

# Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa

Emily Oster\*

University of Chicago and NBER

First Draft: March 25, 2007

This Draft: December 5, 2007

## Abstract

I generate new data on HIV incidence and prevalence in Africa based on inference from mortality rates. I use these data to relate economic activity (specifically, exports) to new HIV infections in Africa and argue there is a significant and large positive relationship between the two: a doubling of exports leads to as much as a quadrupling in new HIV infections. This relationship is consistent with a model of the epidemic in which truckers and other migrants have higher rates of risky behavior, and their numbers increase in periods with greater exports. I present evidence suggesting that the relationship between exports and HIV is causal and works, at least in part, through increased transit. The result has important policy implications, suggesting (for example) that there is significant value in prevention focused on these transit-oriented groups. I apply this result to study the case of Uganda, and argue that a decline in exports in the early 1990s in that country appears to explain between 30% and 60% of the decline in HIV infections. This suggests that the success of the Ugandan anti-HIV education campaign, which encouraged changes in sexual behavior, has been overstated.

## 1 Introduction

In 1347, a ship arrived from the Black Sea in the port of Marseille carrying a full cargo, a sick crew, and the Bubonic Plague. Over the next several years, the plague spread throughout Europe, killing as much as a third of the population. Although many aspects of life in that period favored the plague, it was international trade that was the ultimate source of the epidemic. With no trade at all, the outbreak of plague in China in the 1330s might well have been confined. More generally, migration, travel and trade are major drivers of many historical epidemics (Wilson, 1995; McNeill, 1976).

Data on individuals suggests that similar forces may be at work in the case of the HIV/AIDS epidemic in Sub-Saharan Africa. Trucker and migrants in Africa are more likely to be

---

\*Ward Cates, Jane Fortson, Daniel Halperin, Norman Hearst, Emir Kamenica, Lawrence Katz, Michael Kremer, Steven Levitt, Kevin Murphy, Ben Olken, James Poterba, Jesse Shapiro, Andrei Shleifer, Rebecca Thornton and participants in seminars at the Gates Foundation, the University of Chicago, UCL and Northwestern provided helpful comments. Dwyer Gunn provided outstanding research assistance.

infected with HIV (Lurie et al, 2003; Brewer et al, 1998; Brockerhoff and Biddlecom, 1999; Arnafi et al, 1997; Arnafi, 1993), and it has been argued (in India, as well as in Africa) that HIV spreads along trucking routes (Klitsch, 1992; Steinbrook, 2007). Motivated by the historical evidence and this individual-level connection, this paper explores whether variations in trade – specifically, in exports – drive HIV infections in Africa, and whether this relationship is large enough in magnitude to be an important driver of overall HIV prevalence there.

The major challenge to this estimation is poor data on HIV infection rates. Until the early 2000s, most HIV testing in Africa focused on pregnant women or high risk groups (drug users, STI patients), meaning we know very little about the level of infection in the general population.<sup>1</sup> More problematic, perhaps, is the fact that testing of even these non-representative groups is inconsistent (variations in the areas or populations tested over time), making the prevalence data very noisy and difficult to compare across years. Noisy data on prevalence, in turn, makes it almost impossible to infer anything reliable about the number of new infections each year (incidence), which is the ultimate variable of interest.<sup>2</sup>

The paper therefore begins by generating new estimates of HIV prevalence and incidence based on inference from mortality data. These new estimates match well to estimates from population-based testing, to the extent that the latter are available, but provide significantly more information about the path of the epidemic over time. I then use these data to estimate the relationship between exports and new HIV infections (incidence). I find a large and significant positive relationship between HIV and exports: a doubling of exports appears to lead to as much as a quadrupling in number of new HIV infections, which suggests that if exports overall had been 25% lower in Africa over the course of the epidemic only about half as many people would have become infected. In addition, this relationship seems to explain at least some of the very large decline in HIV prevalence in Uganda in the 1990s, which is typically attributed to an anti-HIV education campaign, which encouraged partner reduction and ABC (Abstain, Be faithful and use Condoms) and has been widely replicated. A decline in the coffee market accounts for between 30% and 60% of the decline in HIV incidence, suggesting the success of this education campaign may be overstated.

The estimation of HIV prevalence and incidence proceeds in two steps: first, estimating deaths from HIV in Africa; second, translating these deaths into prevalence sometime in the past.

---

<sup>1</sup>In the last few years, excellent data based on testing of a random sample of the population have come out from the Demographic and Health Surveys (Halperin and Post, 2004). However, this does not address issues of lack of consistent data over time.

<sup>2</sup>Incidence is the share of the population who becomes infected in a given year (the flow). Prevalence is the share that *is* infected in a given year (the stock).

The concept behind the first step is illustrated in Figure 1, which shows death rates by age in Egypt in the 1990s (unaffected by HIV) and Botswana in 2001 (affected by HIV). Although death rates are similar among young children and older people (suggesting similar levels of overall development), the death rates in Botswana among prime-age adults are much higher than in Egypt. Virtually no other disease has this pattern over the life-cycle, so the gap in mortality between the two countries is likely to be due almost entirely to HIV.<sup>3</sup> The difference in death rates generates an estimate of the number of deaths due to HIV in Botswana.<sup>4</sup>

Translating deaths into infections requires an assumption about time from infection to death. In the simplest case, if everyone died ten years after infection, then infections ten years ago would be equal to deaths this year. In fact, time to death varies, with a significant fraction of individuals dying at any time between 4 and 20 years after infection. With this structure, infections and deaths are linked by a system of equations, which can be solved given information on time to death from infection. This paper is certainly not the first to suggest that this type of methodology might be used to calculate HIV infections. Deuffic-Burban and Costagliola (2006) use a similar methodology (“backcalculation”) to calculate incidence in France, as does Ong et al (1998) in Malaysia and Liao and Brookmeyer (1995) in the United States, among many others. In all of these cases, the papers use data on official reported cases of HIV or AIDS, rather than data on mortality, which is obviously problematic in terms of extending the analysis to Africa. Weinreb (1999) suggests that mortality data might be used in Africa to undertake this type of methodology. To my knowledge, however, this paper is the first attempt to use this type of methodology to calculate prevalence rates in Sub-Saharan Africa.

I estimate incidence and prevalence from the early 1980s to the late 1990s for nine countries in Sub-Saharan Africa.<sup>5</sup> I validate the new estimates of prevalence by comparing them to estimates from population-based testing (from the Demographic and Health Surveys) in cases where the latter are available. The match is quite close in magnitude. In addition, even within a country, mortality-based prevalence by age and gender group is strongly correlated with population testing for that group. Consistent with recent evidence from population testing (Halperin and Post, 2004;

---

<sup>3</sup>Other authors, notably Feeney (2001) and Timeaus and Jasseh (2004) have used the existence of this gap to discuss the effect of HIV on mortality patterns.

<sup>4</sup>Clearly, the quality of estimate of number of deaths relies heavily on the quality of the mortality data in Africa. Official data on mortality in Africa are not widely available, so I make use of sibling mortality histories from the Demographic and Health Surveys to construct mortality over time. Sibling histories have been used by other authors to calculate death rates for countries with limited official reporting (Bicego, 1997; Timeaus and Jasseh, 2004; Gakidou, Hogan and Lopez, 2004; Stanton, Abderrahim and Hill, 2000). Generally, they are thought to be reliable and match well to official statistics. In the data section I discuss this in more detail and perform my own validation checks.

<sup>5</sup>In particular, I provide estimates for all countries for which mortality data is available.

ORC Macro, 2006), the estimates here suggest that UNAIDS estimates in the late 1990s were overestimated by as much as a factor of two.

Turning to the motivating question of the relationship between HIV and exports, I first show a simulation model of the epidemic that serves both to pin down the theory behind the HIV-export connection and generates hypotheses about variations in this relationship. In the model, a share of the male population travels in each period; these travelers have more sexual partners and also encounter partners with higher prevalence, both of which will drive up their infection rates. In addition, the wives of these travelers are more likely to have non-marital partners than the wives of men who do not travel (all of these assumptions are motivated by micro-level data on behavior). In between periods, the travelers return to their wives, and this further interaction also drives infection rates. The model makes a number of predictions. First, when the share traveling increases, the number of new infections also increases. Second, there is an interaction with existing prevalence rates: the higher the prevalence, the stronger the relationship between infection and share traveling. Finally, the model suggests that there should be a relationship between both male infection rates and share of men traveling *and* between female infection rates and share of men traveling, but the latter relationship should be smaller. The model focuses on the share of men traveling as driving the variation; in this section I also provide empirical evidence that travel is greater in period of higher exports, thus linking (indirectly) infection rates to exports.

I use the new measures of HIV to estimate the HIV-export relationship. Consistent with the predictions of the model, I find a significant relationship between the two series, controlling for country and year fixed effects, as well as country-specific trends. The magnitude suggests that doubling exports leads to approximately a quadrupling of new infections. Put differently, the results suggests that if exports were 25% lower in all years in the sample HIV prevalence would be only about half as high. Consistent with the secondary predictions of the model, I also find that the relationship between exports and new infections is stronger when the baseline infection rate is higher, and the relationship is smaller in magnitude when we consider female infections, although still positive and significant.

A central concern with interpretation is the possibility that it is not the export-transit mechanism which drives the result but, rather, some omitted variable (for example, GDP) which drives both exports and HIV. Although this is potentially consistent with the primary results, I argue that there is evidence in favor of a causal interpretation of exports. First, the relationship between HIV and exports is stronger in areas with a greater density of roads and areas closer to

major cities, which is consistent with the transit mechanism. Second, instrumenting for export volume with world commodity prices also points to a positive and significant relationship. Third, the relationship between exports and HIV is stronger in countries where the major export is more closely linked to trucking. Finally, new HIV infections can be linked directly to truck imports, which strongly points to an effect of transit.<sup>6</sup>

It is important to note in interpreting these results that I estimate the *short-term* effect of changes in exports on HIV incidence, within a given country. This is distinct from an estimate of the long-term effect of exports: in the long run, higher levels of economic activity are likely to translate into economic growth. As individuals become richer the epidemic may get worse (if money buys sex), or better (if richer people have more incentives to avoid HIV, or more information to do so (Oster, 2007)). Neither of these forces will be captured here. In addition, the estimate here is distinct from estimates of what determines cross-country patterns in HIV. Africa has, on average, lower levels of exports than much of the rest of the world, and higher HIV, and variations in HIV across countries within Africa do not always line up with aggregate variations in exports. This clearly points to some other factor – for example, circumcision or overall health – which drives HIV, and is not captured here. Although the estimates here may point to an important role for exports in explaining within country variations – they by no means explain all of the cross-country or cross-region variation.

The final section of the paper applies the intuition from the export-HIV connection to understanding prevalence declines in Uganda in the 1990s. There was a large drop in HIV prevalence in Uganda in the 1990s, which is unique among Sub-Saharan African countries; this drop is generally credited to educational programs, which promoted partner reduction (“zero grazing”), and to elements of ABC (Abstain, Be faithful and use Condoms) (Green, 2004; Green et al, 2006; Slutkin et al, 2006). This anti-HIV education campaign was among the first successful HIV prevention strategies, and this type of education campaign has become a central tenant of HIV prevention in Africa, with extensions to Kenya, Tanzania, Malawi and elsewhere. Both UNAIDS and PEPFAR have taken this as a major focus of their prevention approaches; in 2004, PEPFAR spent \$100 million of their budget on ABC-related programs.

During this period in Uganda, however, there was also a very large decline in exports, due primarily to a decline in world coffee prices. The timing of this decline lines up extremely closely with the decline in HIV incidence, as can be seen in Figure 2. The magnitude of the relationship between

---

<sup>6</sup>In a robustness section I also present evidence that various assumptions used to generate the data on HIV are not driving the export results.

HIV and exports suggests that between 30% and 60% of the incidence decline in Uganda during this period can be attributed to changes in economic activity, which in turn suggests that some of the success of the education campaign has been overstated (although, of course, a large share of the incidence decline may well be due to this education campaign, and that should not be discounted).

This result on Uganda has obvious, potentially important, policy implications. The connection between exports and HIV in general also suggests specific avenues of HIV prevention. In particular, the implication that truckers and other migrants are a very important driver of the overall epidemic supports targeting prevention activities at that group (similar to the targeting done of prostitutes in Thailand). In addition, if increases in economic activity make the HIV epidemic worse, it suggests that aid groups aiming to increase growth should do so in collaboration with those seeking to decrease HIV.

In addition to the export-HIV result, this paper also contributes new estimates of HIV prevalence. Historically, those interested in the magnitude of the HIV/AIDS epidemic in Africa have relied on prevalence estimates reported by UNAIDS. This is true in policy circles (PEPFAR and the World Bank websites both link directly to UNAIDS data on the epidemic), media reports (articles in the New York Times (Editorial Desk, 2005; Altman, 2006a; Altman, 2006b) and the LA Times (Maugh, 2006)) and academic research (Kalemi-Ozan, 2006; Werker, Ahuja and Wendell, 2006; Young, 2006). The reliability of these estimates has been called into question in recent years by new population-based testing (ORC Macro, 2006) and, although population-based testing has expanded our knowledge about the current state of the epidemic, consistent data on the past are virtually impossible to glean from existing sources. These data are crucial, however, to much of the academic work on the epidemic (Kalemi-Ozan, 2006; Werker, Ahuja and Wendell, 2006; Young, 2006; Forston, 2006), as well as to understanding trends over time. The prevalence estimates here provide an alternative to relying on either inconsistent testing data or model-generated prevalence for analyses requiring a time series of HIV prevalence.

The rest of the paper is organized as follows. Section 2 discusses the methodology for estimating HIV prevalence, the data used and presents the results of the estimation. Section 3 discusses the simulation model of the epidemic and derives predictions; Section 4 estimates the relationship between HIV incidence and export volume. Section 5 discusses the case study of Uganda, and Section 6 concludes.

## 2 Estimating HIV Incidence and Prevalence

This section presents the methodology for estimating HIV incidence and prevalence in Subsection 2.1, and the data used in Subsection 2.2. Subsection 2.3 then presents the resulting prevalence estimates, with a focus on validating these estimates by comparing to existing HIV testing data.

### 2.1 Methodology

The methodology used here begins with deaths from HIV in a given year. The output is incidence or prevalence rate, sometime in the past. Section 2.2 below describes how deaths from HIV are calculated; this section starts from the assumption that deaths are known. Denote the number of deaths from HIV among age group  $i$  in year  $t$  as  $\mu_{i,t}$ .

As discussed in the introduction, time to death from HIV is a central input to this problem. Detailed data on time to death is difficult to generate, particularly in developing countries, since it requires knowing (roughly) the time of infection. For this reason, the best available data is drawn from developed countries, from time periods before HIV treatment was available, and I will use these data. It is possible, of course, that time to death is different (in particular, perhaps faster) in Africa, but there is no empirical evidence suggesting this is true. In the robustness section I will explore the robustness of the export results to different assumptions about time to death.

Figure 3 shows a graph of time to death from infection, drawn from existing work, by age group (Collaborative Group on AIDS Incubation and HIV Survival, 2000; Stover, 2003; Statistics South Africa, 2004). This distribution shows a large mass at virtually all periods between 4 and 20 years after infection. Since 90% of deaths occur during this period, I will focus on those 17 years.<sup>7</sup> The figure also demonstrates that time to death is longer for individuals who are infected at younger ages. The analysis across age groups will incorporate these differences. One important note is that in an era with HIV treatment, these paths of time to death are likely to be very different. However, in the period covered by this data, in Africa, treatment levels were extremely low, implying that it seems reasonable to use the time path for a situation of no treatment.

Assume that a share  $d_\alpha$  of infected people die  $\alpha$  years after infection, where  $d_\alpha$  is known (for example, taken from Figure 3) and, for the moment, that no one dies from anything else (this assumption is relaxed below, and in generating the estimates). Denoting  $b_{i,t}$  as the number of

---

<sup>7</sup>Extending the time frame to 1-20 years does not make much difference; however, since very early deaths are rare and poorly measured, we begin at 4 years where the share of deaths is large enough to be more reliable. Similarly, truncating the end at 17 years also makes no difference.

infections among age  $i$  in year  $t$ , we can write the total HIV deaths among age group  $i$  in year  $t$  ( $\mu_{i,t}$ ) as:

$$\mu_{i,t} = d_4 b_{i-4,t-4} + d_5 b_{i-5,t-5} + \dots + d_{20} b_{i-20,t-20} \quad (1)$$

This is a single equation with seventeen unknowns, and it is therefore not identified. Adding another year of data on mortality generates another equation (see below), but also another unknown ( $b_{i-21,t-21}$ ).

$$\mu_{i,t} = d_4 b_{i-4,t-4} + d_5 b_{i-5,t-5} + \dots + d_{20} b_{i-20,t-20} \quad (2)$$

$$\mu_{i-1,t-1} = d_4 b_{i-5,t-5} + d_5 b_{i-6,t-6} + \dots + d_{20} b_{i-21,t-21} \quad (3)$$

The system of equations will be identified only given data such that

$$\mu_{i-x,t-x} = d_4 b_{i-x-4,t-x-4} \quad (4)$$

That is, we must observe a year in which the only deaths are from people infected four years ago. Assuming that  $d_4$  is known, equation (5) will identify  $b_{i-x-4,t-x-4}$  and we can solve the system of equations backwards to solve for the entire  $\mathbf{b}$  vector.

There are two ways to fulfill the requirement detailed in equation (5). One is to get data from far enough back in time that there were no infections more than four year ago. The other is to get data far enough back in the life cycle that this holds (i.e. from age groups who are likely to have had their sexual debut four or fewer years ago). Either procedure (I will focus on the first) will make it possible to solve the system by backward induction.

The discussion above focuses on the case in which individuals infected with HIV die only from HIV, which is a simplification. In reality, some of these individuals would have died anyway, even without HIV infection. If the expected non-HIV death rates are known (as they will be from the data used to calculate deaths from HIV, described in the data section below) then adjusting the procedure to take this into account is relatively straightforward. Define the probability that an individual would have died even without HIV  $m$  years after infection as  $q_m$ . In the first year (when all deaths from HIV are due to infections four years before), equation (4) becomes

$$\mu_{i-x,t-x} = d_4 \prod_{i=1}^3 (1 - q_i) b_{i-x-4,t-x-4} \quad (5)$$

This adjustment increases the estimate of HIV infections, since more initial infections are necessary to map into a given number of deaths when we assume that some people who were infected died



before they were killed by HIV. In later years, when deaths depend on more than one year of infection data, the adjustment follows the same principles (although with more elements of the  $q_i$  vector).

Having solved for the  $\mathbf{b}$  vector, calculating incidence and prevalence is straightforward. Number of new infections (incidence) is the elements of the  $\mathbf{b}$  vector. Incidence rate among age  $i$  in year  $t$  is  $b_{i,t}$  divided by the population of age  $i$  in year  $t$ . The calculation of prevalence relies on the  $\mathbf{b}$  vector and the vector of death rates by time from infection. If we denote  $c_i$  the share of individuals still alive  $i$  years after infection, then total infections among age  $i$  in year  $t$  ( $p_{i,t}$ ) is

$$p_{i,t} = b_{i,t} + c_1 b_{i-1,t-1} + c_2 b_{i-2,t-2} + \dots + c_{20} b_{i-20,t-20}$$

Prevalence rate is generated by dividing  $p_{i,t}$  by the population of age  $i$  in year  $t$ .

With noiseless data on mortality rates solving the system of equations for the  $\mathbf{b}$  vector will work perfectly. In practice, however, data on deaths will be noisy. This produces inappropriately high volatility across years – intuitively, a small amount of noise is translated into very large deviations between periods because of the very limited additional information in each of the simultaneous equations.<sup>8</sup>

To avoid this issue, I approach the problem as a minimization with a smoothness restriction, rather than strictly as a set of simultaneous equations. Denoting the matrix of death probabilities ( $d_\alpha$ ) as  $\Theta$  and the vector of death rates ( $\mu_{i,t}$ ) as  $\Lambda$ , I solve the following minimization.

$$\min_{b_1 \dots b_n} ((\Lambda - \Theta \mathbf{b})' (\Lambda - \Theta \mathbf{b})) + \gamma \left( \sum_{i=2}^{n-1} \left( b_i - \frac{b_{i-1} + b_{i+1}}{2} \right)^2 \right) \quad (6)$$

Without the second part (the expression after  $\gamma$ ), this is simply minimized by the solution to the system of equations  $\Lambda = \Theta \mathbf{b}$ . The second element imposes a cost on the function if any given element of the  $\mathbf{b}$  vector differs from the two surrounding values, which will smooth out the elements of  $\mathbf{b}$ ; the degree of smoothness will depend on the size of  $\gamma$ . In the primary analysis, I use a  $\gamma$  value of 0.05. In the section on exports I also explore whether the results are sensitive to higher or lower values of  $\gamma$  (more or less smoothing).<sup>9</sup>

<sup>8</sup>This issue is discussed in more detail, including a simple example, in Appendix A.

<sup>9</sup>Related to the issue of smoothness is the possibility that the function may be minimized when some of the elements of the  $\mathbf{b}$  vector are negative. Since it is not possible for there to be negative HIV infections, I solve the minimization subject to the constraint that all values of  $b_1 \dots b_n$  are positive. Obviously, if the model was frequently suggesting negative values for elements of the  $\mathbf{b}$  vector, this might call into question the technique. In reality, this is extremely unusual, and the unconstrained results are very similar to the constrained results. To the extent that there are differences, they seem to arise only in early years of the epidemic, and the unconstrained results generally do not allow us to reject zero.

## 2.2 Data on HIV Deaths

The methodology described above requires data on deaths from HIV. Because of poor reporting, there are no direct data in Africa on HIV deaths, so deaths from HIV will be calculated by comparing total mortality to expected mortality (based on a non-HIV environment). That is, we will rely on the intuition in Figure 1: because HIV deaths have a very unusual age pattern, if we observe total deaths in an HIV-infected country and total deaths in a country with no HIV but a similar level of development, we can subtract to estimate the number of deaths from HIV.

*Total Mortality in HIV Affected Countries* One possible source for total mortality is official mortality statistics – for example, those reported in the United Nations Demographic Yearbook. Unfortunately, there are no consistent official statistics for Sub-Saharan Africa. As an alternative, I use information from the Demographic and Health Surveys (DHS) sibling mortality histories to calculate death rates. The DHS are household surveys run in a number of African countries beginning in the late 1980s and early 1990s. The focus of the surveys is fertility behavior, maternal and child health. A number of the surveys included “sibling history” modules in which women were asked to list all of their siblings and give information about their gender, their date of birth and date of death (if deceased). Using these reports, it is possible to construct mortality rates over time. For example, if we are interested in the mortality rate of 25 year old men in the last year, we can use these data to figure out how many living 25 year old male siblings the survey participants had one year ago, and then observe how many have died in the last twelve months. Using the same technique, the data can be extended into the past: the mortality rate of men aged 25 ten years ago is constructed by counting the living 25-year-old siblings eleven years ago and then counting deaths over the year after that.

Other researchers have used these data to construct mortality profiles in Africa and argued that, on average, the sibling mortality history data matches relatively well with official data on mortality (Bicego, 1997; Timaeus and Jasseh, 2004; Stanton, Abderrahim and Hill, 2000). However, at least one paper has argued that relative to United Nations life tables, sibling histories underestimate mortality (Gakidou, Hogan and Lopez, 2004), although it is possible this reflects *overstatement* by the UN life tables.. In Appendix B I discuss in more detail the validity of this data and provide evidence that – where we can compare directly (the Philippines and Zimbabwe) – the sibling mortality histories line up very closely with official mortality data. Together with the support from the existing work, this should provide confidence in the use of these data.

*Expected Non-HIV Mortality* The other important component of the calculation of HIV deaths is expected mortality from non-HIV causes. I draw these comparison data on mortality from the United Nations Demographic Yearbook Historical Supplement. I rank country-years from those data based on their child mortality, and use the highest-mortality one-third as the comparison set. I assume that adult mortality in the HIV-affected countries in the absence of HIV would be the same as the adult mortality in this comparison set.<sup>10</sup>

This procedure effectively ignores country-specific non-HIV mortality causes, assuming instead that all countries in the sample have similar non-HIV mortality. Of course, in reality we would expect some differences in the non-HIV mortality rates across areas. This analysis relies on the fact that, on average, total adult mortality in areas not affected by HIV seems to be very low. For prime age adults, the expected death rate is generally between 2 and 4 in 1000, which is very small compared to the estimated death rates in HIV-affected countries (as high as 20 in 1000 in some country-years). This large differences means that small perturbations in the non-HIV mortality rate are unlikely to make a large difference in the estimates.<sup>11</sup>

Web Appendix W.1 (available on the author’s website) shows the all of the mortality data – actual and predicted mortality rates, as well as the excess deaths per 100,000 implied by these figures.

*Existing Data on HIV Prevalence* I compare the HIV rates estimated from the mortality data to two sets of existing HIV prevalence estimates. The first are from UNAIDS. These are reported yearly in the UNAIDS fact sheets and are estimates for adults aged 15-49. These estimates rely on data from studies done in each country, as well as epidemiological models of the epidemic and imputation. Exactly how this is done in each year is not obvious, although the reported prevalence does rely heavily on estimates of HIV rates for women in ante-natal clinics.

The second type of data come from the DHS surveys themselves, and are based on testing done in six of the nine countries in the most recent survey year. The testing is of the general

---

<sup>10</sup>An alternative to using the United Nations data would be to use DHS sibling histories from the few countries with sibling history data that are unaffected by HIV (only Brazil and the Philippines). Although this may have some advantages, it has the disadvantage that the results could be driven by events in those two countries, which is not ideal. The procedure here smoothes such idiosyncratic events over a much larger number of observations.

<sup>11</sup>One thing that is could, in principle, make a large difference here is deaths from war. If the countries in the sample were actively involved in major wars during this time period, this could skew the estimates of HIV rate up, since some of the deaths of prime-age adults would, in fact, have happened in war. Empirically, however, in the countries in this sample, during the sample period, the data suggests there are not many deaths from war (available at <http://users.erols.com/mwhite28/warstat3.htm>). The one place where, early on, there would have been some conflict-related deaths is Northern Uganda. To address this, I have generated data for Uganda excluding the Northern province. This does not change the results or conclusions, either on the export relationship or on the education campaign. Further, war is likely to be a cause of excess deaths primarily for men; the relationships estimated later also hold for women, also suggesting this is not a major issue in these data.

population (all survey respondents are asked to submit to a test), so the estimates should be much more representative and comparable to the mortality-based estimates. The major drawback is that the testing data represent prevalence from a slightly later period than the mortality data (an average of 5 years later) so it is possible that the mortality-based prevalence rates will be somewhat lower. In addition, although these data are clearly more representative than ante-natal clinic testing, they do suffer from relatively high refusal rates (as high as 40% in some groups). This could bias the estimates either upward or downward, depending on the source of the refusals. Despite this, the potential comparability between the mortality-based estimates and the DHS is clearly higher than between the former and the UNAIDS numbers.

### **2.3 Estimates of HIV Prevalence**

Using the methodology and data described above, I generate estimates of HIV prevalence for nine countries in Africa (Burkina Faso, Cameroon, Kenya, Malawi, Mali, Namibia, Uganda, Zambia and Zimbabwe) beginning in the mid-1980s and going through the late 1990s or early 2000s (exact years vary by country). The specific choice of countries is determined by data availability: I use all countries for which there are DHS sibling histories, and therefore mortality data available, starting in the 1990s. Table 1 shows all of the estimates of HIV prevalence generated, for individuals aged 16-35.

In general, the data generated seem to line up well with levels and trends observed in the epidemic over time. The countries which are known to be high HIV in the sample (for example, Zambia and Uganda) have higher prevalence estimates than those known to be low-prevalence (for example, Burkina Faso and Mali). Most countries show increasing prevalence through the 1980s and early 1990s, and then there is some flattening out, consistent with, for example, modeling of the epidemic by UNAIDS. The data also show a large decline in prevalence in Uganda in the early 1990s, which echoes what we see in consistent testing data from that country (i.e. as described in DeWalque (2005)).

These broad observations provide support for the validity of the mortality-based procedure. As a more formal test, I focus on comparing these mortality-based estimates to data from DHS population testing. In particular, I look at the match between age-gender groups in the two samples. This match will only be possible for six of the countries in the sample with DHS-based testing (Burkina Faso, Cameroon, Kenya, Mali, Zambia and Zimbabwe). I generate estimates by five-year age group (15-20, 20-25, 25-30, 30-35, 35-40 and 40-45) for men and women, and match these with the appropriate estimate from the DHS. I generate data for the most recent three years available,

and take the average, to eliminate some of the noise. It is important to recall that the DHS estimates therefore come from an average of 5 years later than the mortality-based estimates so they may, on average, be higher.

The comparison of estimates is shown in Figure 4, which graphs the mortality-based estimates against the DHS estimates and includes the 45 degree line. For a large share of the sample, the mortality-based estimates lie very close to the 45 degree line. The primary exception comes at very high HIV rates, where the DHS estimates are generally somewhat higher than those based on mortality data. This may reflect growth in the epidemic over time (since the testing data are from four years later), or it may reflect underestimates by the mortality data at these high levels (or, less likely, overestimates by the DHS).

Even with the slight mismatching levels at the highest rates, the correlation between the two series is extremely high. This can be seen statistically in Table 2, which shows a regression of the mortality-based prevalence on DHS testing rates by group. Columns 1 and 2 include the whole sample: Column 1 controls for age and gender fixed effects, and Column 2 includes country fixed effects. In both cases, the relationship is positive and highly significant (t-statistic between 10 and 20). Columns 3 and 4 limit the sample to women and men, respectively, and also include age group and country fixed effects. The significant coefficients in these latter columns suggest that the match between the datasets is close enough that it can be identified off of differences in the age pattern of the epidemic in different countries. From the perspective of the export analysis, this close link is quite important. It suggests that the mortality-based estimates are accurately picking up even relatively small variations in HIV prevalence, similar to what we will identify off of in the analysis of the HIV-export relationship.<sup>12</sup>

A second check on the reliability of these data relies on a falsification test: what prevalence does the data generate for a country with virtually no HIV? Similar sibling mortality data is available for the DHS survey in the Philippines. Using the same mortality-based procedure, I estimate an HIV prevalence of 0.1% among 15-25 year-olds and 0.4% among 25-35 year-olds in 1999 (neither is statistically distinguishable from zero), providing further support for this approach. Mortality rates by year for the Philippines are also shown in Web Appendix W.1: the data suggest that actual and predicted death rates are extremely close, with little or no apparent “excess” death.

The close match with the DHS and the falsification test above both support the validity of

---

<sup>12</sup>In the export analysis I will rely on the incidence data that generates this prevalence as the dependent variable. Variations in prevalence are generated by variance in incidence, so matching the small prevalence variations provides confidence in the incidence data.

these mortality-based estimates. However, there are at least two specific patterns in the data that do not seem to match as well with existing evidence, which deserve mentioning. The first is that the data suggest that prevalence increases a bit toward the end of the 1990s in Uganda. Observing Figure 2, we can see this is driven by higher incidence levels between 1993 and 1995. Existing data following antenatal clinics in Kampala and elsewhere (see, for example, Stoneburner and Low-Beer, 2004) does not suggest an increase of this type at the end of the 1990s. This may, of course, reflect errors in the data. It may also be the case that the patterns for women of childbearing age are different than for the overall population. We see some evidence in support of this in Appendix Figure 1, which graphs incidence for the overall population and for women of childbearing age. Women in this group also have an increase toward the end of the 1990s, but it is of much smaller magnitude (for example, compared to the late 1980s) than the overall population. It is possible, therefore, that the differences simply reflect differences across populations.

A second issue is the estimates for Namibia, which are much lower than what we would expect based on testing data. This pattern is more difficult to reconcile. Testing data in Namibia is fairly limited, which may mean it is more selected, but this is largely speculation, and it is also possible that this represents a situation in which mortality reporting does not accurately capture deaths for some reason.<sup>13</sup> Despite this issue, however, the bulk of the evidence in cases where we can make a good comparison *do* point to a very strong relationship between the mortality-based estimates and data from population testing.

Before moving on to exports, it is worth looking at the relationship between the HIV prevalence generated by mortality data and the year-matched prevalence reported by UNAIDS. The recent population testing data from the DHS has suggested that UNAIDS overestimates the prevalence in the general population. I find a similar result comparing with the mortality data. Figure 5 graphs the prevalence generated by mortality data alongside the estimates from UNAIDS. The series are positively correlated, but the UNAIDS numbers seem to be twice or three times as large as the mortality numbers. Interestingly, Uganda is a fairly close match in the two series. This may be due to the fact that Uganda is one of the few areas with reliable population testing early on in the epidemic; this may have allowed UNAIDS to report accurate overall prevalence.

The discussion above, in particular Figure 4 and Table 2, should provide confidence in these mortality-based estimates of HIV rates. The remainder of the paper focuses on the relationship

---

<sup>13</sup>In terms of the analysis of exports, any issue that is specific to Namibia will not affect the results, since Namibia is excluded from the export analysis due to lack of good data on exports.

between the measure of HIV incidence and exports.

### 3 Epidemic Model: Exports and HIV

This section outlines a simple model of the HIV epidemic which makes explicit how HIV and economic activity may be linked. In addition to generating an overall connection between HIV and exports, the model also makes additional predictions about how this relationship should vary across groups and initial conditions. The model builds off three observations. First, truck drivers and other migrants (i.e. those who spend time living or traveling away from home) tend to have more sexual partners than the average in the population (Lurie et al, 2003a; Brewer et al, 1998; Brockerhoff and Biddlecom, 1999; Arnafi et al, 1997; Arnafi, 1993; Orubuloye, Caldwell and Caldwell, 1993). Second, the sexual partnerships these people have away from home tend to be higher risk than those they have at home, largely because their partners are more likely to be infected: for example, are more likely to be bar girls or commercial sex workers (Orubuloye et al, 1993). Finally, the partners (for the most part, wives) of those who travel may be more likely to have additional sexual partners while their spouses are away (Lurie et al, 2003b).

These observations suggest that in times when there are more people traveling, there may be more HIV infections; that link is formalized in the model here. I begin by describing the model setup and then discussing the predictions. At the end of this section I provide evidence for the other necessary link here – namely, a positive relationship between economic activity (exports) and the share traveling away from home.

#### Model Setup

*Sexual Behavior* There are two types of individuals, men and women, all of whom are matched to someone of the opposite sex in marriage. Men may travel (trucking, migrant work), but women do not. Men are of two types: high-sex and low-sex, which captures their desire for non-marital partners (all individuals partner with their spouse). Women are of only one type, but their sexual behavior varies: if their spouse is home, they have  $s$  non-marital partners; if their spouse is traveling, they have  $j > s$  sexual partners. Low-sex type men never travel, and always have  $s$  non-marital partners.<sup>14</sup> When high-sex type men travel, they have  $t$  non-marital partners, who will be drawn

---

<sup>14</sup>The assumption that low-sex types never travel follows directly from an alternative (sensible) assumption that non-marital sex is easier to obtain when you are not living at home. If this is true, then those for whom non-marital sex is a larger perk will be willing to pay more (take a lower wage) for a job that allows them to travel. As long as the share traveling is lower than the share in the high group, no one in the low group will ever travel.

from an outside population. When they are home, they would like to have  $j$  sexual partners, with  $s < j \leq t$  (more than the low-sex type, but less than when they travel, since they are also partnering with their spouse), drawn from the home population. Since the sexual market must clear, however, they may be restricted in their number of partners if there are not enough women with husbands away to match with the high-sex men who are at home. In this case, only some share of the high-sex men will have  $j$  partners, and the rest will have  $s < j$ .

In the calibration, I make assumptions about the values of  $s$ ,  $j$  and  $t$ . Based on overall data from the DHS surveys, which show relatively low rates of non-marital sexual behavior, I assume that  $s = 0.10$ . For individuals in these surveys with at least one non-marital partner (which we will assume are the high types), the average number of partners is 1.5, so I will assume  $j = 1.5$ . In choosing a value of  $t$ , I draw on work which indicates that  $t$  may be very high. For example, in a sample of truck drivers from Nigeria the average number of partners in the last year is 12, and the average current partnerships is 6.3 (Orubuloye et al, 1993). This leads me to consider values of  $t = 6$ ; sensitivity tests suggest that the results are similar with values of  $t$  anywhere between 1.5 and 10.

*Timing* Each period in the model has two sub-periods. In the first sub-period (sub-period  $A$ ), a share  $p$  of individuals are chosen to travel (all from the high-sex group). During this period, these individuals have sex with partners outside of the model; those partners have an average HIV prevalence of  $\hat{h}$ . Individuals who do not travel during this sub-period have sex with their spouse (assuming their spouse is not traveling) and with either  $s$  or  $j$  non-marital partners (depending on their type), chosen randomly from the population of those at home.<sup>15</sup> At the end of this sub-period, infection status is realized based on the sub-period behavior. In the second sub-period (sub-period  $B$ ), no individuals are traveling, and sex happens only with one's spouse.

The timing of the model is intended to capture the idea that individuals who travel tend to return home periodically, and have sex with their spouse. The second sub-period represents this periodic return.

*HIV Transmission* The probability of HIV infection given  $x$  sexual partners, a transmission rate of  $\beta$  and an HIV rate of  $h$  is equal to  $\beta hx$ .<sup>16</sup> In the model, I assume that the HIV transmission rate,  $\beta$ , varies based on the recency of infection, consistent with data on the epidemic. Wawer et al (2005) find transmission rates within 10 weeks after infection are roughly 8 times higher than

<sup>15</sup>Since  $j > s$ , on average people will be more likely to meet a high type for sex than a low type; this is captured in the model.

<sup>16</sup>In fact, the infection probability is  $(1 - (1 - \beta h)^x)$ . However, as long as  $x$  is not too large – i.e. lower than 50 or so – this will be a good approximation, and it simplifies the expression of the comparative statics.



transmission rates more than 10 weeks after infection. I assume a low transmission rate for mature infection (per-partnership transmission rate of 0.01) and 8 times higher for new infections (0.08). New infections are defined as occurring one sub-period before the current period. In other words, if your spouse travels in sub-period 2A and becomes infected, you face a high transmission rate in sub-period 2B. The HIV rate  $h$  among the general population is, of course, calculated based on the observed infections in each period. I assume that the HIV rate among outside partners ( $\hat{h}$ ) is equal to 50%.

*Traveling Calibration* I randomly generate serially correlated values of  $p$  (the share traveling) for 15 periods for 10 “countries”.<sup>17</sup> I seed the model with an HIV rate of 0.005 by assigning 0.5% of the individuals in the sample to be HIV positive before the first period. I then simulate the model forward and generate data on expected HIV prevalence and number of new infections in each period. Using these model-generated “data”, I mimic the regressions I will run later with the actual data. Specifically, I regress the log of new infections on the log of number of people traveling. As with the regressions later, I include period fixed effects, “country” fixed effects and “country”-specific trends.<sup>18</sup>

## Model Predictions

Figure 6 gives a sense of the data generated by the model. This figure graphs the number of infections in each subperiod, for high-sex and low-sex (including women) groups, under the assumption that travel is constant at 2% of men traveling, except in a single year where the share jumps to 10%. The dynamics of the model can be seen clearly. In the sub-period in which travel is elevated, infections are higher for the high sex type because a larger share of them are traveling. Infections are also higher for the non-travelers, since women whose husbands are traveling have more sexual partners; with more men traveling, the number of partnerships at home increases. In the sub-period immediately following the higher travel times, when people have sex only with their spouse, we also see elevated infection rates. This comes from the fact that individuals newly infected in the last sub-period are more infectious, so there is an immediate effect of new infections.

Table 3 estimates this relationship statistically (by period, rather than by subperiod).

Column 1 estimates the baseline relationship: new infections this year regressed on number traveling

---

<sup>17</sup>I do this by randomly generating a beginning value of  $p$  in period 1, uniform over the interval from 0 to 20%. For each year the increase or decrease is equal to 0.03 times a random draw from normal distribution with mean 0 and standard deviation 1.

<sup>18</sup>The use of logs is primarily to parallel the later analysis, in which this is necessary, since the levels of exports and HIV are wildly different across countries, and log-ing both series normalizes them.

this year. The relationship is positive and strongly statistically significant, which is consistent with the model setup, as well as with Figure 6. Columns 2 and 3 consider the issue of lags. Figure 6 shows strong evidence of lags across sub-periods; in fact, there are lags across complete periods as well, as illustrated in Column 2, which includes both travel this year and last year as independent variables. Column 3 regresses new infections on the average of travel this year and last year, which has a positive and significant effect, larger than the effect of either year alone. The regression in Column 3 mimics the regressions that will be run later with the empirical data, which use the average exports over a two year period as the independent variable to allow for lags.<sup>19</sup>

Column 4 of Table 3 includes an interaction between number traveling (in this case, the two year moving average) and the HIV rate in the area. The results suggest another prediction of the model: the effect of increases in the number traveling should increase with the overall HIV prevalence. The interaction between HIV prevalence and number traveling is positive.<sup>20</sup>

Finally, Column 5 of Table 3 shows the relationship between traveling and HIV infections just among those who do not travel. The relationship is still positive and significant, but slightly smaller in magnitude than the relationship with overall infection rates (i.e. small compared to Column 3). This suggests that for those not involved in migration (for example, women) the relationship between infections and economic activity should still appear, and be positive, but may be smaller.

Based on Table 3 there are a number of predictions about the data. If this model accurately captures the world, we should expect to see a positive relationship between travel and HIV, positive interaction with existing infection rates, and smaller (but still positive) effects for those who do not travel or migrate. The regressions in the next section, of course, focus on exports rather than the share traveling as the independent variable. The subsection below therefore discusses briefly the evidence for a link between exports and travel.

## **Exports and Travel**

Thus far, I have discussed specifically two types of travel – migration for work and truck driving. For migration, data that would test for a connection between HIV and economic activity are very limited. Certainly in specific cases (for example, mining in Zambia) we know that increases in the

---

<sup>19</sup>When using actual data, the average also captures the fact that the timing of both measured exports and HIV infections is likely to be imprecise.

<sup>20</sup>The level effect of HIV is negative, although this is driven by the inclusion of country-year trends, which make it difficult to interpret the level effect in this type of generated data where the correlation between level and time is very high.

demand for copper increase the demand for migrant labor to work in the mines. In agricultural areas this link may be less strong, but there are little data to test it in any context.

In the case of trucking, however, it is possible to get a sense of how the volume of goods trucked, or number of trucks directly, varies with exports. In the first case, the World Development Indicators report data on export value and on the volume of goods trucked by country-year for a number of countries worldwide (although generally not those in the sample used here). Column 1 of Table 4 shows the relationship between log export value and log volume of goods transported, with controls for country and year fixed effects, plus country-specific trends. The coefficient is positive and significant, with a magnitude of around one. This suggests that doubling exports approximately doubles the volume of goods trucked, which is not surprising.

Within the sample of countries used here, we can look directly at the relationship between volume of exports (from the data discussed below) and the weight of trucks and truck parts imported (also from the data below). If more exports mean more demand for trucks, this will translate into more demand for truck drivers. The relationship between log export volume and log number of imported trucks is shown in Column 2 of Table 4. Here, the magnitude is around 0.70, slightly smaller (although the point estimate is noisier) but still positive and significant.

In general, these results provide evidence in favor of the link between exports (economic activity more generally) and trucking and, combined with the results in Table 3, a theoretical link between exports and HIV.

## 4 HIV Incidence and Exports

This section begins by presenting the primary results in the paper, showing regressions of the number of new HIV infections on exports. I then discuss the mechanism and identification and argue that there is evidence that the relationship in the first subsection is being driven by the mechanism described in Section 3. The third subsection briefly discusses the robustness of the results to variations in the assumptions in Section 2 used to generate incidence.

The data on HIV used is described above. For exports, I rely on the NBER-United Nations Trade Data (NBER data) from Feenstra et al (2004).<sup>21</sup> I use several measures of exports. First, for the period from 1980 onward I sum the value of all exports reported in the database. However, the reporting of some exports is inconsistent over time, leading to concerns that this methodology

---

<sup>21</sup>The data for Namibia are extremely limited, so it was necessary to leave this country out of the analysis.

produces noisy exports measures. I therefore also generate two measures based on the major exports only, beginning by identifying the major exports in each country based on these data. I aggregate the export data for 1997, 1998 and 1999 by country. For each country, I then identified the top two exports, by value (in nominal US dollars), over that three year period. I compared the value of these exports to the total export value. For countries in which the top export represents at least 50% of the total export value (Burkina Faso, Malawi, Mali and Uganda) I focus only on this top export. For the remaining countries I focus on both of the top two exports. In all cases this covers at least 50% of the total value of exports from the country, and in most countries it is much higher than that. I then use these same NBER data to calculate the trade value and volume beginning in 1983 (the first year in which the volume data are available) for these top exports in each country.

#### 4.1 Primary Results

Table 5 shows the baseline relationship between log of new HIV infections and the three measures of exports used. As discussed, in all case the measure of exports is the log of the average of the current and previous year’s exports, to capture the possibility of a lag in the relationship between exports and HIV, as well as eliminate some of the noise in the timing.<sup>22</sup> Column 1 shows the relationship between HIV infections and the value of all exports reported in the NBER database; Columns 2 and 3 show the relationship with export value and volume of the most important exports, respectively, from the NBER data. In all three columns there is a highly significant positive relationship between exports and HIV. In terms of magnitude, the results suggest that doubling exports will lead to around a quadrupling of new HIV infections.

Figure 7 shows these results graphically. For each country, the residual of new HIV infections and export volume (after removing country and year fixed effects and country-specific trends) are graphed over time. In most cases, the link between the two series is clear – for example, in Burkina Faso we see (again, with the trends removed) both series increasing, then decreasing, then increasing again.

Panels A and B of Table 6 turn to testing two of the other predictions based on the model. Panel A replicates Table 5, but includes an interaction between the level of HIV prevalence and export volume. As predicted by the model, the interaction between exports and baseline prevalence is positive and significant: the effect of changes in exports is larger in years when the baseline rate of

---

<sup>22</sup>Appendix Table 1 shows this regression with separate controls for this year’s exports and last years; in all three cases, both are significant.

HIV is higher. In Panel B, the dependent variable is infections for women only. The model suggested that for individuals who did not travel, the effect of increases in exports should be positive, but smaller. Indeed, the coefficients here are positive and significant, but smaller (on average) than those seen in Table 5.

The relationship estimated here is quite large in magnitude and, hence, the effect of changes in exports overall on HIV may be big. Based on the coefficients in Table 5, I create a simple counterfactual: if exports in all years in the sample were lower by 25%, what effect would that have on HIV infections? Although the exact value depends on which measure of exports is used, on average the counterfactual suggests that a 25% decline in exports would mean a reduction of roughly 60% in the number of new infections. This effect is somewhat larger than what we would expect based on the model in Section 3. However, given that the model is fairly stylized, it is difficult to draw strong conclusions about the expected magnitude of the relationships in the data. The qualitative predictions are likely to be more portable.

## 4.2 Mechanisms and Identification

The results in Tables 5 point to a relationship between HIV infection and exports, and the evidence in Table 6 is consistent with the secondary predictions of the model in Section 3. However, these results may also be consistent with other mechanisms. For example, increases in overall income may be correlated with increases in exports, but may also increase individual ability to purchase sex, which could increase HIV. Alternatively, the introduction of a government that focuses on economic growth and not on HIV prevention could, in principle, drive both of these effects. This subsection focuses on linking the reduced form export-HIV connection more directly to the transit mechanism, and (broadly) making the case for a causal effect of exports.<sup>23</sup> I provide a number of pieces of evidence in favor of an export mechanism – exploiting variations across areas and across export types, instrumenting for exports with commodity prices and looking directly for a link with trucking specifically.

It is worth noting that GDP is perhaps the most obvious confound here. Given that, one simple approach would be to control for GDP (for example, from the World Development Indicators) in the regressions shown previously. It is worth noting that in at least one salient case – Uganda – the overall economy was not doing well in the late 1980s, although exports were high (Figure 2), and

---

<sup>23</sup>It is worth noting first that the controls for country-specific trends in Table 5 rule out some alternative stories. Specifically, any changes which will lead to gradual changes over time (like a new government) are not going to drive the results.

HIV seems to follow exports, not GDP. More generally, inserting GDP as a control into the previous regressions makes almost no difference to the results (for example, the coefficient on export volume in Table 5 moves from 2.79 to 2.96 with the same level of significance), and the control for GDP is not significant (results not shown). This test is not, however, sufficient in general since official measures of GDP are so flawed and inaccurate. So, while this simple test certainly does not support a role for GDP rather than exports, it is difficult to conclude anything firm.

*Variations Across Region: Road Traffic* To the extent that the relationship described above is a product, in part, of increases in road traffic, we would expect it to be stronger in areas where changes in economic activity have a larger effect on the movement of people. This is likely to be urban areas and areas with a greater density of roads. In the extreme, areas with no roads are unlikely to be affected much by an increase in traffic. To test this hypothesis, I generated incidence estimates over time for each region in the countries in the sample.<sup>24</sup> I then explore how the relationship between incidence and exports varies with roads and urbanization.<sup>25</sup>

Panel A of Table 7 shows the relationship between exports and HIV in regions that do, or do not, contain the capital city. Columns 1 and 2 focus on the relationship between HIV and export value, with Column 1 including regions that contain the capital and Column 2 other regions. Columns 3 and 4 focus on the relationship between HIV and export volume, with Column 3 including regions with the capital and Column 4 the other regions. Although in the case of export volume the coefficients are very similar, with export value the relationship is much larger in areas that contain the capital.

Panel B of Table 7 looks at variation in the magnitude of the relationship based on road density, only for areas outside of the capital region. Panel B mimics Panel A, splitting areas based on whether they have the highest road density (Columns 1 and 3) or not (Columns 2 and 4). Here, for both export value and volume, the relationship with HIV incidence is larger in areas with higher road density. The data here are somewhat noisy, since regions tend to be relatively small and there are only a limited number of deaths per region in a given year. Nevertheless, the direction of the effects support an interpretation consistent with the mechanism described in Section 3.

*Variations Across Export Type* A second source of variation, this time across countries

---

<sup>24</sup>The procedure for this is the same as described above with one additional assumption. Respondents are asked only about their own region of residence, not about their siblings. It is necessary, therefore, to assume that individuals live in the same region as their siblings. Obviously this will not always be true, which may introduce noise into the data. However, as long as there is correlation between sibling region of residence (which is presumably the case) then there is some signal included here, making this analysis appropriate even if the results may be attenuated.

<sup>25</sup>Data on road density comes from the ArcGIS database, which provides information on square miles of area and miles of roads by region. Dividing these gives a crude measure of road density.

rather than regions, is export type. Major exports vary across countries in this sample: to the extent that the primary export is more or less dependent on trucking, we would expect the relationship between HIV and exports to be more or less strong.

To identify which exports are linked to trucking, I rely on input-output matrices. The ideal would be to have these matrices from the sample countries, or at least from Africa. With detailed enough input-output data I could identify the contribution of trucking to each export category. Input-output matrices from Africa are, however, limited, and those that are available do not differentiate trucking from other transport (i.e. trains, air transport, pipelines), making it difficult to use them for this purpose. Instead, I use input-output matrices from the United States and calculate, for each export type, the percentage contribution of road trucking to total output. Although, of course, using data from the United States is not ideal, the hope is that with broad categories the patterns will be the same.<sup>26</sup>

Most of the countries in the sample export either agricultural products (coffee, tea, cotton) or forestry products (logging). In the input-output matrices from the United States, trucking seems to make similar contribution to all of these industries, so I combine these countries into one “farming” category. Two of the sample countries, however, have different export types. Zambia exports primarily mining products. In the input-output matrices from the United States, trucking contributes a much larger share to mining than to farming. Cameroon, in contrast, exports primarily oil, for which trucking is only a very small input. I therefore separate these countries out, and run three regressions: HIV on exports for the sample minus Zambia and Cameroon, HIV on exports for Zambia and HIV on exports for Cameroon; in each of the latter cases, I include a control for a linear time trend.

These regressions are reported in Table 8. I limit to value and volume for major exports since this hypothesis focuses on variations in those exports. In both cases (for value and volume) we see the expected patterns. The relationship for the sample minus Zambia and Cameroon is similar to what we see overall and is positive and significant. For Zambia, where the major export involves significantly more trucking activity, the relationship is much larger in magnitude. For Cameroon, where the major export does not involve much trucking, the relationship is actually *negative*, and we can reject a sizable positive relationship. To some extent, we can see these differences in the graphs in Figure 7. For Zambia, the sharp drop in exports accompanies the sharp drop in HIV infections

---

<sup>26</sup>An alternative way to explore the relative contribution of trucking is to look across countries at the correlation between number of trucks imported and exports. This yields similar results.

toward the end of the sample. In Cameroon, if anything the two series seem to move opposite from one another.

It is important to note, of course, that at some level the results here simply suggest that two countries differ from the rest, and this could be for reasons unrelated to the differences in export type. Nevertheless, the results are at least supportive of the overall hypothesis that exports drive trucking and trucking, in turn, drives HIV prevalence.

*Instrument for Exports with Commodity Prices* Moving away from variations across space, an alternative way to identify whether we are truly observing a relationship with exports is to use an instrument. I instrument for export volume with world commodity prices. Short run variations in world commodity price will drive exports (this is especially true here since exports from the countries in this sample tend to be raw commodities like coffee and tea), but are less likely to drive things like GDP (except through exports) and should not affect governance or other internal variables in the very short run. Using these prices as an instrument therefore allows me to uncover the causal effect of exports. To perform the instrumental variable analysis, for each country, I generate a price measure that is a weighted average of the world prices for each export, with the weights determined by the share of the total exports accounted for by each commodity. Historical commodity prices are drawn from the IMF's International Financial Statistics database (year 1980-2001).

Column 1 of Table 9 shows the first stage regression of export volume (average of this year and last year) on commodity price (average of last year, two years ago and three years ago, to allow for lags). The coefficient is positive and highly significant. Note that since the dependent variable is export *volume*, not value, this relationship is not mechanical. Column 2 of Table 9 shows the second stage reduced form relationship, also positive and significant: increasing the average world prices of a country's major exports increases the number of new HIV infections. Column 3 shows the IV coefficient, which is very large and also significant. This coefficient seems much too large, which may well be due to a too-small magnitude in the first stage, driven by noise in the export and price data. However, the positive and significant relationship, especially in the reduced form, provides significant support for the role of exports and, by extension, the trucking/migration mechanism.

*Truck Imports* As a final mechanism test, I can test directly the link between HIV and trucks imports. Table 4 finds a positive relationship between truck imports and export volume; the same data on imports can be used to look directly at the relationship between HIV incidence and number of trucks. Column 1 of Table 10 estimates the relationship between new HIV infections this year and the log weight of of trucks and truck parts imported this year. The coefficient is positive



and significant, with a magnitude of around 0.62, which suggests that doubling the number of trucks imported leads to about a 60% increase in the number of new HIV infections.

*All Imports* Although it is not necessarily a direct test of the mechanism, it is worth noting that the mechanism described here may also suggest a relationship between *imports* (as opposed to exports) and HIV. Since imports are transported around the country in the same way exports are, in principle they should have a similar effect. There are at least two reasons, however, why we may expect a weaker relationship with imports than with exports. First, since urban areas are richer, imports will be disproportionately consumed in these central areas. Relative to exports, which require movement from more remote areas of production (agricultural areas, mines) to denser areas, imports are more likely to simply arrive in the denser areas and stay. The link to transport traffic is therefore likely to be weaker (indeed, there is no significant relationship between import volume and truck imports in these data). Second, the measure of imports over time is much noisier than the measure of exports, since the exports of these areas focus largely on only a few easily tracked items, but the imports cover a wide variety of goods.

With these caveats in mind, Columns 2 and 3 of Table 10 shows the relationship between new HIV infections and imports, again focusing on the major imports for each country. The results are smaller and statistically weaker than in Table 5, but they are both positive and significant at at least the 10% level.

In sum, although it is difficult to *prove* that the mechanism described in Section 3 is responsible for the overall relationship seen in Table 5, the results here provide significant support for that mechanism, particularly over an alternative story that both exports and HIV are driven by some third variable.

### 4.3 Robustness to Assumptions Generating Incidence

One important concern with all of the analyses above is that the relationship between HIV and exports is being driven by the assumptions that generate incidence, rather than by the underlying relationship. This section briefly considers the relationship between HIV and exports if we vary three of the assumptions: the assumption about the degree of smoothing ( $\gamma$  in equation 5), the assumption about time to death after infection, and the assumption about the level of non-HIV mortality.

The value of  $\gamma$  used in the main results is 0.05. I re-estimate the regressions if this is varied either up or down ( $\gamma = 0.01$  or  $\gamma = 0.5$ ). Panel A of Table 11 shows the coefficients from regressions of the form reported in Table 5, with either of these alternative assumptions. The coefficients remain

significant and of similar magnitudes, even with these large changes in the smoothing parameter.

To address the issue of time to death, I generate incidence based on two alternative paths of time to death. One of the alternative paths assumes a faster time to death: the faster time path assumes that everyone has the time to death profile of the oldest age group. The other alternative path models time to death as flatter but with similar speed: 8% of people die every year between 5 and 14 years after infection, rather than the more peaked shape observed in Figure 3 (both alternative paths are illustrated in Appendix Figure 2). Panel B of Table 11 shows the coefficients on exports in regressions using each of these alternative assumptions. Although the coefficients differ a bit across specifications, they are all in a similar range, positive and significant. Of course, while these results in Panel B suggest that the HIV-export connection is robust to small changes in assumptions about time to death, they are unlikely to be robust to large changes in these time paths. For example, the results are likely to be quite different if we assume that all deaths happen in years 9 and 10. The conclusions here rest on the claim that, broadly, the time path assumed is a reasonable one.

Finally, Panel C of Table 11 explores whether the coefficients vary if we change the assumed level of non-HIV mortality. In particular, I consider what happens if we assume *no* non-HIV mortality (i.e. all observed deaths are due to HIV) and if we assume that non-HIV mortality is 25% higher than in the baseline specification. Again, as in Panels A and B, the results are very similar with these changes.

Overall, the results in Table 11 indicate that while the exact magnitude of the relationship between HIV and exports is obviously affected by the assumptions about generating incidence, the general conclusions of this section are generally robust to changes in any of these three elements of the estimation.

## 5 Application: Prevalence Declines in Uganda

The previous section provides evidence that, at the macro level, new HIV infections and economic activity are linked. This section considers whether that link may be at least partially responsible for one of the more striking patterns in HIV prevalence in Africa over the 1990s.

HIV prevalence declined sharply in Uganda in the early 1990s, a decline unique among Sub-Saharan African countries (this can be seen in Table 1, as well as in existing data). This decline is generally attributed to anti-HIV educational efforts, including those by the Uganda National AIDS

Program (NACP) and the ABC campaign, both of which encouraged changes in sexual behavior, including partner reduction, “zero-grazing”, abstinence and condom use (Green, 2004; Green et al, 2006; Slutkin et al, 2006).<sup>27</sup> Among the more important components of the educational efforts were in-school education, communication with faith-based organizations and billboards and other forms of public advertising (for a good summary of the specific aspects of ABC, see Green, 2004; Slutkin et al, 2006 provide a summary of non-ABC partner reduction efforts, for example, NACP). Recent work has argued that changes in concurrency as a result of this program were important in the decline in HIV (Epstein, 2007).

Uganda was among the first countries to address the epidemic head on and the subsequent decline in prevalence was seen as a signal of the success of the behavioral education approach. Subsequently, a number of other countries adopted this type of prevention strategy (Kenya, Tanzania, Rwanda, Malawi and others). Outside aid groups have pursued similar interventions. PEPFAR, in particular, has adopted this campaign as a major part of their anti-HIV strategy: roughly one hundred million dollars of their funding was spent on ABC programs in 2004.

However, in the early 1990s Uganda also saw a major decline in exports. This was likely due in large part to a decline in coffee prices and production.<sup>28</sup> Figure 2 demonstrates that the decline in export value lines up extremely closely with the changes in incidence in Uganda, mimicking both the decline in the early 1990s and the increase afterwards; in Figure 7, we can see also that the decline in export volume lines up closely with this decline (in other words, volume of exports also fell when prices did).<sup>29</sup>

Given the relationship estimated in Section 5, the decline in export volume in Uganda during this period suggests that HIV incidence would have declined somewhat even without any education campaign. Using the coefficients from Table 5, it is possible to evaluate the expected magnitude of the decline between the 1980s and early 1990s due solely to changes in exports. The methodology for this calculation is straightforward. Based on the regressions behind Table 5, I predict the expected new HIV infections in the late 1980s in Uganda and in the early 1990s. Subtracting these two predicted values generates the predicted decline based on exports; by

---

<sup>27</sup>Although there is a fair amount of agreement on the success of education in general in Uganda (although see Gray et al, 2006), there is less agreement on the relative success of the partner reduction elements, versus ABC specifically, and within the ABC on the various merits of A versus B versus C; for a summary and citations on this controversy, see Green et al, 2006.

<sup>28</sup>There was also a significant decrease in airport usage over this period, after Uganda Airlines shut down. Between 1985 and 1991, the number of flights out of Entebbe fell from 717 to 205.

<sup>29</sup>This change in exports did not reflect a collapse of the Ugandan economy in general. It was driven by a world coffee price shock, which was caused by factors outside of Uganda, and decreased exports.

comparing this to the actual decline, we can calculate what share would be expected based on changes in exports alone.

The results are shown in Table 12. The share explained depends on the choice of export measure, and the range is between 30% and 60%. It is worth noting that the changes in exports over time also match with the apparent increase in incidence in the 1993-1995 period, which is seen in Figure 2, and perhaps reflected in other literature suggesting an increase in prevalence toward the end of the decade (Shafer et al, 2006). This increase is harder to explain with the education campaign, which continued at similar levels (if anything, is likely to have increased) during this period (although others have argued that this is due to less preventative behavior after prevalence declines).

It is important to note, of course, that the decline in exports certainly does not explain all of the decline in prevalence in Uganda in the 1990s; there is clearly still a role for changes in the choices of sexual behavior. Further, in some sense the decline due to exports comes from the same underlying source as the decline due to an education campaign: a drop in sexual behavior. The crucial distinction between the two, however, is that the success of an education campaign lies in changing individual (unconstrained) *choice* of non-marital partners, whereas the “success” of a decline in exports lies in changing individual *opportunities* for such relationships.

This result may have important implications for overall HIV policy in Africa. If this education campaign was half as successful as previously thought there may well be other policies which overtake it in terms of cost-effectiveness. In particular, policies that seek to change viral transmission rates (for example, male circumcision) may appear much more attractive in light of these findings. Funding for HIV in Africa is, unfortunately, too limited to make all types of interventions feasible simultaneously. Understanding the relative cost-effectiveness of various inputs is therefore crucial to formulating optimal policy.

## 6 Conclusion

This paper addresses the connection between HIV and economic activity, arguing that increases in economic activity actually significantly increase HIV incidence. This result – particularly the magnitude – may have important implications for HIV prevention. Interventions that target those involved in trucking, or those more generally targeted at migrants, may have an even larger effect than previously expected. Further, HIV is only one focus of development aid to Africa. Significant

amounts of aid are also spent trying to increase trade. Policies like the African Growth and Opportunity Act (AGOA) are designed specifically to encourage exports of African textiles, and smaller efforts, like those made by Fair Trade, also aim to increase trade and contact with the west. The results here suggest that, while those policies may well improve trade, they may actually make the HIV epidemic worse. Combining other prevention efforts (male circumcision, for example) with trade policies may ameliorate some of these interaction concerns. It is, of course, worth noting that in the long run increases in economic growth are likely to help the HIV epidemic, both because people who are richer have more incentive to avoid the virus (Oster, 2007) and because health care improves with economic growth, and with that improvement will come changes in the ability of these countries to deal with the virus. It is in the more short run that the negative interactions are most salient.

Outside of the specific analysis of HIV and economic activity, this paper provides a new set of estimates of HIV prevalence and incidence that could be used to analyze the connection between HIV and other outcomes – for example, fertility or human capital attainment. Existing papers on these topics (Kalemi-Ozcan, 2006; Young, 2006; Forston, 2006) generally rely on somewhat flawed data from testing of pregnant women over time or on model-generated estimates of prevalence. The estimates here are likely to be superior to either – they have the consistency of model-generated data, but are based on actual data in the population rather than assumptions about epidemic growth over time.

## References

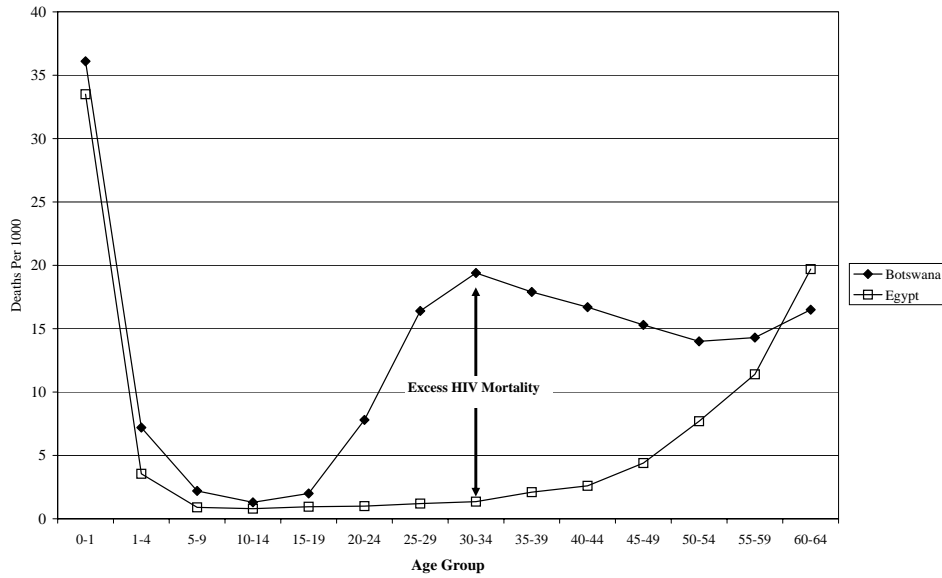
- Altman, Lawrence**, “AIDS Is on the Rise Worldwide, U.N. Finds,” *The New York Times*, 2006, November 22, 2006.
- , “Report Shows 2005 to Be ‘Least Bad Year’ of AIDS Epidemic,” *The New York Times*, 2006, May 31, 2006.
- Anarfi, John, Ernest Appiaha, and Kofi Awusabo-Asareb**, “Livelihood and the risk of HIV/AIDS infection in Ghana: the case of female itinerant traders,” *Health Transition Review*, 1997, 7-5.
- Arnafi, John**, “Sexuality, migration and AIDS in Ghana-A socio-behavioral study,” *Health Transition Review*, 1993, 4, 273–295.
- Bicego, George**, “Estimating adult mortality rates in the context of the AIDS epidemic in sub-Saharan Africa: analysis of DHS sibling histories,” *Health Transition Review*, 1997, 7.
- Brewer, Toye, Julia Hasbun, Caroline Ryan, Stephen Hawes, Samuel Martinez et al.**, “Migration, ethnicity and environment: HIV risk factors for women on the sugar cane plantations of the Dominican Republic,” *AIDS*, 1998, 12, 1879–1887.
- Brockerhoff, Martin and Ann Biddlecom**, “Migration, Sexual Behavior and the Risk of HIV in Kenya,” *International Migration Review*, 1999, 33 (4), 833–856.
- Collaborative Group on AIDS Incubation and HIV Survival**, “Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis,” *The Lancet*, 2000, 355, 1131–1137.
- Deuffic-Burban, Sylvie and Deminique Costagliola**, “Including Pre-AIDS Mortality in Back-Calculation Model to Estimate HIV Prevalence in France, 2000,” *European Journal of Epidemiology*, 2006, 21, 389–396.
- DeWalque, Damien**, “How Does the Impact of an HIV/AIDS Information Campaign Vary With Educational Attainment: Evidence from Rural Uganda,” 2005. Mimeo, World Bank.
- Editorial Desk**, “The State of AIDS,” *The New York Times*, 2005, December 1, 2005.
- Epstein, Helen**, *The Invisible Cure*, New York, NY: Farrar, Straus and Giroux, 2007.
- Feeney, Griffith**, “The Impact of HIV/AIDS on Adult Mortality in Zimbabwe,” *Population and Development Review*, 2001, 27 (4), 771–780.
- Feenstra, Robert, Robert Lipsey, Haiyan Deng, Alyson Ma, and Hengyong Mo**, “World Trade Flows: 1962-2000,” *NBER Working Paper no. 11040*, 2004.
- Forston, Jane**, “Mortality Risks and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa,” 2006. Mimeo, Princeton University.
- Gakidou, Emmanuela and Gary King**, “Death by Survey: Estimating Adult Mortality without Selection Bias from Sibling Survival Data,” *Demography*, 2006, 43 (3), 569–585.
- , **Margaret Hogan, and Alan D Lopez**, “Adult mortality: time for a reappraisal,” *International Journal of Epidemiology*, 2004, 33, 1–8.

- Gray, Ron, David Serwadda, Godfrey Kigozi, Fred Nalugoda, and Maria Wawer**, “Ugandas HIV Prevention Success: The Role of Sexual Behavior Change and the National Response. Commentary on Green et al. (2006),” *AIDS and Behavior*, 2006, 10 (4), 347–350.
- Green, Edward**, *Rethinking AIDS Prevention: Learning from Successes in Developing Countries*, Westport, CT: Praeger Publishers, 2004.
- , **Daniel Halperin, Vinand Nantulya, and Janice Hogle**, “Ugandas HIV Prevention Success: The Role of Sexual Behavior Change and the National Response,” *AIDS and Behavior*, 2006, 10 (4), 335–346.
- Halperin, Daniel and Glenn Post**, “Global HIV Prevalence: The Good News Might be Even Better,” *Lancet*, 2004, 364 (9439), 1035–1036.
- Kalemi-Ozcan, Sebnam**, ““AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa”,” *NBER Working Paper, No. 12181*, 2006.
- Klitsch, M**, “Rural Ugandan Women’s HIV Infection Rates Seem Related to Truck Routes,” *International Family Planning Perspectives*, 1992, 18 (2), 79.
- Liao, J and R Brookmeyer**, “An Empirical Bayes Approach to Smoothing in Backcalculation of HIV Infection Rates,” *Biometrics*, 1995, 51 (2), 579–588.
- Lurie, Mark, Brian Williams, Khangelani Zuma, David Mkaya-Mwamburi, Geoff Garnett, Michael Sweat, Joel Gittelsohn, and Salim Abdool Karim**, “Who Infects Whom? HIV-1 Concordance and Discordance Among Migrant and Non-Migrant Couples in South Africa,” *AIDS*, 2003a, 17 (15), 2245–2252.
- Lurie, MN, BG Williams, K Zuma, D Mkaya-Mwamburi et al.**, “The Impact of Migration on HIV-1 Transmission in South Africa: A Study of Migrant and Nonmigrant Men and Their Partners,” *Sexually Transmitted Diseases*, 2003b, 30 (2), 139–156.
- Maugh, Thomas**, “AIDS Growth Slowing Worldwide, U.N. Finds,” *The Los Angeles Times*, 2006, May 31, 2006.
- McNeill, William**, *Plagues and Peoples*, Garden City, NY: Anchor Press, 1976.
- Murphy, Kevin and Finis Welch**, “Empirical Age-Earnings Profiles,” *Journal of Labor Economics*, 1990, 8 (2), 202–229.
- Ong, HC, SH Quah, and HC Low**, “An Application of the Backcalculation Method to Estimate Past HIV Infection Rates in Malaysia,” *Medical Journal of Malaysia*, 1998, 53 (4), 385–391.
- ORC Macro**, “HIV Notes from Measure DHS,” Technical Report, USAID 2006.
- Orubuloye, I.O., Pat Caldwell, and John Caldwell**, “The Role of High-Risk Occupations in the Spread of AIDS: Truck Drivers and Itinerant Market Women in Nigeria,” *International Family Planning Perspectives*, 1993, 19 (2), 43–48.
- Oster, Emily**, “HIV and Behavior Change: Why Not Africa?,” 2007. Mimeo, University of Chicago.
- Shafer, L.A., S. Biraro, A. Kamali, H. Grosskurth, W. Kirungi, E. Madraa, and A. Opio**, “HIV prevalence and incidence are no longer falling in Uganda a case for renewed prevention efforts: evidence from a rural population cohort 1989-2005, and from ANC surveillance,” *Abstract at XVI International AIDS Conference*, 2006.

- Slutkin, Gary, Sam Okware, Warren Naamara, Don Sutherland, Donna Flanagan, Michel Carael, Erik Blas, Paul Delay, and Daniel Tarantola**, “How Uganda Reversed Its HIV Epidemic,” *AIDS and Behavior*, 2006, 10 (4), 351–361.
- Stanton, Cynthia, Nouredine Abderrahim, and Kenneth Hill**, “An Assessment of DHS Maternal Mortality Indicators,” *Studies in Family Planning*, 2000, 31 (2), 111–123.
- Statistics South Africa**, “Mid-year population estimates, South Africa,” Technical Report, Statistics South Africa 2004.
- Steinbrook, Robert**, “HIV In India – A Complex Epidemic,” *New England Journal of Medicine*, 2007, 356 (11), 1089–1093.
- Stoneburner, Rand and Daniel Low-Beer**, “Population-Level HIV Declines and Behavioral Risk Avoidance in Uganda,” *Science*, April 2004, 304, 714–718.
- Stover, James**, “AIM version 4. A computer program for HIV/AIDS projections and examining the social and economic impacts of AIDS,” *Spectrum system of Policy Models. The Futures Group International.*, 2003.
- Timeaus, Ian and Momodou Jasseh**, “Adult Mortality in Sub-Saharan Africa: Evidence from the Demographic and Health Surveys,” *Demography*, 2004, 41 (4), 757–772.
- Trussell, James and German Rodriguez**, “A Note on the Sisterhood Estimator of Maternal Mortality,” *Studies in Family Planning*, 1990, 21 (6), 344–346.
- UNAIDS**, *AIDS Epidemic Update: Special Report on HIV/AIDS: December, 2006*, UNAIDS and WHO, 2006.
- United Nations**, *Demographic Yearbook*, United Nations, 2001.
- Wawer, Maria, Ronald Gray, Nelson Sewankambo et al.**, “Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda,” *Journal of Infectious Diseases*, 2005, 191, 1403–1409.
- Weinreb, Alex**, “Estimating HIV+ Incidence from Mortality Rates: A Method and Agenda,” 1999. Presentation, Durbin South Africa.
- Werker, Eric, Amrita Ahuja, and Brian Wendell.**, “Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis,” 2006. Harvard Business School Working Paper.
- Wilson, Mary**, “Travel and the Emergence of Infectious Disease,” *Emerging Infectious Diseases*, 1995, 1 (2), 39–46.
- Young, Alwyn**, “In Sorrow to Bring Forth Children: Fertility amidst the Plague of HIV,” 2006. Mimeo, University of Chicago.

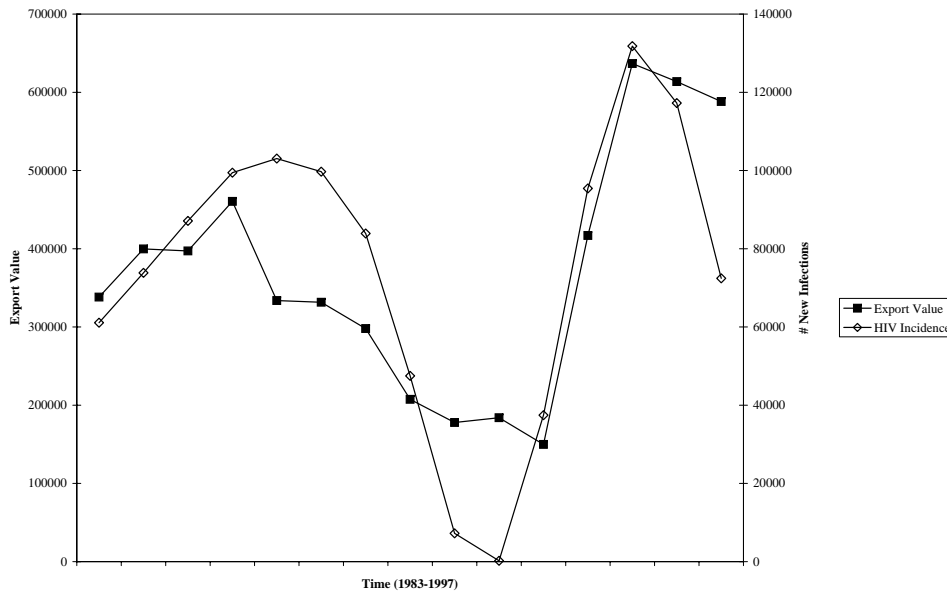


**Figure 1:**  
**Death Rates by Age, Botswana in 2001 and Egypt, 1990s**



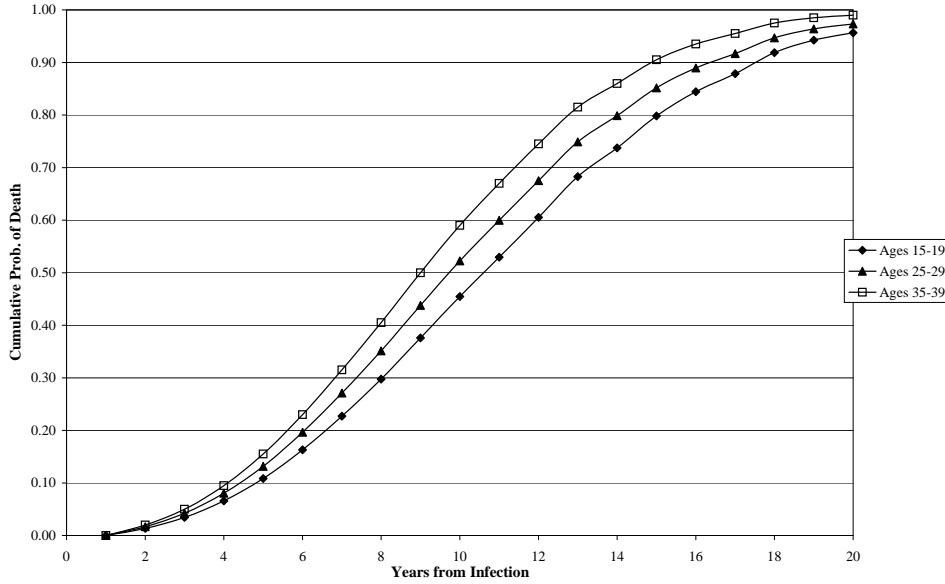
Notes: Death rates are from the United Nations Demographic Yearbook, 2002 edition and Historical Supplement. The Botswana data is for 2001, and the Egypt data is an average for the 1990s. Death rates are deaths (all individuals, both genders) per 1000 people in the age group. Botswana is affected by HIV; Egypt is largely not in this period.

**Figure 2:**  
**HIV Incidence and Export Value in Uganda**



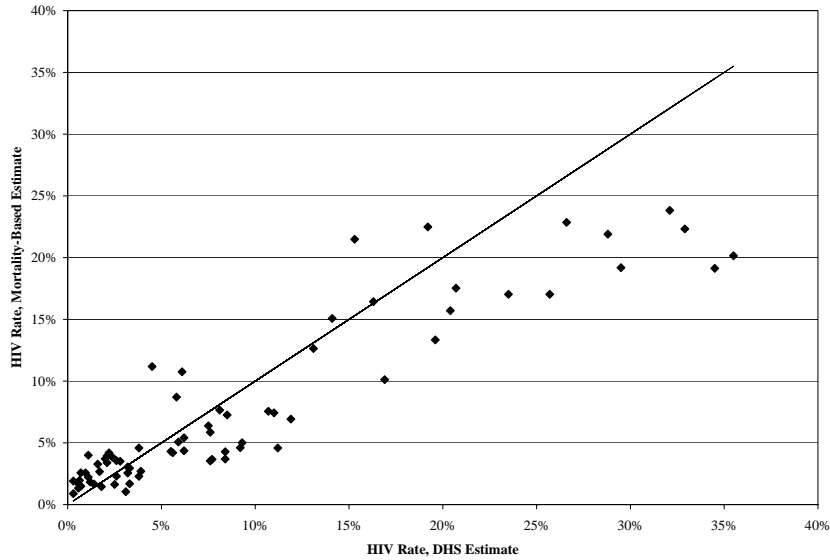
Notes: This figure shows export value for Uganda and HIV incidence for Uganda, as calculated using the mortality-based methodology in this paper for 1983 through 1997. Export value comes from the NBER trade database (sourced from Comtrade, described in Feenstra et al, 2004).

**Figure 3:**  
Distribution of Time from HIV Infection to Death



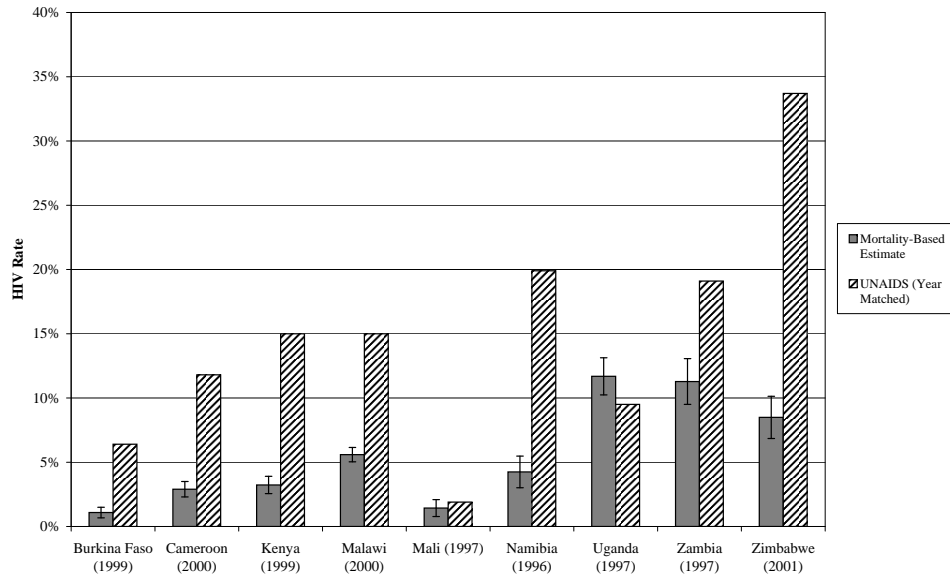
Notes: Figure presents the cumulative probability of death as a function of years from infection with HIV, by age group. The figure is based on data from Collaborative Group on AIDS Incubation and HIV Survival, 2000.

**Figure 4:**  
Mortality-Based Estimates and DHS Testing Estimates by Country-Gender-Age Group



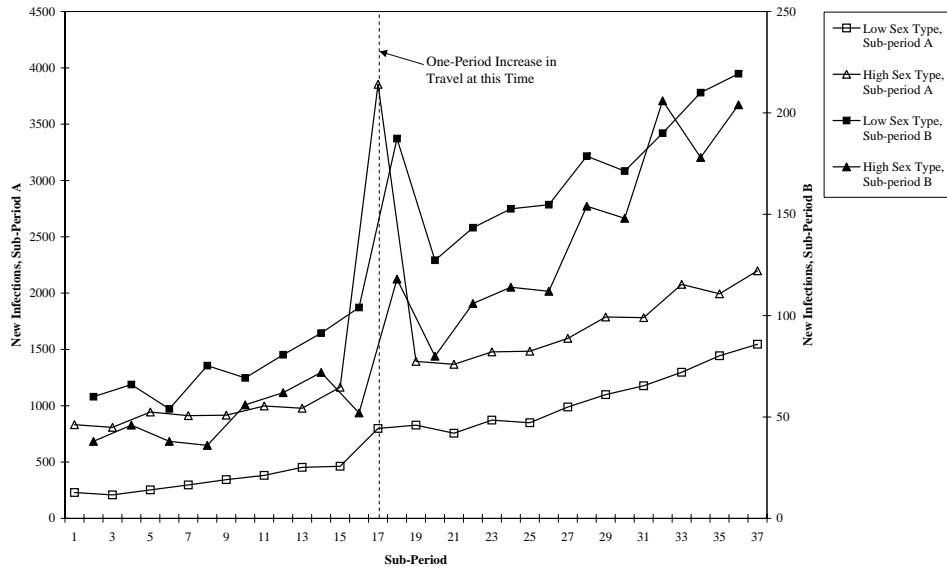
Notes: This figure shows HIV prevalence estimated by the mortality data for each country-gender-age group (5-year age groups) for the most recent year in the sample graphed against the DHS estimated prevalence from an average of five years later (average of three years of mortality estimates). Countries include: Burkina Faso, Cameroon, Kenya, Mali, Zambia and Zimbabwe.

**Figure 5:**  
**HIV Rates Estimated From Mortality Data Versus UNAIDS Estimates**



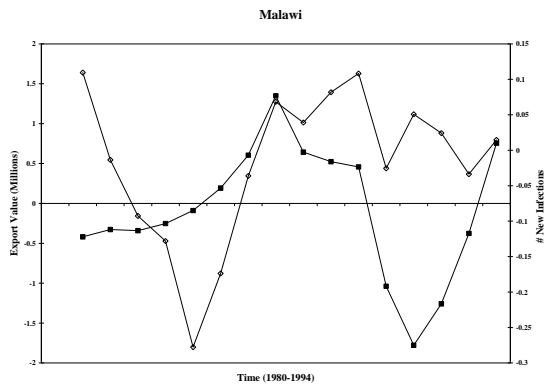
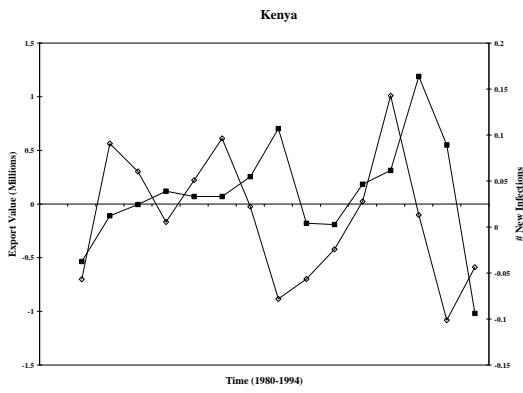
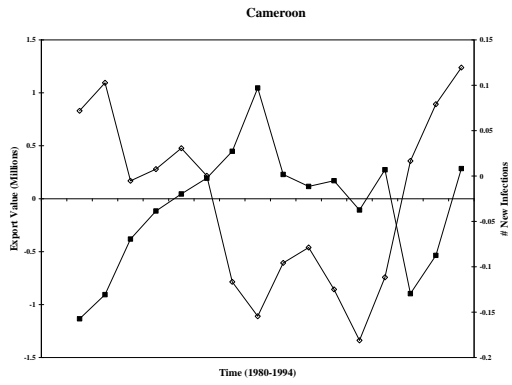
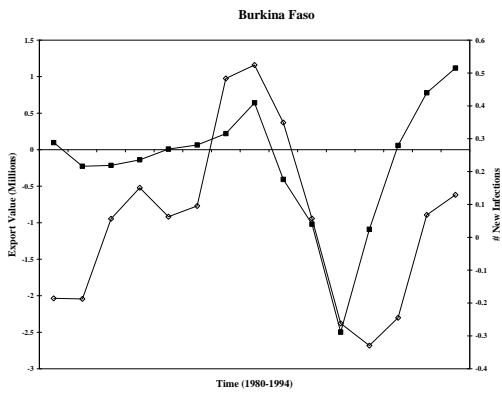
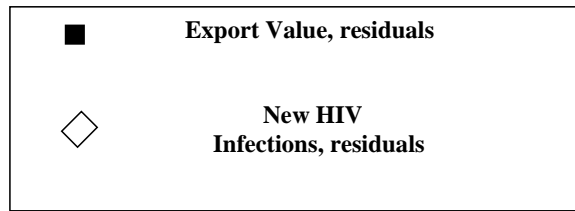
Notes: Solid bars are the estimated HIV rates for the general population (adults aged 16-35) based on mortality data. The black striped bars are the HIV rates in the same years as reported by UNAIDS. Standard errors are provided around the mortality estimates.

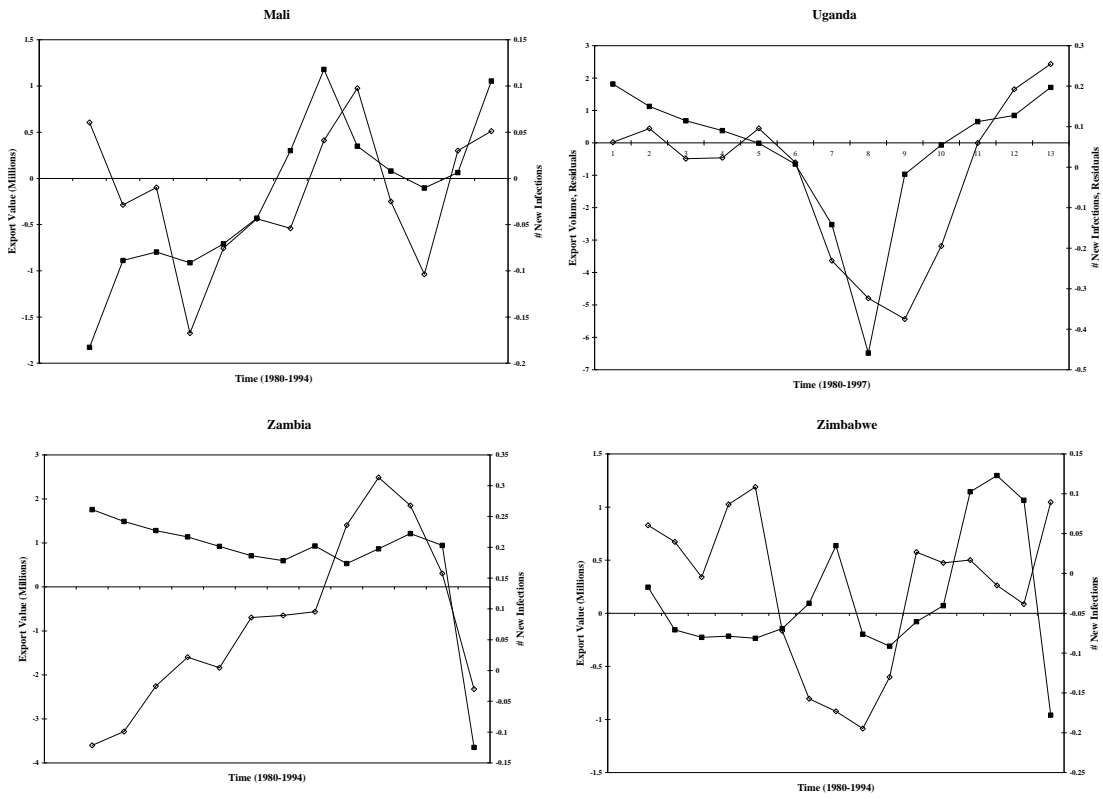
**Figure 6:**  
**HIV Infections over Time, Model-Generated Data**



Notes: This graph illustrates the output of the model, graphed over time. Sub-Period A takes place in the odd-numbered periods; Sub-period B in the even numbered ones. Travel takes place during sub-period A; in sub-period B everyone has sex with their spouse. The model assumes that 2% of men travel in all periods other than period 17, when 10% travel. 50% of men are high sex types; everyone else (the other 50% of men and all women) are low-sex types.

**Figure 7:  
HIV Infections and Exports, by Country**





Notes: This figure shows the time path of exports and new HIV infections, by country, for each country in the sample. Both series have country fixed effects, year fixed effects and country-specific trends removed.

Table 1. *Estimated HIV Rates over Time*

Country	Year	HIV Rate	5% Lower	95% Upper
Burkina Faso	1987	0.75%	0.31%	1.61%
Burkina Faso	1989	1.02%	0.43%	1.84%
Burkina Faso	1991	2.58%	1.53%	3.53%
Burkina Faso	1993	2.39%	1.46%	3.34%
Burkina Faso	1995	1.77%	1.18%	2.42%
Burkina Faso	1997	1.38%	0.94%	1.92%
Burkina Faso	1999	1.08%	0.70%	1.45%
Cameroon	1986	0.00%	0.00%	0.29%
Cameroon	1988	0.09%	0.00%	0.33%
Cameroon	1990	0.24%	0.02%	0.94%
Cameroon	1992	1.27%	0.33%	2.25%
Cameroon	1994	2.99%	1.80%	4.38%
Cameroon	1996	4.05%	2.76%	5.28%
Cameroon	1998	4.41%	3.02%	5.11%
Cameroon	2000	2.93%	2.35%	3.48%
Kenya	1985	0.23%	0.00%	0.53%
Kenya	1987	0.08%	0.00%	0.19%
Kenya	1989	0.11%	0.03%	0.69%
Kenya	1991	1.82%	0.80%	3.02%
Kenya	1993	2.04%	1.19%	3.05%
Kenya	1995	2.21%	1.50%	2.87%
Kenya	1997	4.29%	2.67%	6.40%
Kenya	1999	3.23%	2.44%	3.91%
Malawi	1986	4.02%	2.94%	5.08%
Malawi	1988	6.24%	4.95%	7.61%
Malawi	1990	5.62%	5.03%	6.21%
Malawi	1992	8.53%	7.71%	9.37%
Malawi	1994	12.23%	10.99%	13.49%
Malawi	1996	10.78%	9.74%	11.89%
Malawi	1998	8.30%	7.43%	9.17%
Malawi	2000	5.59%	5.01%	6.21%
Mali	1987	0.80%	0.37%	1.44%
Mali	1989	0.94%	0.42%	1.47%
Mali	1991	0.61%	0.31%	1.02%
Mali	1993	0.58%	0.27%	1.18%
Mali	1995	1.22%	0.74%	1.83%
Mali	1997	1.43%	0.84%	2.03%
Namibia	1986	4.17%	2.86%	5.50%
Namibia	1988	1.83%	1.16%	2.59%
Namibia	1990	1.24%	0.82%	1.78%
Namibia	1992	0.89%	0.54%	1.26%
Namibia	1994	2.46%	1.65%	3.35%
Namibia	1996	4.25%	3.23%	5.22%

*continued on the next page*

<b>Country</b>	<b>Year</b>	<b>HIV Rate</b>	<b>5% Lower</b>	<b>95% Upper</b>
Uganda	1985	5.29%	4.19%	6.41%
Uganda	1987	10.16%	8.75%	11.60%
Uganda	1989	10.08%	8.82%	11.35%
Uganda	1991	15.31%	13.89%	16.72%
Uganda	1993	8.37%	7.35%	9.34%
Uganda	1995	7.02%	6.03%	7.98%
Uganda	1997	11.69%	10.39%	13.06%
Zambia	1985	1.74%	0.97%	2.47%
Zambia	1987	6.29%	5.16%	7.47%
Zambia	1989	12.71%	11.21%	14.09%
Zambia	1991	14.79%	10.76%	20.31%
Zambia	1993	9.95%	8.82%	11.05%
Zambia	1995	12.76%	11.15%	14.14%
Zambia	1997	11.47%	10.29%	12.59%
Zimbabwe	1985	0.04%	0.00%	0.29%
Zimbabwe	1987	0.69%	0.00%	1.65%
Zimbabwe	1989	2.71%	1.67%	4.09%
Zimbabwe	1991	3.84%	2.68%	5.33%
Zimbabwe	1993	6.60%	4.91%	8.40%
Zimbabwe	1995	8.64%	7.01%	10.23%
Zimbabwe	1997	10.05%	8.26%	11.94%
Zimbabwe	1999	13.98%	9.90%	15.62%
Zimbabwe	2001	8.49%	7.53%	9.23%

Notes: These are estimates of HIV rates in the nine sample countries over time. The rates are estimated for 16-35 year old individuals of both genders. Column 3 is the average, Column 4 is the bootstrapped fifth percentile and Column 5 is the bootstrapped ninety-fifth percentile.

**Table 2. Mortality-Based Estimates and DHS Population Testing**

<i>Dependent Variable: Mortality-Based Prevalence Estimate, by Group</i>				
	(1)	(2)	(3)	(4)
	All	All	Women	Men
Explanatory Variables:				
DHS Testing Estimate	.7111*** (.069)	.5972*** (.027)	.5372*** (.043)	.6562*** (.022)
Age Group 20-25	.0158 (.011)	.0194 (.012)	.0173 (.015)	.0238* (.012)
Age Group 25-30	-.009 (.013)	.0002 (.01)	-.0084 (.016)	.0114 (.013)
Age Group 30-35	-.0167** (.006)	-.0044 (.01)	-.0172 (.013)	.009 (.015)
Age Group 35-40	-.0178* (.01)	-.0057 (.005)	-.0196** (.008)	.0077 (.011)
Age Group 40-45	-.0238 (.013)	-.014 (.009)	-.0206 (.014)	-.0098 (.007)
Female	-.0056 (.004)	-.0024 (.003)		
constant	.022*** (.006)	.023*** (.005)	.034*** (.006)	.01 (.009)
Country FE	NO	YES	YES	YES
Number of Observations	72	72	36	36
R <sup>2</sup>	.88	.93	.94	.95

standard errors in parentheses, clustered by country

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table shows the relationship between the mortality-based estimates of HIV prevalence and the DHS testing data. An observation is an age group-gender (age groups 15-20, 20-25, 25-30, 30-35, 35-40).



**Table 3. HIV and Travel: Model-Generated Data**

	<i>Dependent Variable: Log Number of New HIV Infections</i>				
	All	All	All	All	Low Sex Types Only
Explanatory Variables:					
Log # Traveling	.3691*** (.017)	.3739*** (.016)			
Log # Traveling Last Year		.0685*** (.018)			
Two Year Average # Traveling			.5205*** (.027)	.4001*** (.027)	.3504*** (.036)
Average # Travel × HIV Prevalence				2.8524*** (.558)	
				-36.8914*** (6.585)	
“Country” FE	YES	YES	YES	YES	YES
“Year” FE	YES	YES	YES	YES	YES
Country-Year Trend	YES	YES	YES	YES	YES
Number of Obs.	179	174	174	174	174
R <sup>2</sup>	.98	.98	.99	.99	.95

<sup>a</sup> standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table shows regressions run using “data” generated by the model, described in Section 3. Columns 1-4 include all infections, Column 5 only infections among low-sex types in the model. The two year average is the log of the average of number traveling this year and last year.

**Table 4. Exports and Road Traffic**

<i>Dependent Variable</i>	<i>Log Weight of Goods Trucked</i>	<i>Log Weight of Trucks and Parts Imported</i>
	Non-Sample, WDI Data	Sample, NBER Import Data
	(1)	(2)
Explanatory Variables:		
Log Export Value (WDI)	1.0813*** (.244)	
Log Export Volume (NBER)		0.6670*** (.185)
Country FE	YES	YES
Year FE	YES	YES
Country-Year Trend	YES	YES
Number of Observations	543	136
R <sup>2</sup>	.86	.82

standard errors in parentheses, Prais-Winston regression

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: Column 1 shows the relationship between the weight of goods transported by truck and export value within countries over time. The data is drawn from the World Development Indicators and covers a wide range of countries, outside of the HIV sample here. Column 2 shows the relationship between log weight of trucks imported and exports within the sample of countries used here. The data on truck imports comes from the NBER trade database (Feenstra, 2004), as does the data on export volume. Both columns include country fixed effects, year fixed effects and country-specific trends.

**Table 5. HIV Incidence and Exports**

<i>Dependent Variable: Log New HIV Infections</i>			
	(1)	(2)	(3)
Explanatory Variables:			
Log Export Value, all Exports (Average of year t, t-1)	2.6046*** (.53)		
Log Export Value, Major Export(s)(Average of year t, t-1)		2.0199*** (.452)	
Log Export Volume, Major Export(s) (Average of year t, t-1)			2.5984*** (.773)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Specific Trends	YES	YES	YES
Number of Observations	117	117	117
R <sup>2</sup>	.77	.76	.72

standard errors in parentheses, Prais-Winston regression

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table estimates the relationship between log new HIV infections, generated by the mortality-data model, and export value of all exports reported in the NBER-United Nations Trade Database (Column 1), export value for major exports from the same database (Column 2) and export volume for major exports from that database (Column 3). In all cases, the measure of exports is the log of the average of this year and last year's exports.

**Table 6. HIV Incidence and Exports: Interactions and Female Infections**

<b>Panel A: All Infections, Interactions</b>			
<i>Dependent Variable: Log New HIV Infections</i>			
	(1)	(2)	(3)
Explanatory Variables:			
Log Export Value, all Exports (2 year MA)	4.7394*** (.651)		
Log Export Value × HIV Rate	.6192*** (.178)		
Log Export Value, Major Export(s) (2 year MA)		4.2548*** (.586)	
Log Export Value × HIV Rate		.5612*** (.143)	
Log Export Volume, Major Export(s) (2 year MA)			4.3877*** (.987)
Log Export Volume × HIV Rate			.4494** (.18)
HIV Rate	−9.1615*** (2.362)	−8.195*** (1.815)	−6.4597*** (2.298)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Specific Trends	YES	YES	YES
Number of Observations	115	113	113
R <sup>2</sup>	.88	.88	.79
<b>Panel B: Female Infections</b>			
<i>Dependent Variable: Log New HIV Infections for Women</i>			
Explanatory Variables:			
Log Export Value, all Exports (2 year MA)	1.3344*** (.434)		
Log Export Value, Major Export(s) (2 year MA)		1.8351*** (.344)	
Log Export Volume, Major Export(s) (2 year MA)			2.5902*** (.421)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Specific Trends	YES	YES	YES
Number of Observations	117	117	117
R <sup>2</sup>	.80	.85	.83
standard errors in parentheses, Prais-Winston regression			
* significant at 10%; ** significant at 5%; *** significant at 1%			
Notes: This table estimates the relationship between log new HIV infections, generated by the mortality-data model, and export value of all exports reported in the NBER-United Nations Trade Database (Column 1), export value for major exports from the same database (Column 2) and export volume for major exports from that database (Column 3). In all cases, the measure of exports is the log of the average of this year and last year's exports. Panel A covers all infections, Panel B only female infections.			

**Table 7. HIV Incidence and Exports, by Region Characteristics**

<i>Dependent Variable: Log New HIV Infections</i>				
	(1)	(2)	(3)	(4)
<b>Panel A: Distance from Capital</b>				
	Capital	Not Capital	Capital	Not Capital
Explanatory Variables:				
Log Exp. Value, (Major Exports)	2.5476*** (.616)	.9891*** (.204)		
Log Exp. Volume, (Major Exports)			1.5742 (1.092)	1.2685*** (.296)
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Country-Specific Trends	YES	YES	YES	YES
Number of Obs.	117	608	117	608
R <sup>2</sup>	.56	.71	.49	.70
<b>Panel B: Road Density (Outside of Capital)</b>				
	Highest Density	Lower Density	Highest Density	Lower Density
Explanatory Variables:				
Log Exp. Value, (Major Exports)	2.1376*** (.514)	.7235*** (.214)		
Log Exp. Volume, (Major Exports)			2.1635** (.869)	1.138*** (.302)
Region FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Country-Specific Trends	YES	YES	YES	YES
Number of Obs.	117	449	117	449
R <sup>2</sup>	.68	.75	.63	.75

Standard errors in parentheses, Prais-Winsten regressions.  
\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table estimates the relationship between log new HIV infections by region within countries, generated by the mortality-data model, and export volume. Panel A splits the sample by distance to the capital city (regions that contain the capital are in Columns 1 and 3; regions that do not are Columns 2 and 4). Panel B splits the sample *outside* of the capital city by road density. Columns 1 and 3 contain the regions with the highest density, Columns 2 and 4 the other regions. Exports are the average of this year and last year.

**Table 8. HIV and Exports by Export Type**

<i>Dependent Variable: Log New HIV Infections</i>						
<b>Major Export</b>	<b>Farming</b>	<b>Mining</b>	<b>Oil</b>	<b>Farming</b>	<b>Mining</b>	<b>Oil</b>
Explanatory Variables:						
Log Export Value, Major Export(s)	2.0797*** (.481)	6.6496** (2.765)	-1.3876 (1.69)			
Log Export Volume, Major Export(s)				2.2366*** (.84)	10.7661*** (3.051)	-8.8008** (3.206)
Country FE	YES	NO	NO	YES	NO	NO
Year FE	YES	NO	NO	YES	NO	NO
Country-Specific Trends	YES	YES	YES	YES	YES	YES
Number of Obs.	88	13	16	88	13	16
R <sup>2</sup>	.74	.65	.26	.66	.71	.51

<sup>a</sup> Standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: This table estimates the relationship between log new HIV infections and exports for countries with differing exports. Zambia is the country with mining as a major export; Cameroon is the country with oil. All countries other than Zambia and Cameroon are in Column 1. Farming includes logging (forestry).

**Table 9. Exports and HIV: Instrumenting with Prices**

	First Stage	Reduced Form 2nd Stage	IV Estimates
<i>Dependent Variable</i>	<i>Log Export Volume</i>	<i>Log New HIV Infections</i>	
	(1)	(2)	(3)
Explanatory Variables:			
Log Commodity Price	.3367*** (.108)	2.163*** (.852)	
Log Export Volume, Major Export(s)			7.651*** (2.553)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Year Trend	YES	YES	YES
Number of Observations	128	117	117
R <sup>2</sup>	.99	.65	.59

standard errors in parentheses, Prais-Winston regression (IV with Newey-West)

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: Column 1 shows the first stage relationship between export volume (average of year  $t$  and year  $t - 1$ ) and commodity price (average of year  $t - 1, t - 2$  and  $t - 3$ ). Column 2 shows the second stage reduced form relationship between new HIV infections (year  $t$ ) and price; Column 3 shows the full IV regression.

**Table 10. HIV and Truck Imports, All Imports**

	HIV and Truck Imports	HIV and All Imports	
	<i>Dependent Variable: Log New HIV Infections</i>		
	(1)	(2)	(3)
Explanatory Variables:			
Log Weight Trucks Imported	.5343** (.259)		
Log Imp. Value, Major Imports		1.195* (.694)	
Log Imp. Volume, Major Imports			.713** (.296)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Year Trend	YES	YES	YES
Number of Observations	125	117	117
R <sup>2</sup>	.63	.64	.64

standard errors in parentheses, Prais-Winston regression

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: Column 1 relates new HIV infections directly to truck imports, as reported in the NBER trade database. Columns 2 - 3 relate new infections to value and volume of major imports reported in the NBER. All total import measures are the average of the current year and previous year, as with the export measures.



Table 11. *Effect of Exports on Incidence, Alternative Assumptions*

<b>Panel A: Change in Smoothing</b>			
<i>Export Measure</i>	<i>Value, All</i>	<i>Value, Major</i>	<i>Volume, Major</i>
<b>Less Smooth (<math>\gamma = 0.01</math>)</b>	1.711*** (.5333)	2.327*** (.4606)	2.672*** (.7989)
<b>More Smooth (<math>\gamma = 0.5</math>)</b>	1.247*** (.4691)	1.633*** (.4029)	2.372*** (.6446)
<b>Panel B: Alternative Time to Death</b>			
<i>Export Measure</i>	<i>Value, All</i>	<i>Value, Major</i>	<i>Volume, Major</i>
<b>Faster, Same Shape</b>	1.928*** (.6244)	2.562*** (.5353)	3.842*** (.9131)
<b>Flatter, Same Time</b>	1.549*** (.5292)	2.476*** (.4111)	3.142*** (.7466)
<b>Panel C: Alternative Assumptions about Non-HIV Mortality</b>			
<i>Export Measure</i>	<i>Value, All</i>	<i>Value, Major</i>	<i>Volume, Major</i>
<b>No Non-HIV Mortality</b>	1.524*** (.4793)	1.051** (.4783)	1.678** (.7628)
<b>Non-HIV Mortality Increased by 25%</b>	1.826*** (.5126)	2.150*** (.4565)	2.720*** (.7931)

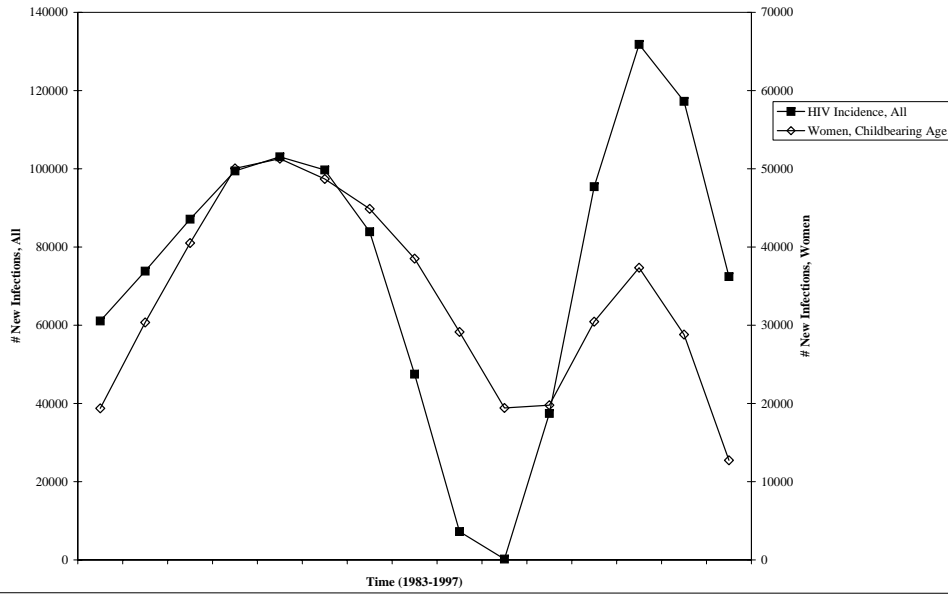
Notes: This table replicates Table 5 using alternative data on HIV incidence, generated based on different assumptions about smoothing, time to death from infection or non-HIV mortality. The assumptions about time to death can be seen visually in Appendix Figure 2.

**Table 12. *Explaining the Ugandan Incidence Decline***

	<b>Pred. Decline in HIV Late 1980s (87-89) to Early 1990s (91-93)</b>	<b>Actual Decline in HIV Late 1980s (87-89) to Early 1990s (91-93)</b>	<b>% Expl. by Exports</b>
<b>Explaining All Infections (Log HIV)</b>			
<b>Export Value All Exports</b>	1.631	3.192	<b>51%</b>
<b>Export Value Major Exports</b>	1.931	3.192	<b>60%</b>
<b>Export Volume Major Exports</b>	1.040	3.192	<b>33%</b>
<p>Notes: This table shows the share of the Ugandan incidence decline that can be accounted for by changes in export volume or value. The predicted decline is based on the regressions in Table 5. % Explained is simply predicted divided by actual decline.</p>			

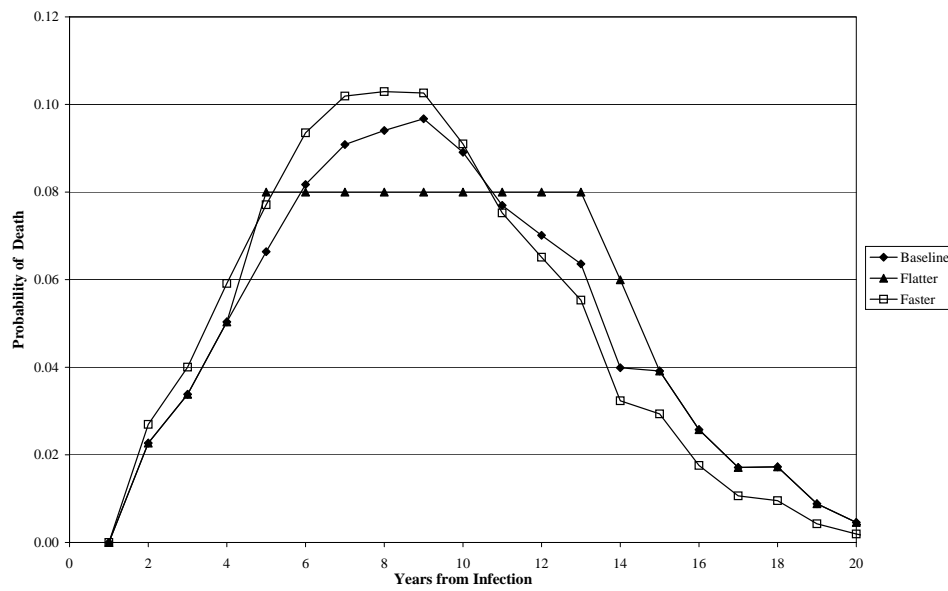
# Appendix Figures and Tables

Appendix Figure 1:  
HIV Incidence in Uganda, Two Groups



Notes: This figure shows HIV incidence for Uganda, as calculated using the mortality-based methodology in this paper for 1983 through 1997 for the entire population, and women of child-bearing age.

Appendix Figure 2  
Time to Death from AIDS, Alternative Assumptions



Notes: Figure shows the PDF of time to death from HIV, under the baseline assumption, the assumption of flatter time path to death and the assumption of faster time to death

**Appendix Table 1: HIV Incidence and Exports, current and lagged**

<i>Dependent Variable: Log New HIV Infections</i>			
	(1)	(2)	(3)
Explanatory Variables:			
Log Value, All Exports, Current Year	1.0253** (.479)		
Log Value, All Exports, Last Year	1.562*** (.486)		
Log Value, Major Exports, Current Year		1.0398*** (.372)	
Log Value, Major Exports, Last Year		1.0161*** (.383)	
Log Volume, Major Exports, Current Year			1.1578** (.527)
Log Volume, Major Exports, Last Year			1.4555*** (.527)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-Specific Trends	YES	YES	YES
Number of Observations	117	117	117
R <sup>2</sup>	.73	.72	.68

standard errors in parentheses, Prais-Winston regression  
\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Notes: Notes: This table estimates the relationship between log new HIV infections, generated by the mortality-data model, and export value of all exports reported in the NBER-United Nations Trade Database (Column 1), export value for major exports from the same database (Column 2) and export volume for major exports from that database (Column 3).

## Appendix A: Details of Smoothing in Estimation

This appendix provides more details on why, when estimating prevalence, it is necessary to impose a smoothing restriction.

Recall that we begin with the concept of solving a system of equations, as below (for simplicity, this appendix notation abstracts away from deaths from non-HIV causes):

$$\mu_{i,t} = d_4 b_{i-4,t-4} + d_5 b_{i-5,t-5} + \dots + d_{20} b_{i-20,t-20} \quad (7)$$

$$\mu_{i-1,t-1} = d_4 b_{i-5,t-5} + d_5 b_{i-6,t-6} + \dots + d_{20} b_{i-21,t-21} \quad (8)$$

$$\dots \quad (9)$$

$$\mu_{i-x,t-x} = d_4 b_{i-x-4,t-x-4} \quad (10)$$

With noiseless data on mortality rates solving the system of equations for the  $\mathbf{b}$  vector will work perfectly. However, noisy data on mortality will produce inappropriately high volatility in predicted infection rates. To see why, consider a simple example: a world with just two years of infections and two years of data on mortality.

$$\begin{aligned} \mu_{i,t} &= d_4 b_{i-4,t-4} + d_5 b_{i-5,t-5} \\ \mu_{i-1,t-1} &= d_4 b_{i-5,t-5} \end{aligned}$$

Assume that the actual values in the  $\mathbf{b}$  vector are  $b_{i-4,t-4} = b_{i-5,t-5} = \psi$  and that  $d_4 = d_5 = 0.1$ . If we observed the exact mortality, we would see  $\mu_{i,t} = 0.2\psi$  and  $\mu_{i-1,t-1} = 0.1\psi$  and would conclude that  $b_{i-4,t-4} = b_{i-5,t-5} = \psi$ . If the mortality rates are observed with noise, however, we could see  $\mu_{i,t} = 0.2\psi + \varepsilon$  and  $\mu_{i-1,t-1} = 0.1\psi + \eta$ . Solving the system of equations will yield

$$\begin{aligned} b_{i-5,t-5} &= \psi + \frac{\eta}{.1} \\ b_{i-4,t-4} &= \psi + \frac{\varepsilon - \eta}{.1} \end{aligned}$$

If we assume that  $\psi = 0.011$ ,  $\eta = 0.002$ , and  $\varepsilon = -0.002$ , then instead of estimating that  $b_{i-4,t-4} = b_{i-5,t-5} = 0.011$ , we will estimate that  $b_{i-5,t-5} = 0.031$  and  $b_{i-4,t-4} = -0.029$ . Intuitively, a small amount of noise is translated into very large deviations between periods because of the very limited additional information in each of the simultaneous equations. The smoothing restriction will address this issue. By penalizing the maximization when any year is far from the two years surrounding it, this does not allow the system to diverge wildly across years (this is similar to an approach in Murphy and Welch, 1990).

## Appendix B: Details on the Reliability of Sibling Mortality Histories

This appendix provides some additional defense of the use of sibling mortality histories by providing direct evidence on the comparison between these sibling histories and official mortality statistics, in a few cases where both are available. I also outline the theoretical issues associated with this type of data.

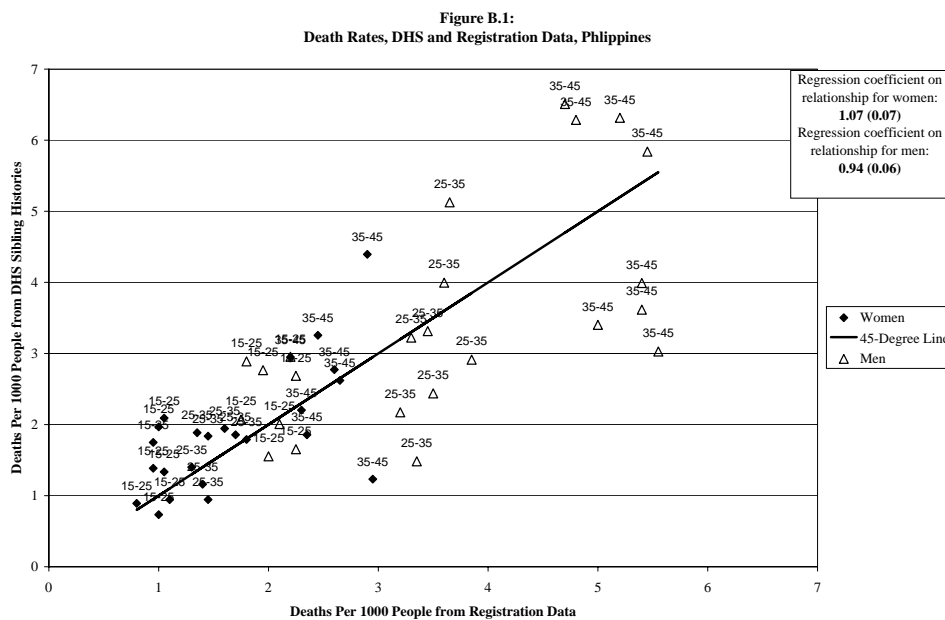
In general, there are two primary concerns with using sibling mortality histories. The first is a simple concern about underreporting, since individuals may forget siblings who have died or may not know that their siblings have died. In addition, even with perfect reporting sibling mortality data have the potential for either upward or downward bias. Sibling reports may overstate the overall death rate since they do not take into account the reporters themselves, who are alive. They may understate the overall death rate since families with low mortality profiles will be misrepresented, and families in which all siblings are dead will not be observed at all (Gadikou and King, 2006). Remarkably, Trussell and Rodriguez (1990) prove that with a random sample of the population *if mortality is uncorrelated with sibship size* these factors will cancel out and the report is unbiased. Although the assumption about independence sibship size is probably unrealistic when considering individual mortality in childhood, in terms of thinking about adult mortality from HIV it may be more appropriate (indeed, in the samples here we do not see a strong correlation).

Rather than relying simply on the result in Trussell and Rodriguez (1990), however, I provide some direct evidence on the match between sibling mortality histories and official death data, using data from two

countries in which I have both sources: the Philippines and Zimbabwe. For the Philippines, I match death rates from sibling mortality histories collected by the DHS to official death rates reported in the United Nations Demographic Yearbook. The data come from 1980, 1983, 1984, 1986 and 1988-1991, and I generate mortality rates for each year by ten year age groups (15-25, 25-35 and 35-45) and gender. The data are shown in Figure B.1, along with the 45 degree line. On average, the death rates from sibling histories are very close to the official rates. A regression of the sibling history death rate on the official death rate for women yields a coefficient of 1.07 (standard error of 0.07) and, for men, a coefficient of 0.94 (standard error of 0.06). In neither case are these statistically distinguishable from a coefficient of one.

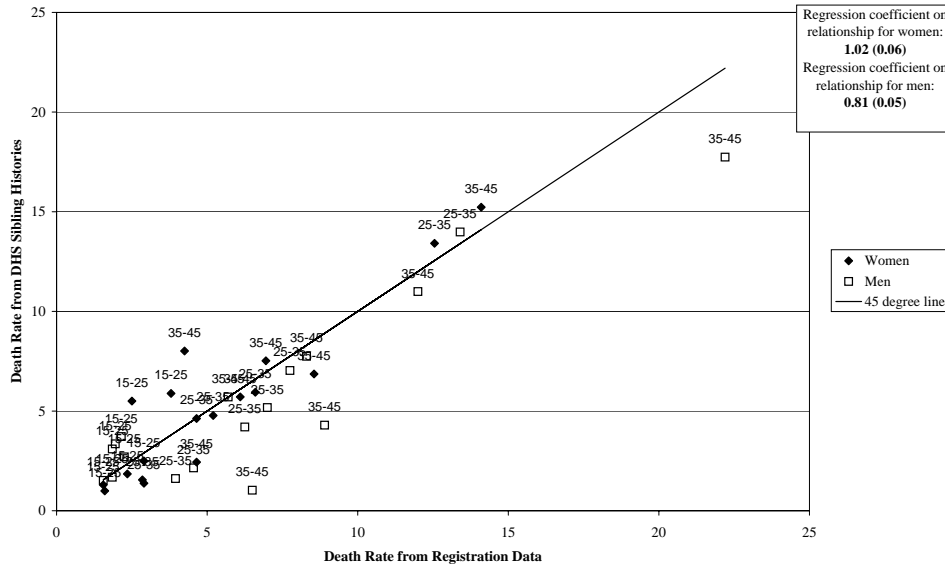
For Zimbabwe, data based on DHS sibling histories are compared with registration data from Feeney (2001), for 1982, 1986, 1990, 1991, 1992, and 1995 (although these are official registration data, it is worth noting that in no year do they cover more than 60% of the population). Figure B.2 replicates Figure B.1. Again, the match between the two datasets is quite strong. For women, the coefficient on the relationship is 1.02, not distinguishable from 1. For men, the coefficient is slightly smaller (at 0.81). This could reflect larger problems with under-reporting when women are asked about their brother, although the absolute magnitude of the errors is still very small. In general, Figures 4 and 5 show strong evidence (although for a limited set of countries) that sibling histories generate accurate mortality rates, similar to what we would see if we had access to official death registration data. This is true even in Zimbabwe, where mortality is heavily affected by HIV.

The evidence in Figures B.1 and B.2 suggests that the naive estimator of death rates from this data (sibling deaths over a period divided by siblings alive at the start) provides a good fit to registration data. It is also worth noting that similar results are obtained if I use the adjustments for bias worked out in Gakidou and King (2006).



Notes: This figure shows the death rate from sibling histories done in 1993 for age-gender groups (15-25, 25-35, 35-45) in the Philippines (years of data match: 1980, 1983, 1984, 1986, 1988, 1989, 1990, 1991), graphed against matched mortality rates from the United Nations Demographic Yearbook Historical Supplement.

Figure B.2:  
Death Rates, DHS and Registration Data, Zimbabwe



Notes: This figure shows the death rate from sibling histories for age-gender groups (15-25, 25-35, 35-45) in Zimbabwe in 1982, 1986, 1990, 1991, 1992 and 1995 graphed against matched mortality rates from registration data. Registration data is from Feeney (2001).