(0.10)		(0.10)
Controls for SES	No	No	Yes	Yes
Observations	4329	4329	4329	4329
	Panel B: 2008 Sample			
	(1)	(2)	(3)	(4)
Urban Residence	-0.14	-0.20 *	-0.09	-0.14
	(0.10)	(0.10)	(0.11)	(0.11)
Migrant Status		0.10		0.13
		(0.10)		(0.11)
Controls for SES	No	No	Yes	Yes
Observations	2893	2893	2893	2893

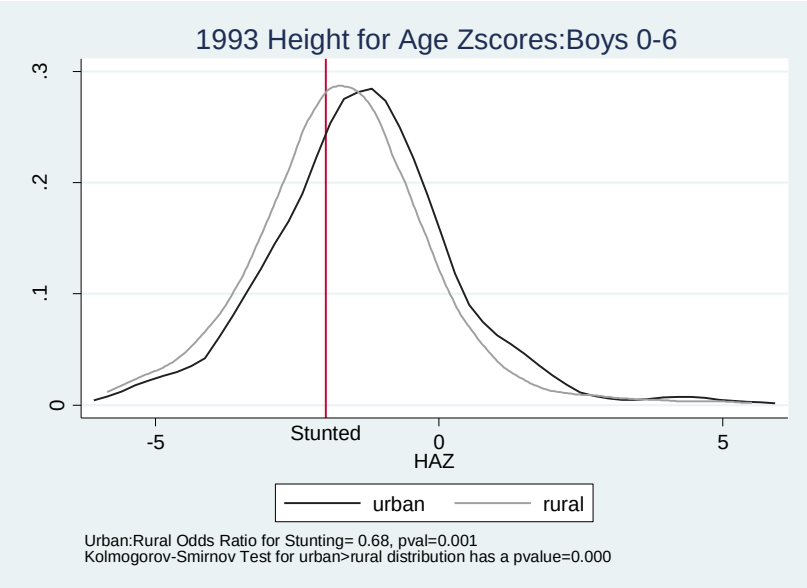
*** p<0.01, ** p<0.05, * p<0.1; Additional controls include province indicators, child's age, age squared and gender; Vector of controls for socioeconomic status includes household income, household income squared, maternal years of completed education, maternal age at child's birth, household asset count, and whether the household had piped water supply. All SES variables are significant predictors of child health individually and jointly when statistical significance is measured at p-value<0.05, with the exception of piped water, which was not significant independently. All regressions are weighted using sample weights, and standard errors (in parentheses) are robust and clustered at household level

Table 3.7: Probit models of Determinants of Wasting among Black Children 0-6

Panel A: 1993 Sample				
	(1)	(2)	(3)	(4)
Urban Residence	-0.11	-0.11	0.06	0.06
	(0.06)	(0.06)	(0.07)	(0.07)
Migrant Status		-0.10		-0.12
		(0.11)		(0.11)
Controls for SES	No	No	Yes	Yes
Observations	4330	4330	4330	4330
Panel B: 2008 Sample				
	(1)	(2)	(3)	(4)
Urban Residence	-0.05	-0.09	-0.02	-0.03
	(0.13)	(0.12)	(0.13)	(0.13)
Migrant Status		0.10		0.06
		(0.12)		(0.13)
Controls for SES	No	No	Yes	Yes
Observations	2859	2859	2859	2859

*** p<0.01, ** p<0.05, * p<0.1; Additional controls include province indicators, child's age, age squared and gender; Vector of controls for socioeconomic status includes household income, household income squared, maternal years of completed education, maternal age at child's birth, household asset count, and whether the household had piped water supply. All SES variables are significant predictors of child health individually and jointly when statistical significance is measured at p-value<0.05, with the exception of piped water, which was not significant independently. All regressions are weighted using sample weights, and standard errors (in parentheses) are robust and clustered at household level

Figure 3.1: Height and Weight for Age among Boys aged 0 to 6:



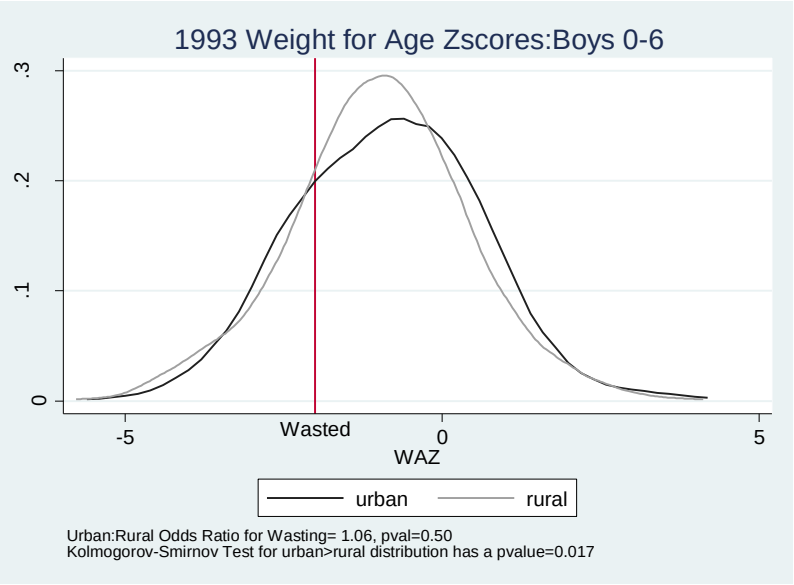
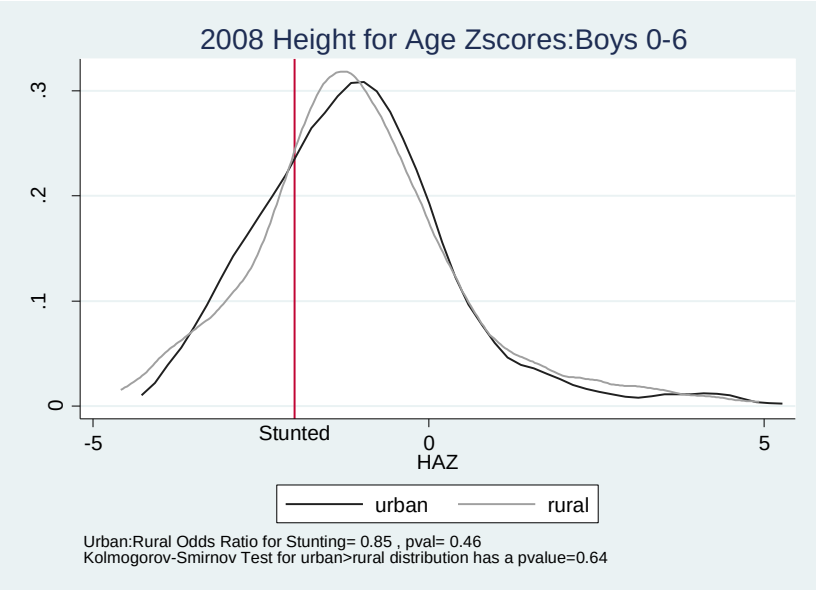


Figure 3.1: Height and Weight for Age among Boys aged 0 to 6:



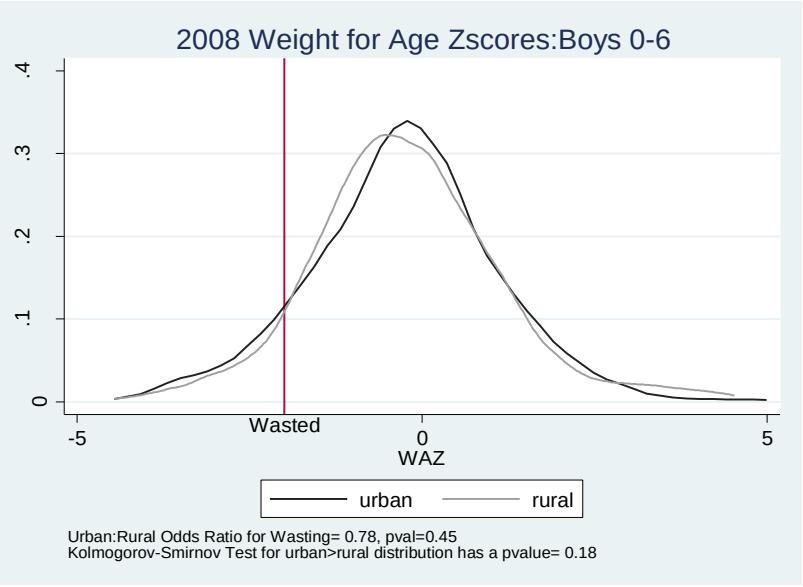
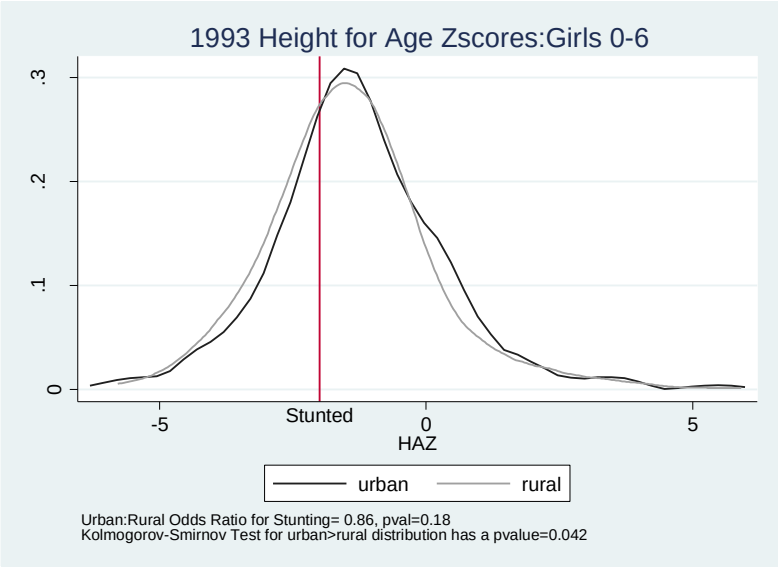


Figure 3.2: Height and Weight for Age among Girls aged 0 to 6:



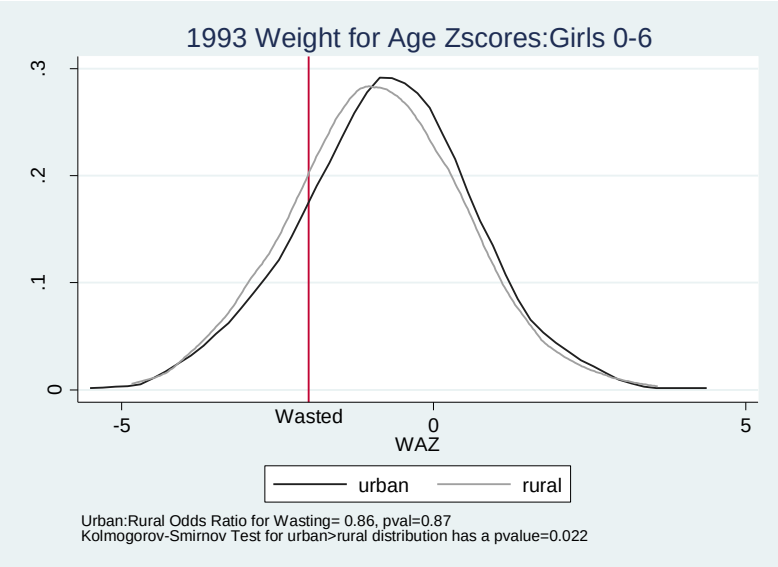
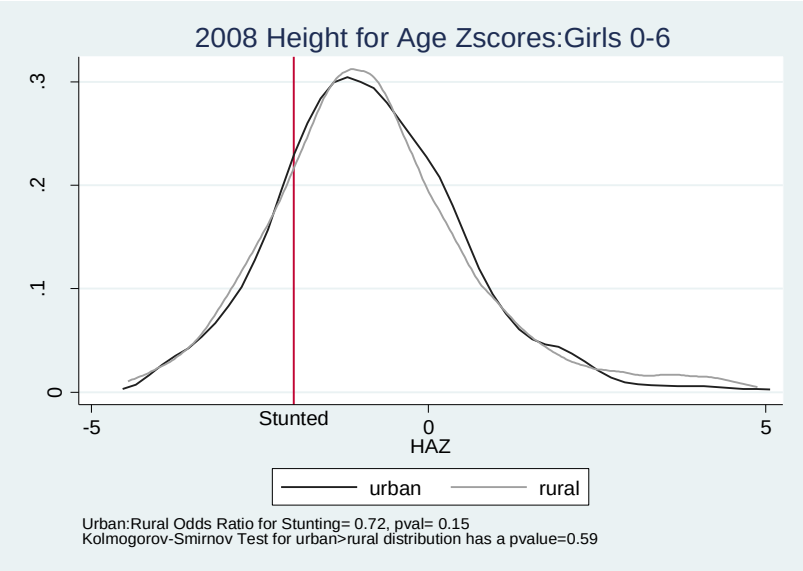


Figure 3.2: Height and Weight for Age among Girls aged 0 to 6:



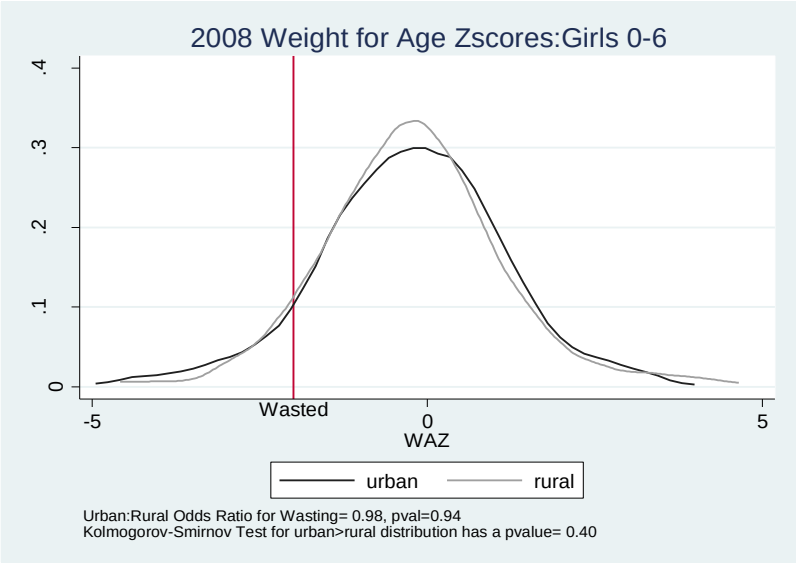
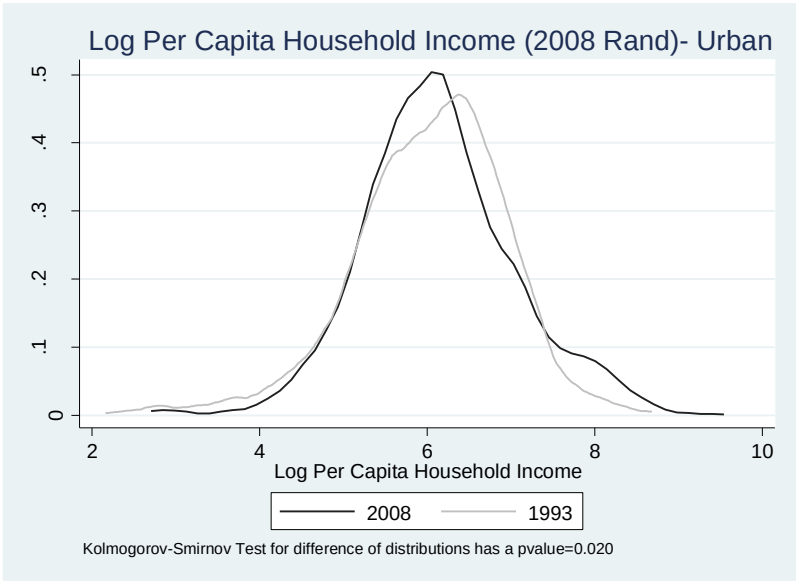


Figure 3.3: Log of Per Capita Household Income in Urban and Rural Households with resident children under 6



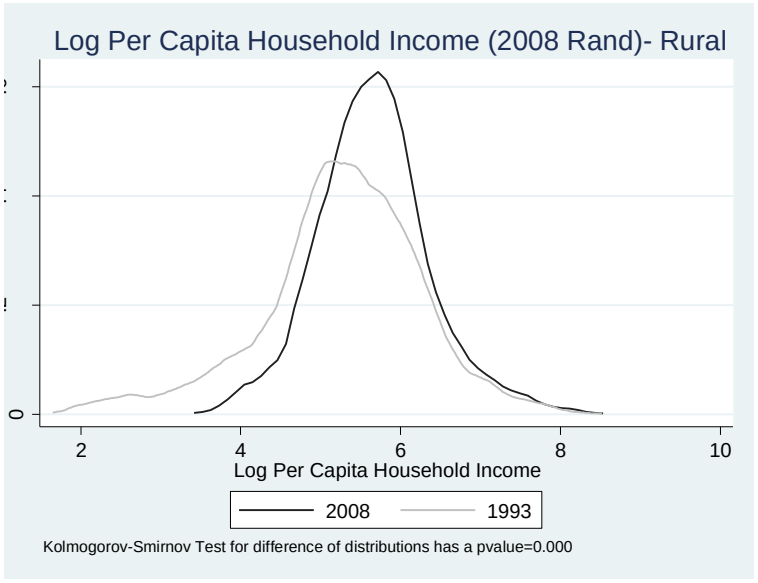
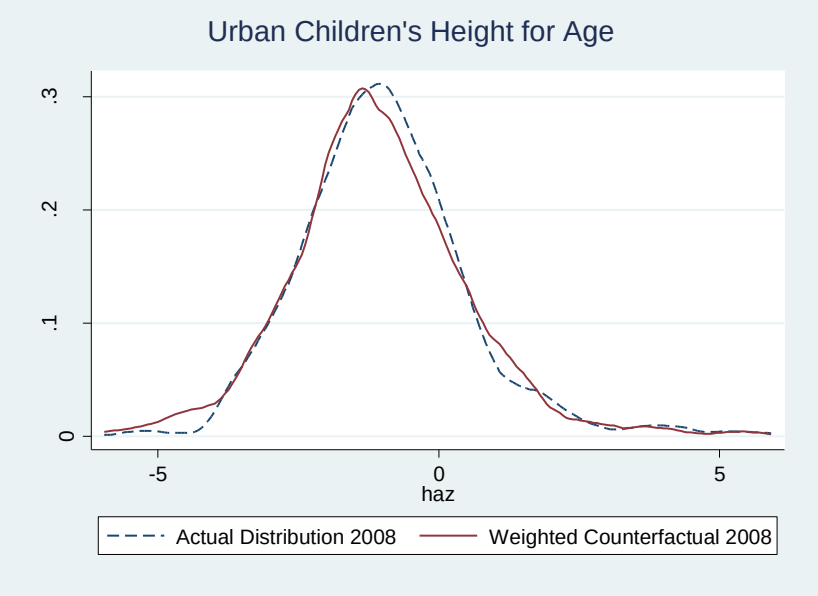


Figure 3.4: Dinardo Fortin Lemeiux Decomposition for Height For Age



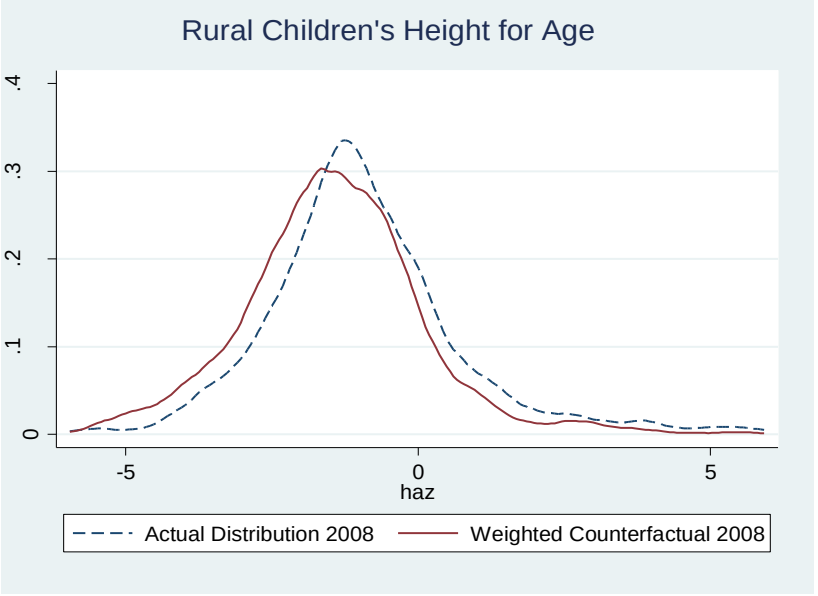
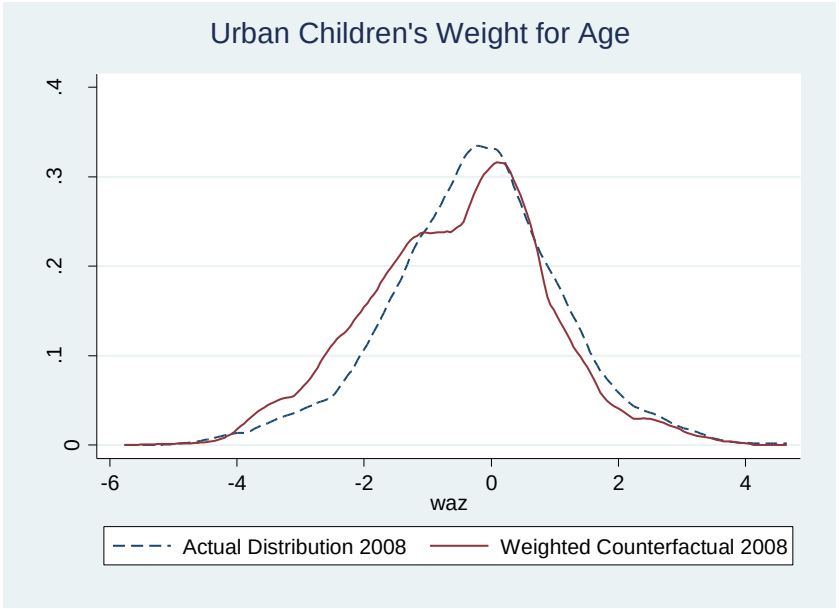
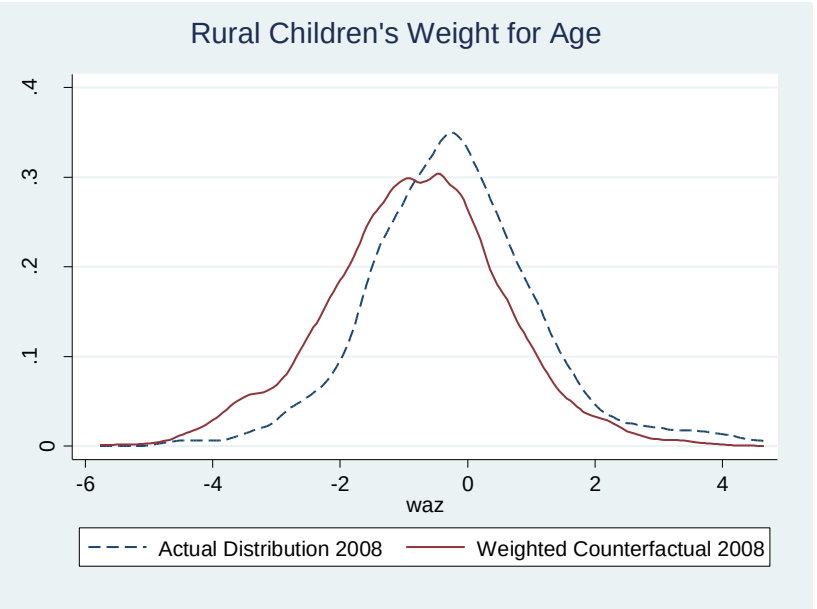


Figure 3.5: Dinardo Fortin Lemeiux Decomposition for Weight For Age





References

Abu Saad, K. and D. Fraser (2010) "Maternal Nutrition and Birth Outcomes" *Epidemiological Review*, Jan 2010, 1-21.

Akbulut-Yuksel, Mevlude, and Belgi Turan. (2011). "Left Behind: Intergenerational Transmission of Human Capital in the Midst of HIV/AIDS." *SSRN eLibrary*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1675637.

Anglewicz, P. (2007). Migration, Risk Perception , and HIV infection in Malawi. January 1. http://repository.upenn.edu/dissertations/AAI3292004/.

Barongo, LR, MW Borgdorff, FF Mosha, A Nicoll, H Grosskurth, KP Senkoro, JN Newell, J Changalucha, AH Klokke, and JZ Killewo. (1992). "The epidemiology of HIV-1 infection in urban areas, roadside settlements and rural villages in Mwanza Region, Tanzania." *AIDS* 6 (12) (December): 1521-8.

Behrman, J., Kohler, H.-P. & R. Castro., (2010). Perception of HIV risk and the quantity and quality of children: The case of rural Malawi. *PSC Working Paper Series*. Available at: http://repository.upenn.edu/psc_working_papers/23.

Blanc, A, K., and A. A. Way. (1998). "Sexual Behavior and Contraceptive Knowledge and Use among Adolescents in Developing Countries." *Studies in Family Planning* 29 (2) (June 1): 106-116. doi:10.2307/172153.

Boerma, J T, M. Urassa, K. Senkoro, A Klokke, and J Ngwesehemi. (1999). "Spread of HIV infection in a rural area of Tanzania." *AIDS* 13 (10) (July 9): 1233-1240.

Bongaarts, John. (2007). "Late Marriage and the HIV Epidemic in Sub-Saharan Africa." *Population Studies* 61 (1) (March 1): 73-83.

Bongaarts, John. (2007). "Late Marriage and the HIV Epidemic in Sub-Saharan Africa." *Population Studies* 61 (1) (March 1): 73-83.

Brockerhoff, M. (1994) "The impact of Rural-Urban Migration on Child Survival" *Health Transition Review* 4:127-149.

Brockerhoff, M. (1995) "Child Survival in Big Cities: The Disadvantages of Migrants" *Social Science and Medicine* v.40, no 10. 1371-1383.

Brockerhoff, M. and E. Brennan. (1998) "The Poverty of Cities in the Developing World." *Population and Development Review* 24:75-114.

Caldwell , J. "Education as a Factor n Mortality Decline: An Examination of Nigerian Data." *Population Studies*, 33, 395-413.

Caldwell, J., and P. Caldwell. (1993). "The Nature and Limits of the sub-Saharan African AIDS Epidemic: Evidence from Geographic and Other Patterns." *Population and Development Review* 19 (4) (December 1): 817-848. doi:10.2307/2938417.

Case, A and C. Paxson. (2009) "Early Life Health and Cogntive Functioning in Old Age" American Economic Review: Papers and Proceedings 2009, v. 99, issue 2, pp 104-109

Case, A and C. Paxson. (2009) "The Impact of the AIDS Pandemic on Health Services in Africa: Evidence from the Demographic and Health Surveys" Forthcoming in *Demography.* Available at http://www.princeton.edu/~accase/papers.html

Case, A and C. Paxson. (2010) "Causes and Consequences of Early Life Health." *Demography*, vol. 47, supplement 2010, pp. S65-S85

Case, A and C. Paxson. (2010) "The Long Reach of Childhood Health and Circumstance: Evidence from Whitehall II study." NBER Working Paper: 15640

Case, A., and C. Paxson. (2011). The Changing Role of Education on the AIDS Crisis in Africa. Mimeo. Princeton.

Case, A. & C. Paxson (2011). The Impact of the AIDS Pandemic on Health Services in Africa: Evidence from Demographic and Health Surveys. *Demography*, 48(2), pp.675-697.

Clark, S J., Collinson, M.A., Kahn, K., Drullinger, K., and S. M. Tollman. (2007). "Returning home to die: Circular labour migration and mortality in South Africa 1." *Scandinavian Journal of Public Health* 35 (69 suppl): 35 -44. doi:10.1080/14034950701355619.

Clark, S., Bruce, J., and A. Dude. (2006). "Protecting Young Women from HIV/AIDS: The Case against Child and Adolescent Marriage." *International Family Planning Perspectives* 32 (2) (June 1): 79-88.

Clark, Shelley. (2004). "Early Marriage and HIV Risks in Sub- Saharan Africa." *Studies in Family Planning* 35 (3) (September 1): 149-160. doi:10.1111/j.1728-4465.2004.00019.x.

Cleland, J.; Bicego, G.; Fegan, G.(1992) "Socioeconomic Inequalities in Childhood Mortality" *Health Transition Review*, 2, 91-95.

Cleland, J.;and J. Van Ginneken, (1988) "Maternal Education and Child Survival in Developing Coutnries: The Search for Parthways for Influence" *Social Science and Medicine*, 30, 33-43.

Cohen, B. (2006) "Urbanization in Developing Countries: Current trends, future projections, and key challenges for sustainability." *Technology in Society*, 28, 1-2, 63-80, Elsevier.

Crimmins, E.M. & Saito, Y., (2001). Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. *Social Science & Medicine*, 52(11), pp.1629-1641.

Cutler, D.M. & A. Lleras-Muney. (2006). Education and Health: Evaluating Theories and Evidence. *National Bureau of Economic Research Working Paper Series*, No. 12352. Available at: http://www.nber.org/papers/w12352 [Accessed March 31, 2011].

De Walque, D. (2006). *Discordant couples : HIV infection among couples in Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania*. The World Bank, June. http://ideas.repec.org/p/wbk/wbrwps/3956.html.

De Walque, D. (2007). "How does the impact of an HIV/AIDS information campaign vary with educational attainment? Evidence from rural Uganda." *Journal of Development Economics* 84 (2) (November): 686-714

De Walque, D. (2009). "Does Education Affect HIV Status? Evidence from five African Countries." *The World Bank Economic Review* 23 (2) (January 1): 209 -233..

DiNardo, J, Fortin, N. and T. Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, Econometric Society, vol. 64(5), pages 1001-44

Feldman, J.J. et al., (1989). National trends in educational differentials in mortality. *American Journal of Epidemiology*, 129(5), pp.919 -933.

Fortin, N., Lemieux, T and S. Firpo (2010) "Decomposition Methods n Economics" NBER Working Paper No. 16045. June 2010

Fortson, J. (2008). "The gradient in sub-Saharan Africa: socioeconomic status and HIV/AIDS."

Fortson, J., (2009). HIV/AIDS and fertility. *American Economic Journal: Applied Economics*, 1(3), pp.170–194.

Fortson, J., 2011. Mortality risk and human capital investment: The impact of HIV/AIDS in sub-Saharan Africa. *The Review of Economics and Statistics*, 93(1), pp.1-15.

Fotso, J. (2007) "Urban Rural Differentials in Child Malnutrition: Trends and Socioeconomic Correlates in sub-Saharan Africa" *Health and Place* 13, 205-223.

Fylkesnes, K, R. Musonda, K. Mubanga, Z Ndhlovu, F Mulanda, L Kaetano, and C. Chipaila. 1997. "The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women." *AIDS* 11 (3) (March 11): 339-345.

Gersovitz, M, Hanan G. Jacoby, F. Seri Dedy, and A. Goze Tape. 1998. "The Balance of Self-Reported Heterosexual Activity in KAP Surveys and the AIDS Epidemic in Africa." *Journal of the American Statistical Association* 93 (443): 875-883.

Gersovitz, Mark. 2005. "The HIV Epidemic in Four African Countries Seen through the Demographic and Health Surveys." *Journal of African Economies* 14 (2) (June): 191 -246.

Glick, P., and D. Sahn. 2007a. Are Africans Practicing Safer Sex: Evidence from Demographic and Health Surveys in Eight Countries. USAID: Strategies and Analysis for Growth and Access.

Glick, P., and D. Sahn.. 2007b. "Changes in HIV/AIDS knowledge and testing behavior in Africa: how much and for whom?" *Journal of Population Economics* 20 (2) (April 1): 383-422.

Gregson, S., H. Waddell, and S. Chandiwana. 2001. "School education and HIV control in sub-Saharan Africa: from discord to harmony?" *Journal of International Development* 13 (4): 467–485.

Gummerson, E. 2011. From predictive to protective? The changing relationship of HIV and Education in sub-Saharan Africa. August 2011

Hargreaves, J.R. & Glynn, J.R., 2002. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medicine & International Health*, 7(6), pp.489-498.

Hargreaves, J R, and J.R. Glynn. 2002. "Educational attainment and HIV-1 infection in developing countries: a systematic review." *Tropical Medicine & International Health* 7 (6) (June 1).

Hargreaves, James R, Christopher P Bonell, Tania Boler, Delia Boccia, Isolde Birdthistle, Adam Fletcher, Paul M Pronyk, and Judith R Glynn. 2008. "Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa." *AIDS* 22 (3) (January): 403-414.

Health Systems Trust, (2008) *South African Health Review 2008*.

Hobcraft, J (1993) "Women's Education, Child Welfare and Child Survival: A Review of Evidence." *Health Transition Review*, 3, 159-175.

Human Sciences Research Council, (2006) *South African National HIV Prevalence, HIV Incidence, Behavior and Communication Survey*, 2005 (Cape Town, South Africa, HSRC Press)

Leibbrandt M, Levinsohn, J. and J. McCrary (2010) "Incomes in South Africa after the Fall of Apartheid *Journal of Globalization and Development*, Volume 1, Issue 1.

Mishra, V. et al., (2006). HIV Testing in National Population Based Surveys: Experience from the Demographic and Health Surveys. *Bulletin of the World Health Organization*, 84(7), pp.537-545.

Montgomery, M. (2008) "The Urban Transformation of the Developing World" Science, 319, 761-764.

Montgomery, M. (2009) "Urban Poverty and Health in Developing Countries," *Population Bulletin* 64, no 2.

Oster, E. (2009). "HIV and behavior change: Why not Africa." *Unpublished manuscript*.

Over, M., and P. Piot. (1993). "HIV infection and sexually transmitted diseases." *Disease control priorities in developing countries*.

Panel on Urban Population Dynamics, (2003) *Cities Transformed: Demographic Change and its Implications in the Developing World*, ed. Mark. R. Montgomery et al. (Washington DC, National Academies Press,)

Preston, S.H. & I.T. Elo. (1995). Are Educational Differentials in Adult Mortality Increasing in the United States? *Journal of Aging and Health*, 7(4), pp.476 -496.

Reniers, G. (2008). "Marital Strategies for Regulating Exposure to HIV." *Demography* 45 (2): 417-438.

Ross, C.E. & Wu, C.-ling, (1995). The Links Between Education and Health. *American Sociological Review*, 60(5), pp.719-745.

Singh, S. & Samara, R., (1996). Early Marriage Among Women in Developing Countries. *International Family Planning Perspectives*, 22(4), pp.148-175.

South African Institute of Race Relations. *South Africa Survey 2008/2009*. (Johannesburg, South African Institute of Race Relations, 2009.)

Statistics South Africa. (2003) "Census 2001 :Investigation Into Appropriate Definitions of Urban and Rural Areas for South Africa"

Tienda, M. Findley S. Tollman S. and E. Preston-Whyte (2006) *Africa On The Move: African Migration and Urbanization in Comparative Perspective,* (Johannesburg, Wits University Press)

UNAIDS. (1998). 1998 Report on the Global AIDS Epidemic. World Health Organization, June 25.

UNAIDS.(2010). Report on the Global AIDS Epidemic. World Health Organization.

Van de Poel, E. O'Donnel, O. and Doorslaer, E. (2007) "Are Urban Children Really Healthier? Evidence from 47 Developing Countries" *Social Science and Medicine* 65, 1986-2003

Van de Poel, E. O'Donnel, O. and E. Doorslaer (2009) "What Explains the Rural-Urban gap in Infant Mortality:Household or Community Characteristics" *Demography* 46: (825-850)

Zaba, B. (2004). "Age at first sex: understanding recent trends in African demographic surveys." *Sexually Transmitted Infections* 80 (suppl_2) (December): ii28-ii35.