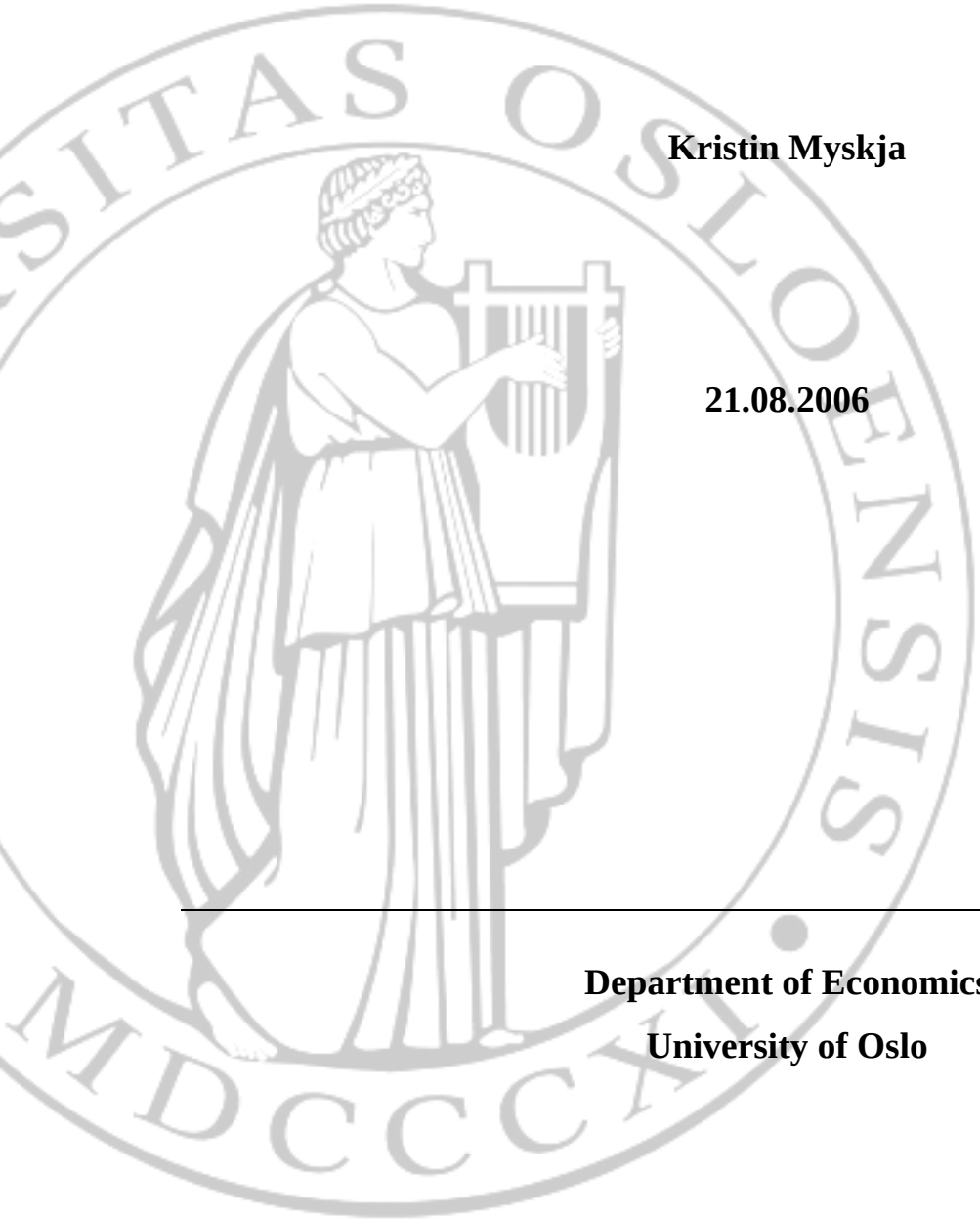
Master thesis for the Master of Philosophy in Economics degree

HIV and Education

a multivariate analysis on macrodata from sub-Saharan Africa



Kristin Myskja

21.08.2006

Department of Economics

University of Oslo

Preface

My motivation for studying economics in the first place was a desire to understand the mechanisms that matters for how the world works. During my study I have realized that this is not possible. What I have discovered is that it is possible to understand quite a lot on many fields that I never thought could be under the economics umbrella.

First I will express my uttermost gratitude to postdoc Kåre Bævre for precise feedback and excellent supervision. I will thank Ingvar Theo Olsen and Sigrun Møgedal for introducing me to development and HIV-topics in the first place. In the final phase Hans Olav Myskja, Kjetil Myskja and Camilla Kristiansen have read through and each contributed with their corrections and comments. Thanks a lot!

Finally, I will thank my parents for their encouragement, patience and for always supporting me, no matter what.

Kristin Myskja

Oslo, August 18th 2006

Content

1. Introduction

Around the world there are approximately 38.6 million people living with HIV (UNAIDS, 2006). The most devastating numbers are found in sub-Saharan Africa, but the disease is also escalating in parts of Eastern Europe and in Central and East Asia. In just 25 years the disease has turned into one of the worst epidemics that mankind has seen.

To be able to halt the HIV-epidemic it is necessary to identify its causes and implications. In sub-Saharan Africa the epidemic is not limited to marginalized groups, but spreads mainly through heterosexual intercourse in the general population. There is a complex web of effects between demographic and socio-economic factors and the HIV-epidemic. The existing literature has paid much attention to education, but also migration and different indicators that picture aspects of the culture can affect the level of HIV-infection.

Several studies have actually shown a positive correlation between higher education level and higher HIV-prevalence on the regional level in sub-Saharan Africa (Crampin, 2003; Abebe, 2003; Hargreaves and Glynn, 2002; Fylkesnes, 1997). Gregson (2001) have shown the same on national level. In his view education "also serves as a useful proxy for the wider effects of socio-economic development". The education effect in his article can contain effects of socio-economic variables not specifically defined. This can be problematic in a methodical sense because the interaction between education and these socio-economic variables can cause biases in the results.

The nature of this positive relationship is intriguing. Education choices are often influenced by cultural aspects and traditions. It is common sense that knowledge will make a person better equipped to prevent oneself from infection. But culture and traditions can hinder necessary precautions to be taken, and in turn reduce the effect of education. Contrary to the common sense-argument above it is shown that higher education actually increases the risk of HIV-infection. This paradox has some problematic implications for politics on the field and for prevention efforts in particular.

Since the results from previous research are so startling it will be natural to go behind the scenes and see if there are other factors that can affect the level of HIV-infection, either in addition to, or hidden in, the education effect. It is, however, natural to use education as departure point since this is the most common explanatory variable. My approach will be an econometric analysis on country level for sub-Saharan Africa. I will use two datasets. One contains primary education as main explanatory variable and the other contains secondary education. The first dataset has 31 observations whereas the second has 25. The other variables are similar in both datasets. I have used PcGive in GiveWin version 2.10 to perform the econometric analysis.

The variables in the datasets are chosen mainly based on research done on regional level. It will be interesting to see if the same mechanisms and effects can be identified on macro level. Education is probably correlated with other variables and there can be biases in the results unless these other variables are included. If there are signs of biases this will confirm the suspicion that it is too simple to focus only on education when examining explanations for variation in the level of HIV-infection.

The outline of my thesis is as follows. I will start in chapter two with a survey over the HIV/AIDS epidemic, in medical and geographical terms. This is relevant to get an overview over the problems and implications of the disease and to place the thesis into an empirical context. Chapter three will have a more theoretical approach. I will sketch the different mechanisms that are believed to be related to the level of HIV-infection. It will be relevant to discuss it in both static and dynamic terms because the analysis will be static, but there are temporal dynamics that must be taken into account in analysing the results. Chapter four will describe the variables and data that are natural to use in an analysis following the discussion in chapter three. The analysis and the main results will be presented in chapter five and I will discuss the implications of my results and conclude in chapter six.

2. Status of the HIV/AIDS epidemic

2.1 The disease and public health concerns

Human Immunodeficiency Virus (HIV) infects the cells of the human immune system and destroys their functions. A person with HIV develops Acquired Immunodeficiency Syndrome (AIDS) when the level of HIVs in the body is so large that the immune system cannot fight infections and symptoms associated with deficiency of the immune system (UNAIDS, 2006c). A person with HIV that does not receive treatment normally develops AIDS within 8-10 years so the time horizon from contamination to the point where the surroundings can observe the disease is very long.

A direct cause of AIDS is that the prevalence rates for other related diseases increases. Since the immune system is more or less destroyed the body is not capable of fighting viruses that a normal person will not notice or at least be able to fight. Tuberculosis is a typical example of a disease that is on the rise in areas with high HIV prevalence.

The first incidence of AIDS was identified by U.S. doctors in 1982. Three years later they discovered that HIV was the cause of the disease. The first cases were found in the gay community in San Francisco, and the disease was therefore believed to be related to gay sexual practice. One of the first names attached to the syndrome was actually Gay Related Immunodeficiency Syndrome (GRID). Dr. Samuel Broder was working in the National Cancer Institute in the United States of America at the time and in a 1997 interview on the National Health Institute response to AIDS he said that after examining the first patient

"Basically, all I remember saying was that we had never seen anything like this before, and I hope we never see anything like this again." (Harden, 1997)

HIV is transmitted through human body fluids. This can be through sexual intercourse, through using contaminated needles in health care or drug injection, and it can be transmitted from mother to infant. It is difficult to assign probabilities to the different ways of transmission, but it is fairly easy to avoid infection. Condom use, use of clean needles and avoiding breastfeeding when infected are ways to halt the epidemic.

Dr. Broder also implies that taboos related to the lifestyle of the infected patients, like homophobia, were an issue for the doctors doing research in this field. In 1983 the first heterosexual epidemic was disclosed in Central Africa. From this point on, all indicators pointed upwards; prevalence rates shot up, incidence rates rose, and death rates increased.

There is no treatment for HIV, but it is possible to postpone the development of AIDS by using antiretroviral treatment. This treatment is relatively expensive. Availability, especially for the poorest, is not very good because of financial and logistical obstacles. In the first 20 years of the epidemic researchers and health workers focused on prevention of transmission rather than treatment. This was not a very beneficial strategy, for many reasons. One is that as long as there is no treatment, people have no incentives to go and get tested for the disease, so it is impossible to know how prevalent it really is.

In the 1980s research on HIV and AIDS was concentrated on the disease itself, how to treat it, and patterns of transmission. The first therapy, azidothymidine, was approved for use in the US in 1987 (UNAIDS, 2006b). Even though the epidemic was on the rise in Africa, the international society had not realized the severity yet, and the research was concentrated on the western world.

The tide was starting to turn at the end of the 80s and the beginning of the 1990s. Paula A. Treichler (1991) criticized researchers at the International AIDS conferences for being too homogenous and western, and she challenged the prejudices towards the black man and HIV-victims all over the world. Ann Larson (1990) also underlined migration and socio-cultural circumstances as decisive for the patterns in which the epidemic spread. During the 1990s people's awareness of the epidemic increased, and research was widened and strengthened. The socioeconomic effects of the epidemic came more in focus. Antiretroviral therapy was launched and incorporated in public health programs in Brazil in 1996 and in Kampala and Abidjan in 1997 (UNAIDS, 2006b). In medical and epidemiological research the focus was then on treatment and monitoring. More reliable tests are developed, and antiretroviral therapy was a huge leap forward towards finding a treatment for AIDS. The next steps were to find ways to make this available to the masses. In the 1990s antiretroviral therapy was mainly for the rich. Even though the therapy was a part of public health programs in some countries in the last half of the decade, the vast majority of patients did not have access to this kind of treatment.

After the AIDS conference in Durban in 2000 the focus shifted from prevention to treatment, and a process began to make antiretroviral treatment more accessible, particularly in Africa which was, and still is hardest hit. The process terminated in the "3 by 5" initiative by WHO. This initiative established the aim to provide treatment for 3 million infected patients in Africa by December 2005. In December 2005 they had provided treatment for 1.3 million. This is undoubtedly somewhat lower than 3 million, but a good improvement from the 400,000 patients treated when the initiative began (The Economist, 2006).

The "3 by 5" inititative is the closest one has come to reaching out with treatment to everyone, but according to the WHO one still has a long way to go (WHO, 2005). Furthermore, there is evidence that a higher availability of treatment increases the number of people who want to be tested for HIV and hence the estimates of prevalence and incidence will be more correct (WHO, 2005).

The last 5-6 years there has been more focus on the importance of prevention and information. Caldwell (2000) argues that "the central plank in the victory over AIDS is the recognition by African governments of social and sexual reality". He also argues that the most effective way will be for risk groups to form their own organizations and leadership, both for distributing condoms and for providing information. Further he concludes:

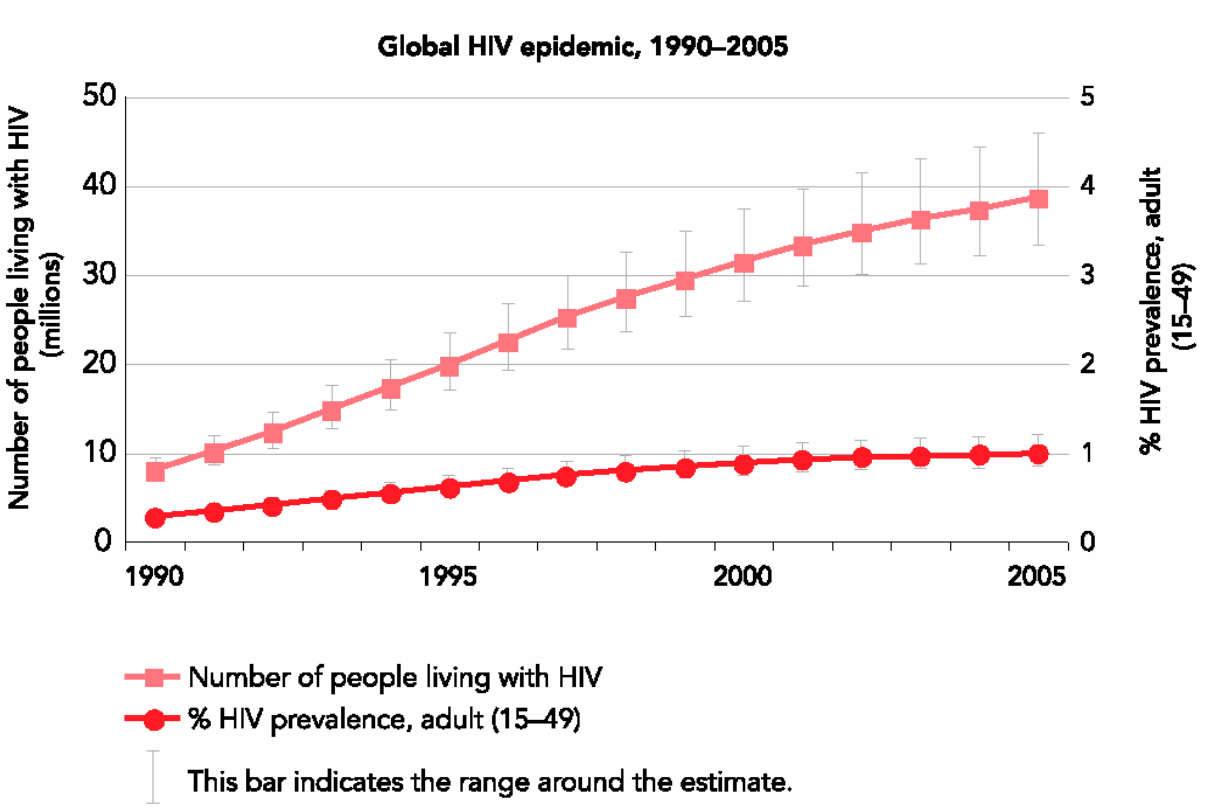
"Finally, strong informational programs must continue to point out the reduction in the risk of AIDS from changed sexual behaviour, but this should not be presented as the only option and vigorous efforts will be needed to make the other options as risk-free as possible"

2.2 Characteristics of the epidemic

Prevalence describes the percentage of the population infected with HIV, while incidence measures the rate at which new infections are occurring. Figure 2.1 shows the development of HIV-prevalence and adults living with HIV in the world from 1990-2005. The estimated number of people living with HIV has increased gradually over the whole period, but the growth rate seems to have decreased somewhat the last five years. The range around the estimates is larger now than they were in the beginning, in relative terms. Even though the numbers are much larger, one should expect the estimates for the last years to be more certain than in the beginning of the 1990s due to better estimation methods. In 1990 the estimated

number of adults living with HIV was lower than 10 millions. For 2005 the estimate is 38.6 millions with a confidence interval of 33.4 - 46 millions. This is close to a fourfold increase. HIV-prevalence, in percentage of adult population, has not increased that much. From approximately 0.3 % of world adult population infected by HIV in 1990 it has grown to 1 %. The growth in prevalence has levelled out the last 3-4 years, but the ranges around the estimates show that there still are uncertainties to which way the arrow points. I will come back to a more thorough discussion of the data in section 4.2.

Figure 2.1 – Estimation of number of persons with HIV and the HIV prevalence in adult (15-49 years old) world population



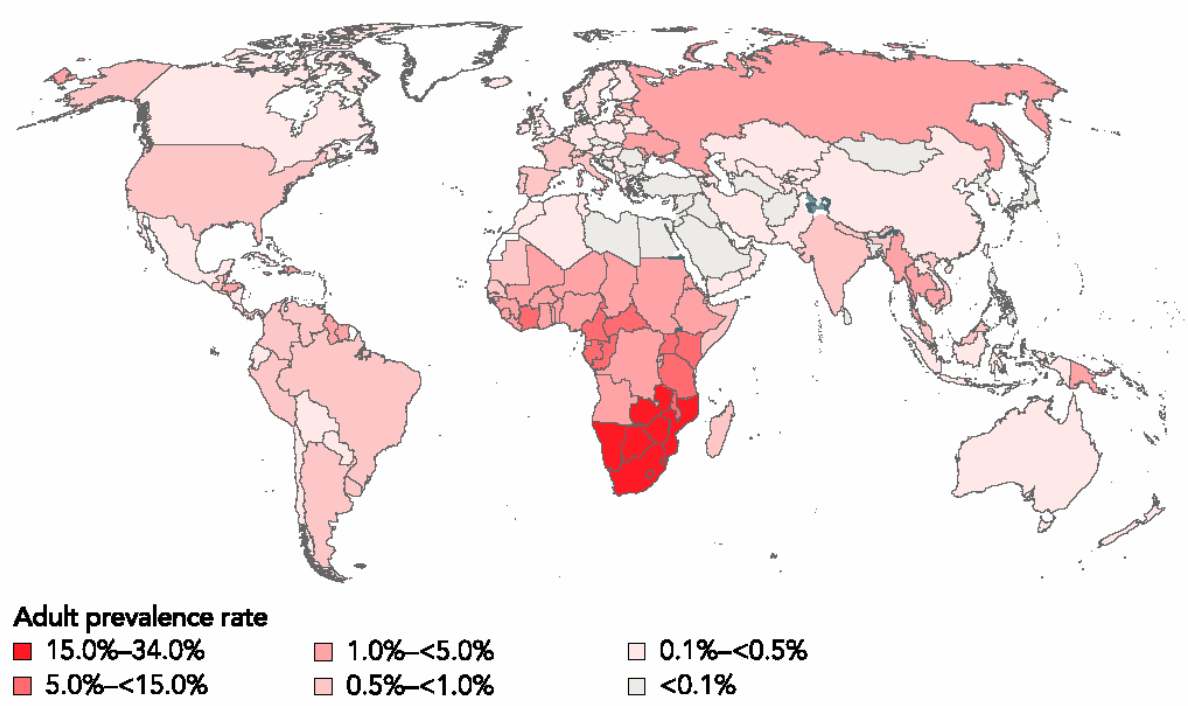
Source: UNAIDS, 2006a

Figure 2.2 shows the geographical spread of the epidemic. Southern Africa is the only region with higher prevalence than 15%. Central Africa, Eastern Europe, South-East Asia and Latin America are regions with countries that have a higher prevalence than 5%. The region least affected is Northern Africa and the Middle East.

2.2.1 The World

Traditionally HIV/AIDS is spread in groups with high-risk behaviour. Men who have sex with men, injecting drug users and the commercial sex industry are focus groups of the epidemic all over the world. Still, the epidemic is transmitted in various ways in different cultures. I will now present a brief overview of the different characteristics of the epidemic over the world.

Figure 2.2 – A global view of HIV infection



Source: UNAIDS, 2006a

Apart from Africa, Asia is the continent that has the highest numbers. Estimates show that 8.3 million people are living with HIV in Asia in 2005, 1.1 million out of which became infected during the past year. China, India and the countries in South-East Asia have the highest prevalence rates. Nevertheless, Thailand is the only country with prevalence above 1%. HIV-positives are mostly injecting drug users, commercial sex workers or men who have sex with men. But the epidemic is in transition, and there are, especially in some areas, reasons to believe that the epidemic will spread to other parts of the population unless efforts are made to stop it.

Stigmatisation and discrimination are huge problems in Asia, as in the other regions. HIV-testing and counselling are not much used and will continue to be rare unless this stigmatisation and discrimination is reduced and people's awareness and knowledge are increased.

The Caribbean also experiences a severe epidemic. The epidemic here is very similar to the one in sub-Saharan Africa. Unprotected heterosexual intercourse is the main mode of transmission, and this happens in an environment of deep impoverishment and gender inequalities. In percentage it is the second most affected region in the world (UNAIDS, 2006a).

In North America there is widespread access to antiretroviral therapy, but immigrants and migrants are often not reached by public services. Prevention, diagnosis and treatment services need to be adapted to the shifts in the epidemic, and especially need to be shifted towards immigrants, ethnic minority groups and men who have sex with men (UNAIDS, 2006a). Again, we find that stigmatisation and discrimination are problems, especially since the transmission of the disease is associated with risky sexual behaviour, often considered morally reprehensible. Several studies have reported "resurgent risk behaviour" among men who have sex with men (UNAIDS, 2006a). There is a disproportion of HIV infections among African Americans and Hispanic Americans. The driving force of the epidemic is unprotected sex and injecting drug use.

In Western, Central and Northern Europe the epidemic has traditionally been limited to men who have sex with men and injecting drug users. Recent developments are, however, that people originating from regions with severe epidemics, mainly sub-Saharan Africa, spread the virus through heterosexual intercourse. Next to this, unsafe sex between men is the most important mode of transmission. Another risk group, injecting drug users, has experienced a decrease in HIV diagnoses, mainly due to harm reduction programmes (UNAIDS, 2006a).

Eastern Europe is severely hit by the epidemic. The countries that are hardest hit are Ukraine and the Russian Federation, but epidemics are on the rise in several of the other former Soviet republics. The starting point for the epidemic in the Russian Federation was the socio-economic environment in the 1990s where drugs flourished and drug use escalated. Risky behaviour related to injecting drug use is by far the single most important mode of

transmission in Russia (UNAIDS, 2006a). Male injecting drug users are spreading the virus to their sexual partners and so more women become infected. There has been a surge in the estimated number of adult women living with HIV from 2003 to 2005.

As in Russia, the driving forces of Ukraine's epidemic are injecting drug use and unprotected sex. Financing treatment is a big hurdle to jump, both for Russia and Ukraine. Prevention efforts, especially among injecting drug users, are also important to be able to hinder further spread of the virus.

2.2.2 Sub-Saharan Africa

The epidemic in sub-Saharan Africa is different from the rest of the world. Even though there are similar epidemics in the Caribbean, the magnitude of the epidemic in Sub-Saharan Africa causes a more severe impact on society. In the early stages it was highly educated people that travelled a lot who got infected. There was a higher prevalence along the main trading routes of the continent. Seasonal workers that stayed away from home for months at a time had a higher prevalence than the rest of the population. There are probably many reasons for this, but one is that African men by tradition claim a physical need for sexual intercourse. The commercial sex industry is large, and there are many grey areas in the business. Infidelity is very common too. Women do not have sex when they breast-feed to secure enough spacing between their children. Their husbands have sex with others in these periods, and very few use condoms. Their wives are then exposed to sexual transmitted diseases even though they do not expose themselves through high-risk behaviour. Despite the claimed physical need to have sex, many men do not want to use condom because it affects the pleasure. The common explanations are "No one showers with a raincoat" and "No one eats candy wrapped in paper" (Treichler, 1991).

Traditionally Africa is a continent where witchcraft and superstition is very widespread. There have also been many myths related to HIV/AIDS. One example of a myth is that if you were infected and had sex with a virgin, you would get rid of the virus. This led to increased rates of rapes, particularly of very young girls, and increased the HIV-incidence rates in many areas. Many do not care whether they protect themselves or not because they have a strong deterministic belief. Their life will end when it is meant to end, and their behaviour cannot

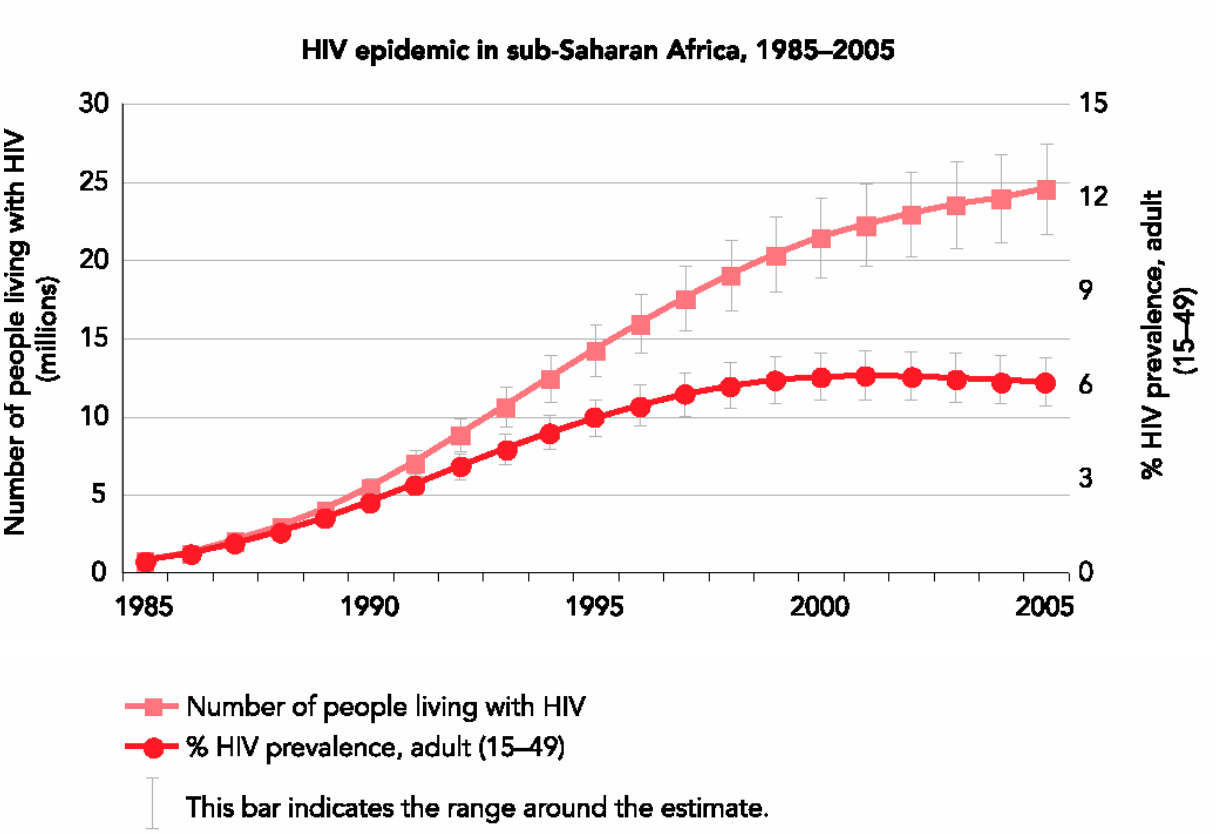
affect this. If they die, they die, and if not they were not meant to die yet. This complicates the planning of prevention efforts.

Another obstacle has been African governments' reluctance to intervene. They will not acknowledge that HIV/AIDS is a problem for their country. Some of them have even insinuated that the disease does not exist and blamed tuberculosis and malaria for the higher death rates. When large generations die, and hundreds of thousands of children are orphans it is vital that the governments see the problem and act. It is possible to look at HIV-prevention as a public good. Because the disease has so fatal effects and is spread so rapidly in these parts of the world countries risk loosing whole generations. One human life is invaluable, but the collective value of a generation for a country is even greater. Hence governments should be concerned with the development and offer treatment. At least they should acknowledge the problem and face the challenges. Studies have shown that for HIV prevention efforts to work they need to be intensive and long-term (UNAIDS, 2005). It is important that treatment and prevention efforts are coordinated and done simultaneously. To achieve the best results, the affected governments need to engage in the issue and devote resources to it.

There are a lot of stigma and discrimination associated with the disease. HIV is in many cases linked to a life-style that is not socially acceptable. Poverty, gender inequality and social marginalization of specific populations are factors that put people at greater risk of HIV. Because of the stigma connected with HIV many people refuse to get tested, and hence finding the ones that need treatment can be very hard. Those who are tested often refuse to admit that they have the disease. In addition it is worth noticing that even when confidentiality is guaranteed many people are reluctant to know their HIV status because they expect severe psychological distress if the result is positive. Stigma and discrimination makes it difficult to control the epidemic and facilitates a good environment for the disease to grow.

Some countries in Sub-Saharan Africa have actually managed to halt the epidemic. Kenya, Uganda and Zimbabwe show declines in HIV prevalence (UNAIDS, 2005). This is due to extensive prevention efforts. Even though the prospects are devastating this shows that it is possible to fight the epidemic as long as it is acknowledged and taken seriously.

Figure 2.3 – Estimation of number of persons with HIV and the HIV prevalence in the adult (15-49 years) population in sub-Saharan Africa[2]



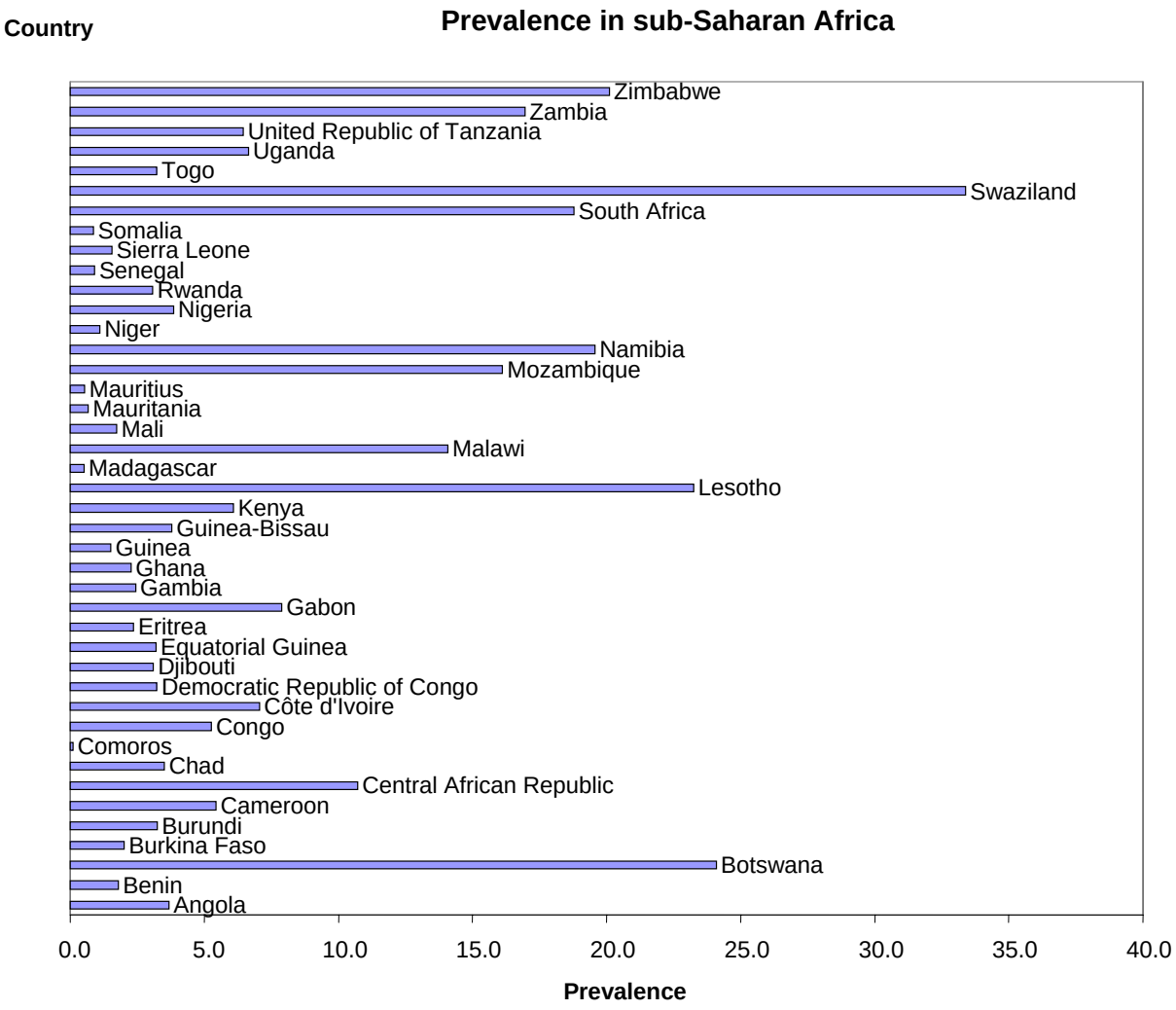
[2]Even though HIV prevalence rates have stabilized in sub-Saharan Africa, the actual number of people infected continues to grow because of population growth. Applying the same prevalence rate to a growing population will result in increasing numbers of people living with HIV.

Source: UNAIDS, 2006

It is worth a remark that data for HIV prevalence actually show a delayed picture and that incidence rates in some cases are a better measure of the development of the epidemic than prevalence. Still the numbers are uncertain and when the epidemic has grown as large as it has in some countries in East and Southern Africa the prevalence can show a grievous equilibrium where approximately the same number of people gets infected as die from the disease. In other words one should be cautious in how one interprets the numbers. The development of the epidemic in sub-Saharan Africa is shown in figure 2.3. HIV prevalence is often based on testing done on pregnant women in antenatal clinics so the risks of overestimation are present since pregnant women by definition have had unprotected sex (Fylkesnes, personal communication). There are also large regional differences inside the countries, and this has to be taken into account when planning strategies for combating the disease.

Figure 2.4 Differences in prevalence across Sub-Saharan Africa

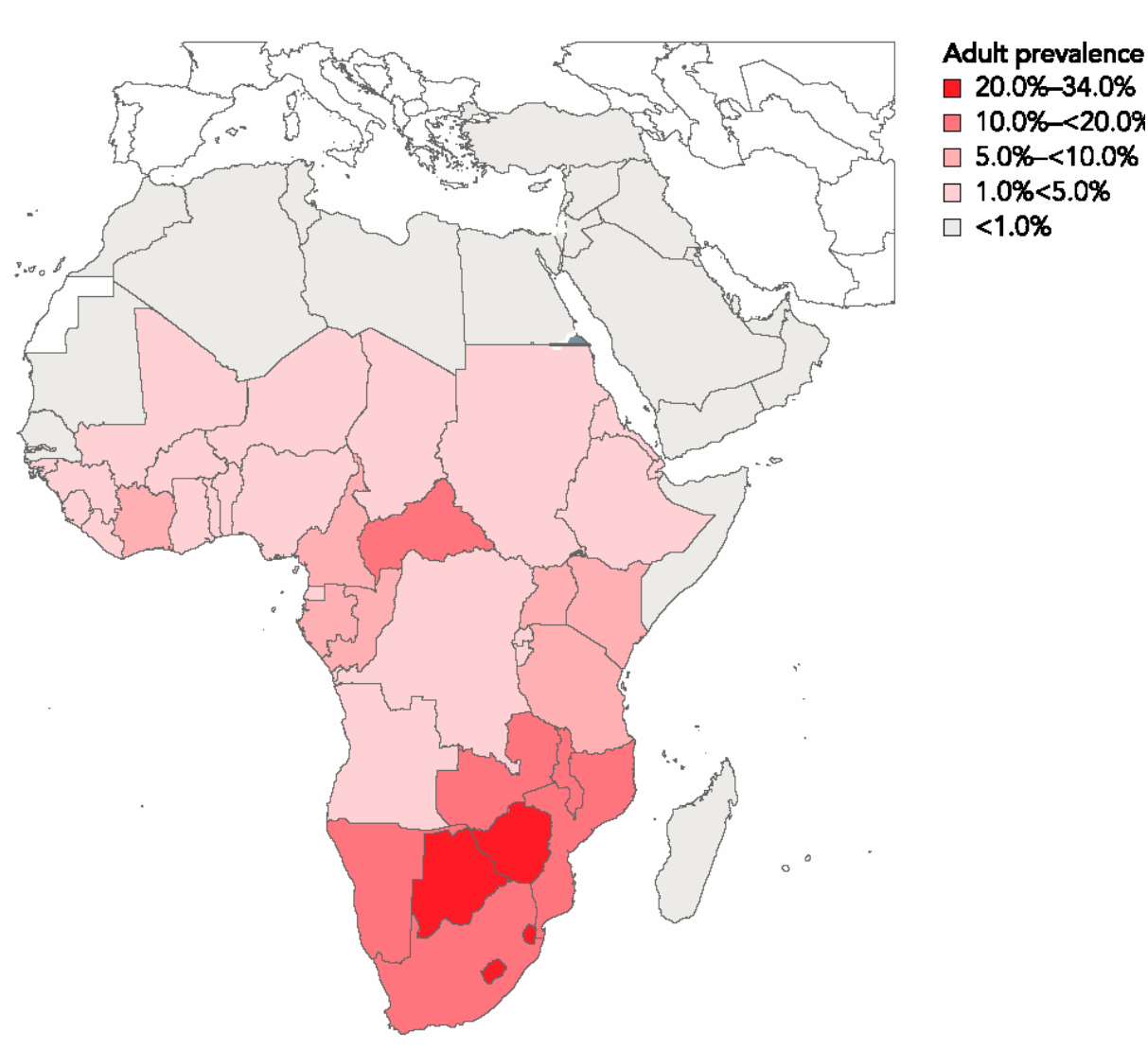
Measures prevalence as a percentage of total population that is infected by HIV, numbers taken from UNAIDS, 2006a.



Observe from figure 2.4 that the highest prevalence is observed in Botswana, Central African Republic, Mozambique, Namibia, Lesotho, Swaziland, South Africa, Malawi, Zambia and Zimbabwe. All these have prevalence higher than 20%. Geographically they lie in a belt stretching from central Africa down towards South Africa, see figure 2.5. There have been observed striking urban-rural differences in the 1990s (Fylkesnes, 1997). Recent research suggests that the urban-rural differences are decreasing and that the epidemic has spread to the rural areas. (Mwaluko et.al, 2003). This underlines the results found by Michelo et.al (2006) that the incidence pattern has changed with respect to education. Another general tendency, as we can see from figure 2.4, is that the countries with lowest prevalence are countries with large Muslim communities.

Sub-Saharan Africa has regional differences in other aspects as well. The heritage from colonial times is for example shown in religion. The countries geographically close to the Arab world are mostly Muslim whereas the countries that were under English or French rule often have large Christian churches. Religion and culture are closely connected to education systems and there are regional patterns with respect to this too.

Figure 2.5– HIV-prevalence (in %) for adults (15-49) in Africa



Source: UNAIDS, 2006a

Changes in sexual behaviour seem to be the most important feature in the countries where HIV prevalence has declined. Increased condom use within casual partnerships and a reduction in reported numbers of sexual partners are the most apparent changes. In some parts

of these countries mortality rates are levelling off. So the HIV incidence is declining and together these factors are driving the decline in prevalence.

HIV/AIDS have a huge impact on mortality in Sub-Saharan Africa. Even though the region has experienced an increase in public health services and improved hygiene average life expectancy has only increased with 2 years from 1970-75 to 2000-05. In the first period it was 46.5, whereas the last years' average life expectancy has been 48 years (UNDP, 2005). In many countries life expectancy has actually fallen for the last 30 years, and this is mainly because of HIV/AIDS.

The progress of making treatment available is unevenly distributed. According to UNAIDS and WHO at least one third of people in need of antiretroviral therapy are receiving it in Botswana and Uganda, while in Cameroon, Cõte d'Ivoire, Kenya, Malawi and Zambia between 10% and 20% of people requiring this kind of treatment where receiving it in mid-2005 (WHO, 2005). At least 85% of South Africans who need therapy were not receiving it and the same goes for 90% or more of those needing it in Ethiopia, Ghana, Lesotho, Mozambique, Nigeria, Tanzania and Zimbabwe.

As we see, the HIV-epidemic is closely linked to different challenges in society. In the countries where we see a decline in prevalence it is due to behavioural change. This shows that it is possible to turn the tide for the other severely hit countries too, but governmental intervention is necessary. Many places socioeconomic dynamics are intensifying the risk for people already in the high-risk group. Poverty reduction and strengthening the public health care are important efforts to diminish the epidemic. Treatment is important, but has to be combined with prevention efforts.

3. Mechanisms at work - education and HIV-prevalence

3.1 Prevalence and incidence

The HIV-epidemic is a dynamic entity. In the early stages prevalence rates can be low, but incidence rates very high. Then prevalence rates increase, and the epidemic roots itself in the population. When the epidemic has lasted long enough for people to start dying, which can be up to 10-15 years, prevalence rates flattens.

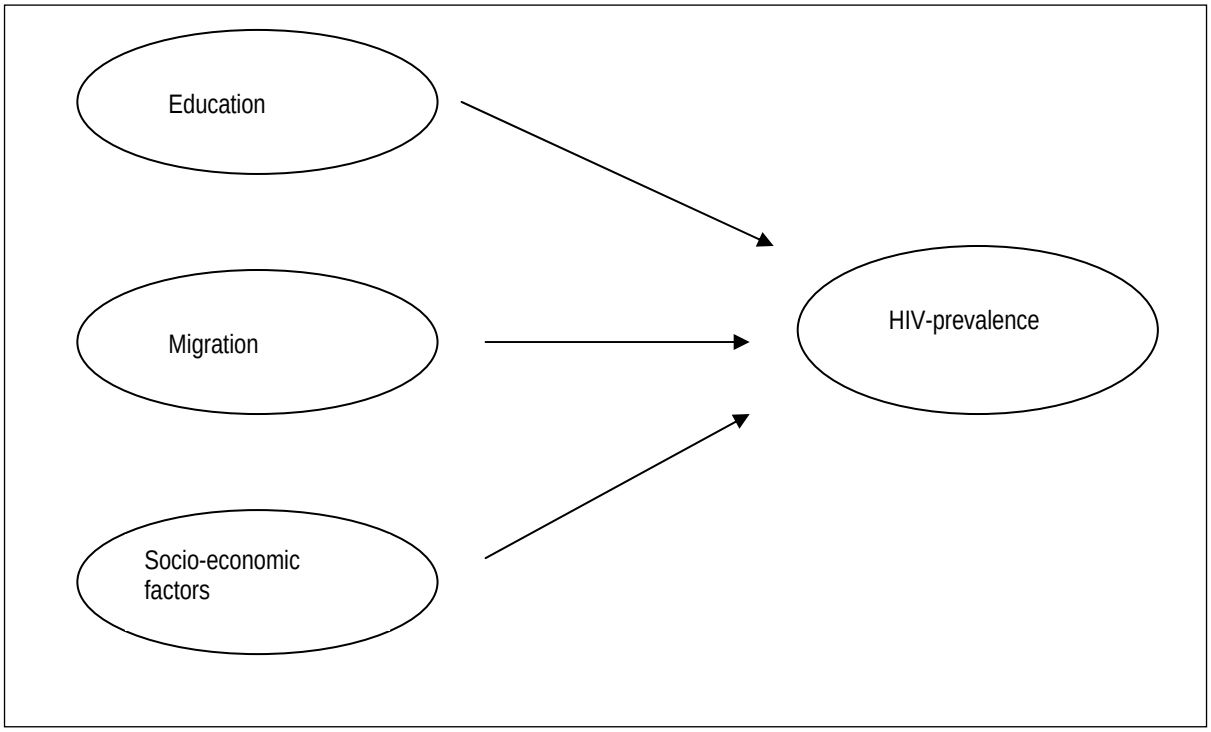
At this stage of the epidemic the number of people that die of AIDS-related conditions can actually exceed the number of people that get infected. The death rates will then be higher than the rate at which new infections are occurring. The number of people living with HIV will decrease and prevalence rates will drop. But incidence rates have not necessarily gone down. In the early stages of the epidemic HIV spreads primarily among those with high-risk behaviour. When they die this group of people is removed from the circuits of HIV transmission, and this could cause HIV incidence to decline and lead to lower prevalence.

This is important to keep in mind. To have a good overview over an epidemic, the best option is to have an overview over both incidence rates and prevalence rates. But prevalence rates alone can give a good enough picture of how deep in the population the epidemic is rooted, so I will concentrate on prevalence rates in the rest of my analysis.

3.2 What matters for HIV-prevalence?

When analysing which variables that affect HIV-prevalence the problem is to separate the effects from different sources. Endogeneity can occur in several ways. We will first examine the three factors usually considered to be the main factors that affect HIV-prevalence. These are education, migration and socio-economic variables, see figure 3.1.

Figure 3.1 Factors that affect HIV-prevalence



Education provides knowledge and makes people better equipped to acquire and use information provided from different sources. Educated women and men are more likely to see the advantages of using health care services and they should know more about hygiene and how to avoid infectious diseases, like HIV. They should also be more aware of aspects of traditions and myths that facilitate the spread of HIV instead of preventing it. But education can lead to higher risk of HIV-infection too. When the education level in a country increases it tends to increase the average age of marriage. This implies that more people engage in high-|risk sex and prevalence will increase.

Another point is how the ratio between male and female attendance rates in education affects HIV-prevalence. In most sub-Saharan countries the mother is the primary care provider of the children in the family. The children's health depends primarily on the mother's knowledge. Therefore female attendant rates can have different effects compared to male attendance rates. Daughters are brought up very traditionally and prepared for a career as care takers and then education is not considered necessary in many cultures. In some cultures children's health even depends primarily on the mother-in-law's education because drawing attention to one's child's sickness before the mother-in-law does is not considered proper behaviour. Besides

this, there is also an economic aspect here. Education is expensive, and families prefer to invest in their sons rather than their daughters.

Male education matters for HIV-prevalence in other ways. Men often have the last word in decision-making processes inside the family and in many cultures they decide how much their wife can decide and whether she can move around freely. In turn, this affects her decision on for instance whether or not to take their child to see health care workers in case of illness. Traditionally, HIV spreads through seasonal workers that engage in extra-marital sex when they are away from their families, and then they bring the infection home to their families. Education is then important to make the husbands aware of the risk to which they expose not only themselves, but also their wives, and to promote the use of condoms.

The community effects of education must be taken into account as well. Studies done on the relationship between education and fertility show that a woman with x years of education living in a village with high average education will have a higher probability of using family planning than a women with the same x years of education living in a village with lower average education (Kravdal, 2002). Since family planning requires behavioural change in a way that can be compared to the behavioural change necessary to avoid being infected by HIV, we can assume that similar community effects will be evident in our study of the relationship between education and HIV-prevalence.

There are other aspects that complicate the relationship between education and HIV-prevalence as well. The most important one is the know-do-gap (Joint Learning Initiative, 2004). There can be a long stretch from knowledge to action. Behavioural changes do not only depend on the information that is provided, either through education or other sources, but also on cultural aspects and especially how the society reacts when confronted with new problems. The know-do-gap has to be narrowed because it has been shown that the most efficient way to reduce incidence rates of HIV is by behavioural change. Education in itself, whether in general or specific education on how to protect oneself from infections, does not reduce HIV-prevalence, but is necessary to make people aware of the behavioural changes that need to take place and thus to narrow the know-do-gap.

Other obstacles can be the lack of access to health care facilities and condoms. Studies have shown that unless health care is being offered close to where people live, education has little

effect (Olshansky, 1997). A mother can know that she should have brought her child to the clinic, but when the clinic is situated a three-hour walk from her village there is a higher probability that she would stay at home than if the clinic is situated in her own village.

The second factor that affects HIV-prevalence is migration. Educated people tend to move to more urban areas, either to get more education or to work in other fields than agriculture. This link between migration and education show the complexity of the interactions between these three main factors. Particularly those with secondary education or higher will migrate from rural to urban areas. Since the HIV-prevalence traditionally have been higher in urban than in rural areas, and since young people moving to urban areas have a higher probability of engaging in high-risk sex, the risk of getting infected in the city has traditionally been higher than in rural areas. When these young people either move back home or spread the disease to their families at home, the prevalence in rural areas rise as well.

The third factor is the socioeconomic characteristics. The know-do-gap already mentioned above is often wide, especially in African cultures. Traditions and myths play major roles and are not easy to change. Religion is also an important constituent here. The main distinction should be made between Muslims and non-Muslims. In sub-Saharan Africa the dominant religions are either Islam or various confessions of the Christian church. Next to this, and often mixed with Christianity is Animism and other traditional religions and beliefs. A study from Ethiopia shows that religion has a significant effect on HIV-prevalence and hence should be included as a control variable (Abebe, 2003). The expected result is that there is a lower prevalence in Muslim countries or countries with a large Muslim community. This is also supported by Gray (2004).

In Islam male circumcision is widespread, and male circumcision is related to a lower risk of HIV-infection (Weiss, 2000; Kelly, 1999). The ritual cleaning after intercourse among Muslims also increases penil hygiene. Christians and animists often mix their faith with superstition and traditional religion and do not have as strict a moral policy. Another important difference is the attitudes towards alcohol. Non-Muslims have a much more liberal attitude towards alcohol consumption and are thus more likely to expose themselves to high-risk sex without condom. In Muslim cultures, girls tend to be strictly protected by their families and they do not have the chance to expose themselves to high-risk sexual behaviour in the same way as boys in the Muslim tradition have.

However Gray (2004) also finds ambiguous results concerning to what extent Muslims in sub-Saharan Africa have a higher moral standard with respect to sexual activity. HIV-contamination is connected to sinful behaviour through what the Muslims refer to as the prism of sin and is thus something that is difficult to discuss. There may also be some degree of underestimation due to lack of testing, especially among females.

Other socio-economic factors that can affect HIV-prevalence relate to how rich a country is and how the HIV-infected are considered in society. Stigma and discrimination towards HIV-infected are evident in many societies and the degree influences to what extent governments see the necessity of prioritizing health care for HIV-infected in particular, and their efforts to prevent the spread of the epidemic. Nevertheless, these factors are closely connected to religion through the cultural aspect and hence religion can work as a good proxy for most socio-economic variables.

3.3 The other way around

However, there are also mechanisms that work so that HIV-prevalence affects education level. When the probability of getting infected by HIV is high, people are less likely to spend time and money on education. Education is an investment in the future, and when life expectancy is low, such an investment may not seem worthwhile. Hence, high prevalence can lead to lower average education level in a population.

Teachers are not spared from the epidemic either. When the prevalence is increasing, so is the prevalence among teachers. The labour force in the educational sector will be reduced, and there will be fewer teachers. This can worsen the quality of the education system and reduce the number of schools, especially in areas where prevalence is high and teachers are scarce, and the average education level can fall because the teachers die from AIDS.

Another aspect that is linked to both education and migration is the education of nurses, physicians and other health care staff. High HIV-prevalence would mean that a country needs a strong health sector, and has incentives to build good education systems for health care workers to avoid loosing whole generations to AIDS. However, educated health care workers

with international degrees have opportunities to move to countries where they get higher salaries, and their country of origin experiences brain-drain. The health care workers that stay in their home country are often the second best, and they have to manage with second best equipment. They have a much higher risk of acquiring infectious diseases and also have a higher risk for HIV-infection compared to someone that does not work in the health sector. This results in lower quality care for infected with limited resources, and lower physician density and health worker density. Education of health care staff is vital, but keeping them at home and in the public sector is also important to get pay-off from the investment in this kind of education. This is another example of the complexity of the effects. Migration effects and socioeconomic factors often work together, and it can be difficult to define where the causality begins.

3.4 The temporal dynamics of explanatory variables

As I have mentioned in the beginning of this chapter, the HIV/AIDS-epidemic is a dynamic entity that changes over time. Epidemiological studies done in Zambia show that there has been a recent change in the socioeconomic characteristics of HIV-infected individuals with respect to educational level (Michelo, 2006). The poor, rural population with low education has the highest incidence rates. This could imply that incidence rates are increasing in countries with a low average level of education and that they are decreasing in countries with a high average level of education. Hence, the education effect can be different now compared to what it was 10 years ago when the epidemic was less mature.

Infant mortality is high in countries with high HIV-prevalence because many pregnant women with HIV pass the virus on to their children. In families where both parents have died from AIDS, infants and small children are left in the care of their older sisters and brothers, and they are in many cases not as good caretakers as their parents would have been. This leads to challenges for governments because the age distribution of the population changes and can affect the composition of the fiscal budgets and other macro indicators. In this way the epidemic not only has huge impact on the individual and community level, but the effects can also be devastating on country level.

What this argument shows, as I have mentioned above, is that HIV-prevalence is the result of a complex web of effects and that these effects change over time. Education, migration and

several socio-economic factors are probably necessary to explain the variation in HIV-prevalence across countries. The studies done on regional level generally do not take the complexity of the effects into account, and should have studied a wider range of explanatory factors. There is, however, a difference between looking at the macro level compared to the regional level. Studies based on surveillance data have a different nature than those done on aggregates and estimates for a whole population and I will now see if the effects that other researchers have found on regional level can be identified on macro level.

4. Data

4.1 Limitations

The first I have to decide is what kind of data to use. In macroeconomic analysis, it is most common to use time-series data. For HIV-prevalence it is impossible to do that kind of analysis because it is only 25 years since the virus first was identified. In addition it is not until recently that the estimates produced have been correct and reliable, so great care needs to be taken when comparing estimates both over time and across countries. The latest estimates have improved, though, so it is possible to use the 2005 and 2003 estimates from the UNAIDS-report for cross-section analysis (UNAIDS, 2006a).

Education is not a volatile variable, and changes in the educational system do not occur often. Time-series comparison of education effects is therefore very difficult, because one would need data for education and prevalence over at least 10-20 years to be able to capture the effects on prevalence of changes in aggregate education on the macro level. When analysing how education can explain the variation in HIV-prevalence the only option left is to use cross-section analysis.

The second limitation is geographical. In econometric analysis the more observations, the better fit we get in the results. But when dealing with empirical data there can be huge variations in the nature of HIV-prevalence and also in educational systems that make comparison across all countries in the world very unstable and little robust. In chapter 2 I have argued that the epidemic in sub-Saharan Africa is mainly a generalized epidemic. Comparing generalized epidemics to concentrated epidemics when the countries are very different in all other senses can be problematic. Sub-Saharan Africa is not a homogenous entity, but the countries have more or less the same history and the same characteristics when it comes to populations and ethnic diversity. In a way we can say that the countries are similar enough to be compared, but different enough to get reliable results. In HIV/AIDS research there has been a focus on this region so a fair amount of recent regional-level research can be found. This is another reason for choosing this particular group of countries. There is also a natural division between sub-Saharan Africa and the countries in northern Africa since the population in North Africa is mainly Arab, whereas in the rest of Africa African ethnic groups dominate.

4.2 HIV-prevalence

Data on prevalence is taken from "2006 Report on the global AIDS epidemic" published by UNAIDS in May 2006 (UNAIDS, 2006). The estimates are produced by UNAIDS/WHO and relate to the status in 2003 and 2005. I mainly use the estimates for 2003 because the data I have on education is from 2002 and hence the time span between estimates for prevalence and the educational level is as narrow as possible.

There are many uncertainties connected to the estimates. The report lists number of people living with HIV/AIDS, confidence ranges of number of people living with HIV/AIDS, a point estimate for the prevalence rate and also a confidence range for the prevalence. In my analysis I will use the point estimate for prevalence because this is comparable across countries, in contrast to number of people living with HIV/AIDS which varies with the population size in a country. A lot has been done to improve estimates and estimation methods the last 5-10 years, and this work is starting to pay off. Earlier estimates often showed an over-estimation, both because they were based on pregnant women who, as stated above, by definition have had unprotected sex and because of other biases in the samples. There are many ways to estimate numbers for HIV-prevalence. One problem may be whether to use population-based surveys or surveys among different population groups. All methods have their advantages and disadvantages and therefore one needs to be aware of what kind of estimation methods are used in each particular case. For a more thorough discussion on different ways to estimate see Feeney (2001).

There can also be systematic biases in the data. Groups of countries that are similar in one sense can have similarities in other variables. These biases can also be systematically correlated with the variables used in the analysis. Systematic biases like these can be hard to identify, but awareness of the possibility that they exist is important.

UNAIDS/WHO separate between a generalized epidemic, where more than 1% of the total population are infected and where the virus mostly spread through heterosexual activity, and a concentrated epidemic, where the epidemic is restricted to mainly high-risk populations like injecting drug users, commercial sex workers, and men who have sex with men. In a generalized epidemic HIV-estimates are mostly based on surveillance among pregnant women seeking antenatal clinics (ANC). Estimates from population-based surveys that

include testing for HIV-antibodies are often the most reliable, but studies done on regional level in several countries shows that ANC-results provides a good proxy where population-based surveys are absent (UNAIDS, 2006a). In concentrated epidemics, HIV-estimates are based on surveillance of key populations most at risk.

In the UNAIDS/WHO dataset which I will use, the sub-Saharan Africa section consists of 44 countries. Of these 42 countries has data on estimated prevalence. Data is missing for Liberia and Ethiopia. The UNAIDS/WHO estimates are based on all available data, from surveys of pregnant women, population-based surveys and other surveillance data. When measuring the number of people infected by HIV they have defined adults as 15 years and older. However, in the prevalence estimates adults are defined as those between 15 and 49 years of age. This has been the norm for all numbers, but the epidemic has a devastating effect on people above 49 years old as well, so they need to be included in estimates.

The lowest estimated adult prevalence in 2005, except for Comoros which has less than 0.1%, is 0.5 with a 95% confidence interval of (0.2-1.2). This estimate is for Madagascar. The second lowest is Mauritius with 0.6, and the third lowest is Mauritania with 0.7. Comoros, Madagascar and Mauritius are island states off the coast of Africa. Mauritania and the next countries on the list are all Western African states. The highest estimated prevalence is 33.4 in Swaziland and 24.1 in Botswana. The estimated prevalence for adults aged 15-49 in the whole sub-Saharan Africa is 6.1 %. So there are both generalized and concentrated epidemics in the dataset, but the vast majority are generalized epidemics.

The 9 countries on the bottom of the table in Appendix A are in southern Africa. This shows us that there are regional differences in Africa on where the epidemic strikes hardest. Southern Africa has the most devastating numbers. This is also seen in figure 2.5 in chapter 2 where southern Africa is dark red. The exception for this is Central African Republic which has a prevalence of 10.1% even though it is situated in Central Africa.

4.3 Education

4.3.1 What to measure?

I need to measure to what extent people are able to transform information they receive into action. This depends, among other issues, on their knowledge or human capital. But knowledge can be hard to quantify. An easier approach can be to measure the level of education. The next problem encountered is that it should be measured on country level. An estimate for knowledge should capture the country's aggregated ability to transform information into action. Number of years of completed education is a good measure in demographic population-based surveys with one observation for each individual, but at country level it is not adequate because averages can hide the spread of education in the population.

Measures for education can be divided into two main groups, related to stock and flow. The first group says something about the general level of education that the population holds, for example the average number of years completed education or literacy rates. On the other hand we have measures like enrolment rates that tell us how many are being educated right now, the flow of education. By separating these measures we can distinguish between the effects that flow data and stock data have on prevalence.

4.3.2 How to measure it?

A possible way to quantify the stock effect could be to use literacy rates, but then we only get information on how many in the population that are able to read, which is basic curriculum in primary school and we are not able to see effects of different education levels. The approach I have chosen in my analysis is related to flow. I have found data for how many in the general population who are enrolled at different levels of education. The data provide information on how wide education reaches in the country. It is also possible to measure effects of higher education because we have numbers for percentage of population enrolled in primary, secondary, and tertiary education.

A disadvantage when using enrolment rates is that they only tell us the percentage of the relevant age group that is enrolled in education. We can get numbers above 100% if there are many enrolled in education below or above their age rank. The data does not necessarily tell

us anything about the age distribution of people enrolled in education, but in most cases it is likely that the majority of students will be in the primary- or secondary school-age population. Another weakness is that it only measures how many students are enrolled, so the data does not say anything about how many students that actually finish primary, secondary, or tertiary education.
Education on macro level are associated with externalities. It is more complex than just the sum of the individual effects. Externalities can be hard to quantify. An example is the community effects of one persons' education as mentioned in chapter 4.1. When someone close to you have walked a path it is easier to walk the same way compared to if you were the first person in your village to aspire to higher education. The externalities can also be viewed in connection with development. Assume that development is correlated positively with education, as is argued by Gregson et.al (2001). In a two-equilibrium model a country will only need a collective level of education over a threshold to end in the good equilibrium which represents higher development.

A question to answer is whether to use primary, secondary, or tertiary education in the analysis. For tertiary education the numbers are too small. The range is 1-5%, except for South Africa with 13%. They are also unreliable because students from abroad will be counted in a country's numbers and students from that country that studies abroad will not be counted.

Even though data on education traditionally has been poor for sub-Saharan Africa, the collected data for primary and secondary education are good enough to be used. There are differences in the countries' school systems across the continent. In most countries the official age in primary education is 6-11 and in secondary education 12-18, but this can vary and awareness of this variation is important when interpreting the results.

In my analysis I will use data from UNESCO in the Global Education Database provided by USAID (USAID, 2006). I have considered gross enrolment rate as the indicator best suited for my purpose. The indicator should tell something about the general level of knowledge in the population, irrespective of at what age it has been acquired. However, age does matter, because many people in sub-Saharan Africa become sexually active at very young ages, and they need to be able to understand the necessity of precautions before they make their sexual debut not to place themselves at risk.

As we can see from table B.1 in Appendix B there are six countries with a primary gross enrolment rate below 50% in this dataset. These are Mali, Burkina Faso, Djibouti, Niger, Guinea and Mauritania. Most of these countries lie in Western Africa, and a large proportion of the population is Muslim. There are twelve countries with a primary gross enrolment rate above 100%. This means that there are more students enrolled in primary school than there are 6-11 year olds. Among these countries are Swaziland, Leshoto, Botswana and Zimbabwe; countries in southern Africa. The average primary gross enrolment rate is 77.825 whereas the median is 71.5. This indicates that there are no large biases towards low or high enrolment rate, but that the countries are evenly distributed between the minimum and the maximum value of the series.

When it comes to secondary gross enrolment rate the minimum value is 7%, which is found in Niger. Zimbabwe and Mauritius are the only countries with higher than 50% enrolment in secondary education. Compared to the primary gross enrolment rate, this is a more grouped series. The average over all sub-Saharan Africa is 22.18% and the median is 18%.

4.3.3 Endogeneity and Instrumental Variables

In chapter 4 I have argued that there can be endogeneity problems with using education as an explanatory variable for HIV-prevalence. This can be solved by finding instruments for education and use Instrumental Variable Estimation (IVE). The formal prerequisites for an instrument are first that the covariance between the instrumented variable and the instrument is different from zero, and second that the covariance between the instrument and the error term in the main equation is zero.

I will use the enrolment rates from 1985 and 1990 as instruments for education. In sub-Saharan Africa the awareness of HIV evolved over the 1990s. Before 1990 very few in the general population had any knowledge of HIV or AIDS. Therefore it can be argued that education data from 1990 or before should be old enough not to be affected by HIV-prevalence and hence there are no covariance between the error term and the instruments. It is also a reasonable assumption that the education system has not changed to an extent that there are no correlation between education in 1985, 1990 and 2002. Then both the formal prerequisites for an instrument are fulfilled.

4.4 Socio-economic factors and control variables

To single out the education effect in a regression analysis we need to control for other variables that can have an impact on HIV-prevalence. This can be religion, what kind of public health care that is provided, availability of condoms, degree of ethnic diversity, the country's GDP, and other socioeconomic and cultural factors. Generally we can say that these variables facilitate the narrowing of the know-do-gap. The population gets knowledge and information (know) through education and the control variables show to what extent the government and society make room for behavioural change (do).

4.4.1 Culture

Culture and norms are important for how a society reacts to changes in the conditions of life. Religion is a good proxy for different cultures and should therefore be included as a control variable. The difference between countries with large Christian communities and countries where traditional religion is most common is not expected to be that large. Hence, the most important is to measure the difference between Muslim and non-Muslim communities. However it is worth mentioning that the Catholic Church's ban of condoms can play a part in the spread of HIV among Catholic Christians, but on the whole this effect is probably too small to be of any importance here. I have solved this by including a variable that measures the percentage of the population that is Muslim. This is taken from CIA's World Fact Book (CIA, 2006). Even though these estimates are from different years and could be estimated with different accuracy, characteristics of religion are not very volatile, so it should be stable and reliable enough to use for comparison in this context.

4.4.2 Nation's ability to fight the epidemic

Another of the socioeconomic factors is a nation's ability to halt the epidemic. This depends on their attitude to the epidemic, but also on how rich they are. A country's GDP can be used as a measure for this. We need to use GDP pr capita to make it comparable across countries with different population sizes. I have used GDP pr capita that is adjusted for different purchasing power parities (PPP) to control for different price levels in the countries. GDP pr capita is measured in PPP US $ and is collected from the Human Development Report Indicators (UNDP, 2005). The data are computed on the basis of GDP and population data

provided by the World Bank. The numbers are from 2003 for all countries except Equatorial Guinea where the data refer to 2001.

A problem is that GDP pr capita does not say anything about how the country's resources are used. There have been tendencies that countries have different attitudes towards the epidemic, and this often affects how the resources are spent. Corruption and bad governance can also affect government spending so that the nations' utility is not maximized. In chapter 3 I have argued that the size and nature of the health sector is a necessary condition for education to have an effect on HIV-prevalence and a variable that tells us how much of GDP that is used on health.

4.4.3 Health expenditure

Total expenditure on health as a percentage of GDP could be a good measure of how health is prioritized. In many African countries there are numerous private clinics for the rich to use, so the question is whether to use total expenditure as a percentage of GDP or government expenditure on health as a percentage of total government spending. The second shows more clearly how the government prioritizes efforts to reduce the HIV-epidemic, but the first one is what matters for the country as a whole and therefore I have chosen to use that in the analysis. Data on expenditure is taken from the World Health Organization and the estimates are for 2003 (WHO, 2006). The data is collected from the different nations' health accounts and OECD-sources, and then consolidated by WHO to make it comparable. It should be noted that the data is not necessarily similar to the official health expenditure from each nation's Ministry of Health, but each nation has had an opportunity to comment on and correct the data collected by WHO. A problem with this variable is that endogeneity problems are very likely. If a country has high prevalence their expenditures on health should also be higher. I will discuss this possibility further in chapter 5. This could be solved by using an instrument for total expenditure on health in Instrumental Variable Estimation, but finding an instrument can be a time consuming task, so I have not been able to do this.

Indicators of the density of nurses or access to public health service, especially in rural areas could also show how good the health care is, but since the datasets are so small that there are only room for four explanatory variables I have chosen to use the total expenditure on health instead.

4.4.4 Urbanization effects

I have argued in chapter 4 that the distribution of rural and urban residents in a population can have an important effect on HIV-prevalence. Several studies show a significant difference in prevalence between rural and urban areas (Fylkesnes, 1997, Abebe 2003). This is because many people stay in urban areas for a shorter or longer period of time. Higher education is also correlated with a later age at marriage and increased pre-marital sex. Many are forced to urban areas for economic reasons too. They are away from their families and often engage in high-risk sexual relationships. However, when they come back to their families their spouses will be exposed to risk of HIV-infection even though they themselves have not placed them there. The result is that incidence rates rise in rural areas, the epidemic matures here too and characteristics are changing.

A possible problem is that there are also differences in educational level between rural and urban areas, so the effect on prevalence can both be a direct link between urbanization and HIV-prevalence and an indirect link that goes from urbanization, via education to HIV-prevalence. I have used percentage of total population that are urban residents to measure if there are differences between countries with mainly urban residents and countries where most of the population live in rural areas. The data is from UNDP's Human Development Report Statistics and is computed by UNs Department of Economic and Social Affairs, Population Divison (UNDP, 2005). The results should be interpreted with caution because the data are based on the countries' own definition of what constitutes an urban area, and these definitions may differ.

Because I have so few observations, I can only use a limited number of control variables. The ones that I will use in the analysis is then GDP pr capita, measured in PPP US $, total expenditure on health as a percentage of GDP, urban residents as a percentage of population and the last one is how many percentage of the population which is Muslim. This gives a good picture of the socioeconomic factors because it measures aspects of cultural differences, results of migration patterns, and how the health sector is prioritized.

5. Analysis

5.1 Basic regression analysis

In this chapter I will test if it is possible to measure the effect of education on HIV-prevalence based on empirical data. I will use Ordinary Least Squares (OLS) regression and Instrumental Variable Regression.[1] In econometric modelling variations in a dependent variable explained by variations in the independent variables and an error term which is included to capture effects that is not from the independent variables. The theoretical counterpart to the empirical residuals is disturbance. A set of assumptions about this disturbance has to be fulfilled for the estimates to be unbiased, consistent and with minimum variance (BLUE-estimates) (Hill, 2001).

1) The expectation of the disturbance is zero.
2) The variance of the disturbances is constant and they are uncorrelated.
3) The values of the independent variables are not random, there are no exact linear relationships between any of the independent variables and there are more observations than independent variables.
4) The disturbance is normally distributed.

A general econometric model will have the following form:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki}, i = 1,..,n$$

Where *k* is the number of explanatory variables and *n* is the number of observations.
In Instrumental Variable Estimation (IVE) one of the right-hand side variables are dependent and hence are correlated with the error term and the OLS-estimates will be biased and inconsistent (Hill, 2001). Then we can add a second equation which is a linear combination of the instruments. We measure the effect of this set of instruments on the dependent variable and then insert this estimate for the dependent variable in the first equation. For this method to work the instruments will have to be correlated with the second dependent variable, but uncorrelated with the disturbance in the first equation.

When dealing with empirical data it is a problem that real world data seldom fulfil all these assumptions. Especially in cross-section regression there are often few observations and the assumptions cannot be fulfilled. Then we have to try to find a best fit and be aware of the

[1] Regressions are done using PcGive in GiveWin version 2.10.

weakness in the data. The estimates are not necessarily BLUE-estimates, but can still be consistent or unbiased.

5.2 The model

There are different ways to model the relationship between prevalence and education. Since prevalence is measured in percentage with a range from 0 to 100, I have transformed it into logs to better facilitate for OLS. The log-models are a more flexible functional form compared to regular linear models. Another possibility could be to use PROBIT or LOGIT estimation. These are binary models that report the possibility that a random drawn observation from the set exhibits a given property. However, I have chosen to use OLS to keep things simple.

There are two main types of log-models, the log-linear model and the log-log model. In the log-linear model the explanatory variables are not logs. Both the slope and the elasticity of the function changes from point to point. The interpretation of the coefficient is that a one percentage point change in the independent variable will lead to a one unit change in the logarithm of the dependent variable. This is a somewhat tricky interpretation, but the important thing for the analysis here is that there will be an impact with the same sign as the coefficient, and the significance can be measured with a regular t-test.

In the log-log model all variables are transformed into logs. The slope changes from point to point, but the point elasticity is constant. A one percentage change in the independent variable leads to a percentage change in the dependent variable equal to the coefficient. Both the log-log model and the log-linear model are monotonic transformations of the data, so it does not change the relationships between the variables, but it can alleviate problems with for instance heterogeneity in the disturbance term. The important point when choosing a functional form is to see which form makes the data compatible with the assumptions for estimation. In the analysis I have tried both log-linear and log-log models. The results using the log-log formulation are at large in agreement with those found by using the log-linear formulation. However, I have chosen only to report the results from the log-linear model since the results from using this formulation appear to be the most reliable and robust.

I have divided the data into two different datasets. One has primary education as the main explanatory variable and the other secondary education. This is because I do not have enough

variables to measure the effect of both primary and secondary education together with all the control variables. I will also loose too many observations in the dataset if I include both primary and secondary education. If I compare the results from the same regression done with primary and secondary education it is possible to see whether there is a difference. If there are no differences between the two, I can conclude that according to the dataset educational level does not seem to have an impact on HIV-prevalence. A weakness here is that the prevalence data relates to stock, it measures the level of HIV-prevalence in the whole population, whereas the education data relates to flow because it measures the part of population that are educating themselves the year in question.

On both models I have first performed OLS-regressions where I include different independent variables to see if some of the education effect is really captured by other explanatory variables. Then I have done Instrumental Variable estimation to see if there are reasons to believe that there are endogeneity problems related to the education variable. The instrumented dependent variable is gross enrolment rate from 2002 and the instruments are gross enrolment rates from 1985 and 1990. I have done the exact same analysis for the datasets with respectively primary and secondary education as the first explanatory variable.

From chapter 3 I will expect to find a significant effect of education, but the coefficient can be both positive and negative since higher education has shown to be correlated with higher prevalence. The coefficient for total expenditure on health measured as percentage of GDP should be negative since higher expenditure on health should prevent the spread of the virus, but a positive coefficient could be an indication of endogeneity problems. The coefficient for urban residents should be positive since a higher degree of urbanization is related to higher prevalence. And the coefficient for percentage of population that is Muslim should be negative due to cultural factors.

5.3 Main results

The results from the regressions are shown in table 5.1 and 5.2. Table 5.1 shows the results for the regressions done with primary education as first explanatory variable and table 5.2 shows the results for secondary education. Model 1 is the point of departure. The coefficients for education here are positive, coherent with what is shown in the literature. The coefficient for primary education is slightly larger than the one for secondary education, but none of them

are significant and they build on different numbers of observations so this difference should not be weighted too much. Explanatory power, however, is very low. R^2 is smaller than 0.1 for both secondary and primary education. This implies that these results are very vague and not reliable. We have to dig deeper into the data to find explanations for the variation in HIV-prevalence.

The next step would be to control for socio-economic variables. When including the variable for percentage of Muslims in the population the results changes. The partial effect of this variable is negative and significant in all models. Explanatory power rise with approximately 0.5 in all models when percentage of Muslims in the population is included compared to when it is not included. This is a very robust result. Another interesting feature related to the inclusion of the proxy for culture is the interaction between the coefficients of education and percentage of Muslims. When the variable for Muslims in the population is included, the sign of the coefficient for education changes from plus to minus. This is another indication that the results in model 1 are unreliable.

A likely explanation for the change in coefficients for the education variables is that there is an omitted variable bias in model 1 in particular, but also in the other models where education is included and the socio-economic variables are not. This bias reflects the fact that there is a correlation between enrolment rates and percentage of Muslims in the population. When the socio-economic variables are excluded from the model the biased coefficient for education inhibits the coefficient for Muslims multiplied with the correlation. Mathematically it is defined as follows where the left hand side β is the one estimated without the Muslim-variable included and the βs on the right hand side are the true coefficients. The covariance and variance are both estimated. The signs of the components in the bias are indicated above and below the components.

$$\hat{\beta}_{educ} = \beta_{educ} + \underbrace{\bar{\beta}_{musl} \cdot \frac{\operatorname{cov}(\bar{educ}, musl)}{\operatorname{var}(educ)}}_{\text{bias}}$$

$$\bar{\beta}_{musl} \cdot \frac{\operatorname{cov}(\bar{educ}, musl)}{\operatorname{var}(educ)}$$

$$+$$

Since the variance of a variable is positive by definition, the true coefficient for percentage of Muslims in the population is negative and the estimated covariance is negative, this bias is

positive. This implies that when we do not include the variable for Muslims in the population the results show a more positive effect of education on prevalence than the true effect is. It is important to remember that the effect of an omitted variable is not just the sum of the true coefficients, but that the correlation between the variables also matters. Even though both true coefficients are negative, the estimated coefficient for education is positive when the variable for Muslims is omitted.

My results point in the following direction. There are likely to be a bias in model 1, the model that corresponds with the literature. If there is an effect of education on the variation in HIV-prevalence, the true effect appears to be negative. This implies that efforts to change the behaviour of individuals must aim wider than just to provide information and increase knowledge; they must take the socio-economic factors into account as well.

As we see from models 2-5 the IVE gives us approximately the same results as the OLS. The coefficients have the same signs and approximately the same significance level. Hence, it appears that the endogeneity problem in the variable for education can be overlooked in this dataset and OLS gives us unbiased and consistent estimates. This also implies that HIV-prevalence does not affect enrolment rates. HIV-status apparently does not affect the choice of whether to go to school or not, at least not when the numbers are aggregated at country level.

Percentage of people living in urban areas seems to have no real effect on the level of prevalence. The coefficient is unstable and only significant in model 3 for secondary education. It has the same sign in the two groups of models, except for model 8 where it is positive when secondary education is used and negative when primary education is used. It does not vary systematically with the variables included or excluded. The migration effect that was expected to be found in the results is not evident on country level with the data examined here, at least not when it is measured as degree of urbanization. But this could also imply that the differences in prevalence between urban and rural areas even out. The epidemic can have come to a level where it does not distinguish between urban and rural areas any longer. This would mean that the migration effect has reached the end of its potential, the ones who moved have brought the virus back to their home villages and families; prevalence in rural areas have increased and is now on the same level as in urban areas.

Table 5-1 – log model with log of estimated prevalence as dependent variable and primary education as explanatory variable. Numbers in parenthesis are t-values. In the IVE primary education enrolment rates from 1985 and 1990 are used as instruments for primary education in 2002.

Log-linear model - primary education

Endogenous variable - log of estimated prevalence 2003

	Model 1 OLS	Model 2 OLS	Model 3 OLS	Model 4 IVE	Model 5 IVE	Model 6 OLS	Model 7 OLS	Model 8 OLS	Model 9 OLS	Model 10 OLS	Model 11 OLS
Primary education 2002	0.0142	-0.0048	-0.013	-0.0214	-0.0132	0.00959	0.00984	-0.0048	-0.0125	-0.00398	0.0128
Gross enrolment rate	1.52	-0.549	-1.64	-0.668	-0.639	0.982	1.05	-0.531	-1.69	-0.47	1.29
GDP pr capita									-0.00013	-0.00014	
PPP US$									-1.92	-1.73	
Total expenditure on health			0.412		0.414	0.27	0.274		0.374		
% of GDP			3.68		3.05	1.69	1.8		3.63		
Urban residents			0.011		0.0109	-0.00176		-1.9E-05			-0.00805
% of total population			0.987		0.959	-0.111		-0.00147			-0.503
Muslims		-0.0267	-0.0313	-0.033	-0.0315			-0.0267	-0.0317	-0.0284	
% of total population		-4.18	-5.61	-2.45	-3.51			-4.06	-6.02	-4.54	
Number of observations	31	31	31	31	31	31	31	31	31	31	31
R-squared	0.0741	0.43	0.625			0.1704	0.17001	0.43	0.659	0.486	0.0824

Table 5-2 – log model with log of estimated prevalence as dependent variable and secondary education as explanatory variable. Numbers in parenthesis are t-values. Possible expansions of the model

Log-linear model - secondary education

Endogenous variable - log of estimated prevalence 2003

Estimation method	Model 1 OLS	Model 2 OLS	Model 3 OLS	Model 4 IVE	Model 5 IVE	Model 6 OLS	Model 7 OLS	Model 8 OLS	Model 9 OLS	Model 10 OLS	Model 11 OLS
Secondary education 2002	0.00544	-0.0166	-0.028	-0.0098	-0.0217	0.00703	0.0041	-0.0223	-0.00787	0.0132	0.0107
Gross enrolment rate	0.311	-1.18	-2.37	-0.582	-1.44	0.428	0.262	-1.41	-0.315	0.454	0.594
GDP pr capita									0.000691	-0.00023	
PPP US$									-0.409	-1.17	
Total expenditure on health			0.474		0.467	0.427	0.455		0.404		
% of GDP			4.21		4.1	2.37	2.64		3,23		
Urban residents			0.0254		0.0224	-0.0113		0.0131			-0.02078
% of population			2.06		1.7	-0.665		0.423			-1.14
Muslims		-0.0292	-0.0342	-0.028	-0.0326			-0.0324	-0.0277	-0.0279	
% of population		-4.36	-5.84	-4.06	-5.15			-4.14	-4.94	-4.14	
Number of observations	25	25	25	25	25	25	25	25	25	25	25
R-squared	0.00418	0.466	0.726			0.259	0.243	0.48	0.67	0.499	0.06

The coefficient for total expenditure on health is significant in most models. The exception is model 6 and 7 with primary education as explanatory variable. The coefficient, however, is not negative as we would expect, but positive. This could be a sign that the total expenditure on health actually depends on HIV-prevalence. Different degree of HIV-prevalence leads to differences in the burden the AIDS-epidemic put on the public and private health budgets. A country with a high prevalence will also have higher total expenditure on health because of an increase in the need for health care treatment of AIDS-victims. The most likely explanation is then that there is an endogeneity problem related to this variable. If there is an endogeneity between expenditure on health and HIV-prevalence, as I believe there is, it is hard to trust the results produced when this variable is included.

Primary education is not significant in any of the regressions. Model 3 and model 9 shows approximately the same results with t-values of -1.64 and -1.69, which we can label a very weak rejection. These models both include four explanatory variables, however in model 3 GDP pr capita is excluded and urban residents are included. The opposite goes for model 9. The coefficient is positive in model 1, 6, 7 and 11, but negative in the rest. The common feature here is that the models 1, 6, 7 and 11 do not include the Muslim variable while the other models do include this variable. The highest explanatory power (R^2) we obtain is 0.659 in model 9. This is remarkably high for a cross-section regression. Again it is the Muslim variable that is the reason for the difference. When it is not included R^2 varies from 0.0741-0.1704, and when it is included R^2 varies from 0.424-0.659. This implies that the percentage of Muslims in a country explains roughly half of the variation in HIV-prevalence.

In the results for secondary education model 3 and model 9 again singles out. Model 3 is the only one where the coefficient is significantly different from zero and it is also the model with the highest R^2 (0.726). But since we have reason to believe that the exclusion of GDP pr capita leads to an omitted variable bias model 9 is the one we should concentrate on. The coefficients of the variables that are included in both models are quite similar and there is only a small decline in R^2, from 0.726 to 0.67. This leads me to conclude that higher enrolment rates are connected with lower prevalence.

Primary and secondary education show similar effects trough the entire analysis. The coefficients have the same signs and approximately the same value in all models except model 10 where the coefficient for primary education is negative and the coefficient for

secondary education is positive. However, none of the coefficients are significant, so it is not possible to conclude that there is a difference in the effects from primary and secondary education. The R^2 is close to 0.5 in both versions. To examine the possible differences in the way primary and secondary education explains the variation in prevalence further, model 11 includes only urbanization as a control variable, since urbanization is expected to be correlated with education. The coefficients for both the education variable and the urbanization variable do not vary a lot, so there are no signs of variation in the effects. This leads to the conclusion that there are no big differences in the effect that primary and secondary education respectively have on HIV-prevalence in this dataset.

GDP pr capita does not seem to matter for HIV-prevalence. This is the only variable with different effects in the two groups of models. In model 10 it is the coefficient for the education variable that changes sign. It is negative for primary education and positive for secondary education. However none of the coefficients are significantly different from zero and since they are quite unstable they should not be given too much weight in this analysis.

There is no reason to expect that the assumptions for OLS are violated in such a way that these results are without value. As long as we do not have heterogeneity in the variances the estimates are BLUE. The residuals show no traces of heterogeneity in any of the models. The RESET test for model specification and omitted variables show no rejection when all relevant variables are included. Even though these kinds of tests should not be taken as absolute truths they give reason to believe that the model is to be trusted.

In sum the main point in the analysis is that it is of huge importance to control for socio-economic variables, and culture in particular, when examining the explanations for HIV-prevalence on the macro level.

5.4 Possible expansions of the model

There are several ways to improve this model. The most important for this kind of analysis would be to improve the variables in the existing model. This is a long and hard process. A better modelling of the health system that does not represent endogeneity is probably the most precarious improvement. Alternatives could be to use physician density or health care

facilities density. It could also be an option to find instruments for total expenditures on health and use instrumental variable estimation with these.

Another weakness is the size of the dataset. 25 and 31 variables are just on the limits of what it is possible to do OLS on. A possible way to solve this could be to include the countries in North Africa. However, these are mainly Arabic countries with a different culture than in sub-Saharan Africa. These countries have a very high percentage of Muslims in their countries, HIV-prevalence is very low and the pattern of infection is very different from the rest of Africa. In addition they represent a fairly homogenous group of countries, so there would not be added much variation in the variables, only more observations.

The third thing that could be done to improve the existing model is to find a better way of modelling religion effects. The relationship between culture and religion and also the differences between Islam, Christianity and traditional beliefs could be measured in a much better way than what is done here. This could be particularly important since this effect apparently explains very much of the variation in HIV-prevalence.

When it comes to other models that could be used there are three different paths. The first is to use panel-data estimation, the second is binary models that measure probabilities, like PROBIT or LOGIT mentioned earlier and the third is bootstrapping methods to take the uncertainty of the estimates of prevalence into account. The first one would give a better treatment of unobserved heterogeneity across countries and improve the modelling of the effects over time. This could be a bit tricky, since HIV-prevalence data good enough to be used in comparable studies at country level only exists for the last five years. The second option, binary modelling, could measure the probability of drawing a person that is infected with HIV in the different populations. This would be a much more complex analysis and would probably have to be done first on each country separately and then maybe it would be possible to compare the odds ratios given certain control variables. The third, bootstrapping methods, are probably the easiest solution if the analysis should be expanded. The dataset already reveal the confidence ranges for the prevalence estimates, and it should also be possible to collect the same for the explanatory variables. The prevalence estimate ranges are quite wide and varies from country to country. By generating new datasets, where the prevalence estimates are drawn at random from these intervals, a bootstrapping method could exploit the full information in the HIV data more carefully.

6. Discussion and conclusion

My main focus in this thesis is to see whether the variables that affect HIV-prevalence on regional level also determine the variation in HIV-prevalence on national level and, in particular, to what extent education can explain the variation. Education has traditionally been viewed as the main explanatory factor. However, in chapter four I argued that there are three main groups of variables that should be used to explain the variation in HIV-prevalence. The first is the effect of knowledge, or education, the second is the migration and urbanization effects and the last group is the socioeconomic factors.[2]

The effect of education is quite difficult to identify. There are indications that the enrolment rate affects prevalence, but whether it is positive or negative is uncertain. It is only significant in one case, with a negative coefficient for secondary gross enrolment rate. This could indicate that there is a decline in prevalence among the higher educated, which would be in accordance with the results found by Michelo (2006). However, the results do point towards a wide distance from knowledge to action, if not education should have had a significant effect. The significance of the results should be interpreted with caution though, due to few observations.

The main conclusion that can be drawn from this analysis is twofold. First, it shows that the socioeconomic factors are the most important variables to explain variation in HIV-prevalence levels. Even though education and migration can have important effects on the individual level, the socioeconomic factors are what matters most on macro level. From this follows that to focus solely on education as an explanatory variable is too simple. The correlation between education and culture causes omitted variable biases in the results when culture is excluded, as has been done in the literature.

Second, even though education does not seem to have a direct impact on prevalence it is a necessary condition to fight HIV/AIDS. Improving the level of education will facilitate an improvement in awareness of HIV/AIDS, but it is not in itself sufficient to stop the epidemic. One of the biggest obstacles that have to be overcome is the stigma and discrimination that many HIV/AIDS patients have to face. The governments play important roles in their

[2] See figure 4.1

willingness to facilitate behavioural change, as well as their willingness to finance good health care systems. Nevertheless, the main responsibility lies in each individual's choices and attitudes towards the risks to which they expose themselves and their close ones.

To shorten the distance from knowledge to action, cultural and religious aspects and traditions are necessary to address. The socio-economic factors considered as the third main group of effects can facilitate this shortening. The unexpectedly strong impact of these variables in the analysis, particularly the percentage of the population that is Muslim, strengthens this theory.

This analysis is done on country level. In sub-Saharan Africa the countries have a large ethnic diversity and different ethnic groups can have unalike attitudes towards behavioural change. The openness towards new ideas depends on tradition and traditional attitude to changes. One government can experience that their programs to prevent the spread of HIV will get different receptions by different ethnic groups. This can be a complicating factor in the planning of their efforts.

On individual level the ability to transfer knowledge into behavioural change depends on personal character and intelligence as well as the culture and tradition that one is a part of. Education could be closely connected to character and intelligence, but, particularly in sub-Saharan Africa, the access to education is not evenly distributed across the population so intelligent people without education are in some cases more likely to be able to instigate behavioural change than less intelligent people with more education. These differences are almost impossible to measure, but should be kept in mind.

An interesting result is the relationship between the percentage of Muslims in the population and prevalence. There are different possible explanations for this relationship, some of them cultural and some religious. As mentioned in chapter 3 Gray (2004) has shown that the two most important reasons why Muslims have a lower prevalence are the high degree of male circumcision and the low alcohol consumption in Muslim communities.

In addition to these explanations it is also a possibility that some of the variation seen can be explained by temporal dynamics. Gregson (2001) argues that HIV-prevalence is correlated positively with development. If we look more closely at the Muslim countries in our survey, they have lower GDP pr capita and lower educational level, which indicates lower

development. The situation could simply be that the Muslim countries have not come to the same level of development that the countries in southern Africa have, and hence that the AIDS epidemic has not rooted itself in the population yet. When the Muslim countries develop, rising HIV-prevalence will be one of the side effects, incidence rates will increase and eventually prevalence rates will escalate. Awareness of this scenario could prevent it from happening as long as the governments take the action that is needed with respect to information and health care.

The other socio-economic factors that were expected to affect HIV-prevalence, urbanization and GDP pr capita do not seem to have an impact. There are two possible explanations for this. One is that the variable is measured or the data is collected in a way that does not incorporate the vital trends, or that there are differences from country to country in how it has been collected so that cross-section analysis gives spurious results. The other explanation is that these variables, contrary to what is shown to be the case on regional level, do not have an impact on HIV-prevalence on country level.

In the results the urbanization effect does not appear to have been of great importance at country level. A possible explanation is that some of the effect is incorporated in the education variables. But it could also be that the latest spread of the epidemic to rural areas actually shows in the results so that the differences between rural and urban areas even out. However, the correlation between percentage of the population that are urban residents and enrolment rate is actually negative in this dataset. That could point in the direction that it is not necessarily the effects that are unexpected, but that the variable should be examined more thoroughly. It could be differences from country to country in how urban areas are defined or other measurement errors that explain the unexpected results.

When it comes to GDP pr capita this does not have as large effect on HIV-prevalence as expected. It is most likely that GDP pr capita actually cannot explain the variation in HIV-prevalence since the coefficient for GDP pr capita is unstable and changes sign in an unsystematic way. Hence, the suggestion from Gregson (2001) that development and high HIV-prevalence is positively correlated is not that evident in the data examined here, at least not if development is measured as GDP pr capita. Then there should be a clear positive relationship. It could be that this effect only matters in certain regions or only in urban areas, not in rural. Then it could be that the variable works in different directions depending on the

characteristics of the country and we will not get significant results when comparing across countries. It is also an important point that how rich a person is does not matter for the risk that he or she is exposed to. The advantage for the rich is that they can pay for the best available treatment and keep the infection at a level were AIDS is not developed. In this perspective the results are reasonable because prevalence then would not depend on GDP pr capita. This means that if an indicator for how many people that die from AIDS was used instead of prevalence, the coefficient for GDP pr capita would be significantly negative.

The development-argument for a positive relationship only holds to a certain extent, though, because higher GDP pr capita should imply that the government had more money to put into information campaigns and condom distribution. Then higher GDP pr capita should have a negative effect on prevalence. The bottom line is that there are many mechanisms between GDP pr capita and HIV-prevalence and they point in different directions so the results do not produce any clear relationship.

Total expenditure on health actually explains quite a lot of the variation in HIV-prevalence, but since this variable probably is endogenous it is impossible to trust the results when this variable is included. However, it shows that countries with high HIV-prevalence spend relatively much on health care.

The awareness of the dimensions that the HIV-epidemic has in sub-Saharan Africa is increasing. Many people have joined the forces to combat HIV, and increasing amounts of financial and human resources are devoted to research and treatment distribution. However, the highest obstacle for sub-Saharan Africa is the stigmatisation and discrimination that HIV-infected face.

The first step is to start talking about the disease and its consequences, not only in the governments' offices, but on the streets, in the shops and on the fields. Empowering women to stand up against suppression, often disguised in traditions and cultural heritage, and take control over their own lives and sexuality is also vital. In the end it all comes down to individuals and their choices. To stop the epidemic people both need to be informed about the consequences of their actions and understand the necessity of behavioural change. Information alone is apparently not enough.

References/Literature

Abebe, Yigeremu et.al (2003): "HIV prevalence in 72 000 urban and rural male army recruits, Ethiopia", AIDS journal Vol 17: 1835-1840

Caldwell, John (2000): "Rethinking the African AIDS epidemic", Population and Development Review Vol 26 No 1: 117-135

CIA (2006): World Factbook. Potomac Books, Inc., Dulles

Crampin, Amelia C. et.al (2003): "Trends and measurement of HIV prevalence in northern Malawi", AIDS journal Vol 17: 1817-1825

Economist, The (2006): "1.3 by 5", The Economist, April 1st 2006

Feeney, Griffith (2001): "The impact of HIV/AIDS on adult mortality in Zimbabwe", Population and Development Review Vol 27 No 4: 771-780.

Fylkesnes, Knut et.al (1997): "The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women", AIDS Journal vol 11: 339-345

Gray, Peter B. (2004): "HIV and Islam: is HIV prevalence lower among Muslims?", Social Science & Medicine Vol 58: 1751-1756

Gregson, Simon, Waddell, Heather and Chandiwana, Stephen (2001): "School education and HIV control in Sub-Saharan Africa: From discord to harmony?", Journal of International Development Vol 13: 467-485

Harden, V and V. Hannaway (1997): "Oral history interview with Dr. Samuel Broder", `http://aidshistory.nih.gov/transcripts/transcripts/Broder97.pdf`, downloaded July 10th 2006

Hargreaves, James R. and Glynn, Judith (2002): "Educational attainment and HIV-1 infection in developing countries: a systematic review", Tropical Medicine and International Health Vol 7 No 6: 489-498

Hill, R.C, W. Griffiths and G. Judge (2001): Undergraduate Econometrics. John Wiley & Sons, Inc.

Joint Learning Initiative (2004): Human Resources for Health; Overcoming the crisis, Harvard University Press, Global Equity Initiative, Cambridge, Massachusetts

Kravdal, Øystein (2002): "Education and fertility in sub-Saharan Africa: Individual and community effects", Demography Vol 39: 321-334

Kelly, Robert et.al. (1999): "Age of male circumcision and risk of prevalent HIV infection in rural Uganda", AIDS Journal Vol 13: 399-405

Larson, Ann (1990): "The social epidemiology of Africa's AIDS epidemic", African Affairs Vol 89 No 354: 5-25

Michelo, Charles et.al (2006): "Marked HIV prevalence declines in higher educated young people: evidence from population based surveys (1995-2003) in Zambia", AIDS Journal Vol 20 No 7: 1031-1038

Mwaluko et.al (2003): "Trends in HIV and sexual behaviour in a longitudinal study in a rural population in Tanzania, 1994-2000", AIDS Journal Vol 17: 2645-2651

Olshansky, S.J et.al (1997): "Infectious diseases – New and ancient threats to world health", Population Bulletin Vol 52 No 2, Population Reference Bureau Inc, Washington D.C

Treichler, Paula (1991): "AIDS, Africa and Cultural Theory", Transition No 51: 86-103

UNAIDS (2005): "AIDS epidemic update: December 2005", UNAIDS, Geneva

UNAIDS (2006a): "2006 Report on the Global AIDS Epidemic", UNAIDS, Geneva

UNAIDS (2006b): "Fact sheet, 25 years of AIDS" `http://data.unaids.org/pub/FactSheet/2006/20060428_FS_25YearsofAIDS_en.pdf`, downloaded May 5th 2006

UNAIDS (2006c): "Fast facts about AIDS" `http://www.unaids.org/en/MediaCentre/References/default.asp#begin`, downloaded July 19th 2006

UNDP (2005): "Human Development Report 2005", United Nations Development Programme, New York

USAID (2006): "Global Education Database", `http://qesdb.usaid.gov/ged/index.html`, downloaded June 12th 2006

Weiss, Helen, M. Quigley and R. Hayes (2000): "Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis", AIDS Journal Vol 14: 2361-2370

WHO (2005): "Progress on Global Access to HIV Antiretroviral Therapy; An update on 3by5, June 2005" World Health Organization, Geneva

WHO (2006): "Working together for health, The World Health Report 2006", World Health Organization, Geneva

Appendix A

Table A.1 – Estimated prevalence in Sub-Saharan Africa

Sub-Saharan Africa	6.1
Comoros	<0.1
Madagascar	0.5
Mauritius	0.6
Mauritania	0.7
Somalia	0.9
Senegal	0.9
Niger	1.1
Guinea	1.5
Sierra Leone	1.6
Mali	1.7
Benin	1.8
Burkina Faso	2
Ghana	2.3
Eritrea	2.4
Gambia	2.4
Rwanda	3.1
Djibouti	3.1
Equatorial Guinea	3.2
Democratic Republic of Congo	3.2
Togo	3.2
Burundi	3.3
Chad	3.5
Angola	3.7
Guinea-Bissau	3.8
Nigeria	3.9
Congo	5.3
Cameroon	5.4
Kenya	6.1
United Republic of Tanzania	6.5
Uganda	6.7
Côte d'Ivoire	7.1
Gabon	7.9
Central African Republic	10.7
Malawi	14.1
Mozambique	16.1
Zambia	17
South Africa	18.8
Namibia	19.6
Zimbabwe	20.1
Lesotho	23.2
Botswana	24.1
Swaziland	33.4
Ethiopia	…
Liberia	…

Appendix B

Table B.1 - Dataset for primary education

Estimated prevalence, education, urban population, percentage of Muslims in the population and total expenditure on health is measured in percentage. GDP pr capita is measured in US $ adjusted for purchasing power parity.

Country	estprev	primeduc1985	primed1990	prime2002	urbpop	muslim	gdpprcap	totexphealth
Benin	2.0	68	58	109	44.6	20	1115	4.4
Botswana	24.0	105	113	103	51.6	0.001	8714	5.6
Burkina Faso	2.1	27	33	46	17.8	50	1174	5.6
Burundi	3.3	53	73	77	10	10	648	3.1
Cameroon	5.5	102	101	108	51.4	20	2118	4.2
Central African Republic	10.8	75	65	66	42.7	15	1089	4.0
Chad	3.4	44	54	78	25	51	1210	6.5
Comoros	0.1	82	75	90	35	98	1714	2.7
Congo	5.4	147	133	80	53.5	2	965	2.0
Côte d'Ivoire	7.0	72	67	78	44.9	40	1476	3.6
Djibouti	3.1	40	38	40	83.6	94	2086	5.7
Gambia	2.2	68	64	85	26.2	90	1859	8.1
Ghana	2.3	75	75	79	45.4	16	2238	4.5
Guinea	1.6	34	37	81	34.9	85	2097	5.4
Kenya	6.8	99	95	92	39.3	10	1037	4.3
Lesotho	23.7	112	112	126	18	0.001	2561	5.2
Malawi	14.2	60	68	140	16.3	12.8	605	9.3
Mali	1.8	25	26	58	32.3	90	994	4.8
Mauritania	0.7	48	49	88	61.7	100	1766	4.2
Mauritius	0.2	109	109	104	43.3	16.6	11287	3.7
Mozambique	16.0	87	67	103	35.6	17.8	1117	4.7
Niger	1.1	26	29	44	22.2	80	835	4.7
Nigeria	3.7	104	91	119	46.6	50	1050	5.0
Rwanda	3.8	63	70	122	18.5	4.6	1268	3.7
Senegal	0.9	56	59	80	49.6	94	1648	5.1
Swaziland	32.4	102	111	98	23.6	10	4726	5.8
Tanzania, U.R	6.6	75	70	84	35.4	35	621	4.3
Togo	3.2	93	109	121	35.2	20	1696	5.6
Uganda	6.8	73	74	141	12.3	16	1457	7.3
Zambia	16.9	105	99	82	35.9	24	877	5.4
Zimbabwe	22.1	136	116	94	35	1	2443	7.9

Table B.2 – Dataset for secondary education

Estimated prevalence, education, urban population, percentage of Muslims in the population and total expenditure on health is measured in percentage. GDP pr capita is measured in US $ adjusted for purchasing power parity.

Country	estprev	seceduc1985	seced1990	sece2002	urbpop	muslim	gdpprcap	totexphealth
Benin	2.0	18	12	28	44.6	20	1115	4.4
Botswana	24.0	29	43	73	51.6	0	8714	5.6
Burkina Faso	2.1	4	7	11	17.8	50	1174	5.6
Burundi	3.3	4	6	11	10	10	648	3.1
Cameroon	5.5	23	28	31	51.4	20	2118	4.2
Chad	3.4	6	8	15	25	51	1210	6.5
Comoros	0.1	28	18	31	35	98	1714	2.7
Djibouti	3.1	12	12	20	83.6	94	2086	5.7
Gambia	2.2	16	19	34	26.2	90	1859	8.1
Ghana	2.3	40	36	39	45.4	16	2238	4.5
Guinea	1.6	13	10	24	34.9	85	2097	5.4
Kenya	6.8	21	24	33	39.3	10	1037	4.3
Lesotho	23.7	23	25	35	18	0	2561	5.2
Malawi	14.2	6	8	33	16.3	12.8	605	9.3
Mauritania	0.7	15	14	23	61.7	100	1766	4.2
Mauritius	0.2	49	53	81	43.3	16.6	11287	3.7
Mozambique	16.0	7	8	16	35.6	17.8	1117	4.7
Niger	1.1	5	7	7	22.2	80	835	4.7
Nigeria	3.7	34	25	36	46.6	50	1050	5.0
Rwanda	3.8	6	8	16	18.5	4.6	1268	3.7
Senegal	0.9	14	16	19	49.6	94	1648	5.1
Swaziland	32.4	39	44	45	23.6	10	4726	5.8
Uganda	6.8	10	13	20	12.3	16	1457	7.3
Zambia	16.9	19	24	28	35.9	24	877	5.4
Zimbabwe	22.1	42	50	40	35	1	2443	7.9