

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

Emily Oster\*  
University of Chicago and NBER

September 10, 2010

## Abstract

This paper estimates whether exports affect the incidence of HIV in Africa. This relationship has implications for HIV prevention policy as well as for the consequences of trade increases in Africa. I estimate this impact using two sources of data on HIV incidence, one generated based on UNAIDS estimates and the other based on observed HIV mortality. These data are combined with data on export value and volume. I find a fairly consistent positive relationship between exports and new HIV infections: doubling exports leads to between a 10% and a 70% increase in new HIV infections. Consistent with theory, this relationship is larger in areas with higher baseline HIV prevalence. I interpret the result as suggesting that increased exports increase the movement of people (trucking), which increases sexual contacts. Consistent with this interpretation, the effect is larger for export growth than for income growth per se and is larger in areas with more extensive road networks.

## 1 Introduction

Economists and policy-makers have traditionally argued that there is a positive link between income and health (see, for example, Pritchett and Summers, 1996). This link has been taken as evidence that an increase in globalization and trade in the developing world will make people better off not only through increases in income but also through improvements in health (Dollar, 2001). In line with this, cross-country analysis has suggested that trade and health are positively linked (Owen and Wu, 2007; Levine and Rothman, 2006). Recent research, however, has called into question whether increases in income really do improve health (Case and Deaton, 2006; Ruhm, 2007). In the case of communicable diseases, historical episodes suggest that trade (while good for income) may be detrimental to health by promoting disease spread. In the case of the Black Plague, there is clear evidence that it spread out of Asia along trade routes. The initial introduction to Europe was

---

\*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, Northwestern, Ohio State and Berkeley provided helpful comments. Dwyer Gunn provided outstanding research assistance.

through a merchant ship from the Black Sea, which arrived in Sicily in 1347 with an infected crew (Wilson, 1995; McNeill, 1976).

In this paper I explore the question of whether economic activity, and trade in particular, can be detrimental to health in an important policy context: HIV in Sub-Saharan Africa. Significant policy emphasis has been put on the importance of trade in increasing income in Africa, following the lead of China and India (Arbache et al., 2008; Dollar, 2008). At the same time, roughly 5% of the adult population is infected with HIV, and if the historical evidence is a guide, increases in trade have the potential to increase the spread of the virus. This concern seems particularly worrying since individuals traditionally involved in trade-related activities (truckers and migrant workers) have very high HIV prevalence (Lurie et al., 2003a; Lurie et al., 2003b; Brewer et al., 1998; Brockerhoff and Biddlecom, 1999; Anarfi et al., 1997; Anarfi, 1993).

The primary analysis in this paper focuses on estimating the impact of export activity on new HIV infections within a country over time. The central question is whether increases in trade increase HIV prevalence. Section 2 briefly outlines the theory and empirical strategy motivating this argument. I begin by positing a relationship between exports and sexual behavior in which increases in exports increase the share of individuals involved in risky sex. I combine this model with a simple model of how increases in risky sex translate into an increase in new HIV infections; the result is an equation connecting exports to new HIV infections. The model suggests the relationship between the two will be positive, and that it will be larger when current HIV prevalence is higher. Further, the model provides a clear way to evaluate the magnitude of the coefficients I observe by connecting them directly to the relationship between exports and sexual behavior. This will allow us to understand whether the magnitude of the results are reasonable.

The significant challenge to this estimation is poor data on HIV infection rates, which make it difficult to calculate a time series of new infections.<sup>1</sup> To address this, I use two newly available sources of data on HIV infections. First, I use recently produced data on trends in HIV infection from UNAIDS to estimate incidence in 35 countries (UNAIDS, 2008). Second, for a subset of 11 countries I also use estimates of infection rates based on inference from mortality data, described in

---

<sup>1</sup>Before the early 2000s, virtually no testing was done in the general population. Testing was done among pregnant women, but the selection of the sample changed significantly between years in most countries, undermining the ability to use these data to estimate trends. 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 arising from a lack of consistent data over time. Further, in order to calculate new HIV infections in the current year it is necessary to have data going back into the past – new infections in year  $t$  is not simply the difference between number of infections in year  $t$  and year  $t - 1$ ; it is also necessary to adjust for deaths, which are affected by infection rates in earlier years.

Oster (forthcoming). Both data sources are described in more detail in Section 3.

I combine these data with data on exports. I focus on three measures of exports: total export value from the World Development Indicators, and export value and volume from the NBER-United Nations Trade Data described in Feenstra et al. (2004). For the last two measures I focus only on export categories which are consistently observed over time (as in Jones and Olken, 2010), and supplement the NBER data series with data for 2000-2007 from Comtrade.

I first estimate the relationship between exports and HIV controlling for country and year fixed effects and lagged HIV prevalence only. I find a fairly consistent positive relationship between exports and new HIV infections: doubling exports leads to between a 10% and a 70% increase in incidence, depending on the HIV measure used. Interpreted in the context of the model, these results imply that doubling exports leads to between a 10% increase and a doubling in the share of people who are engaged in risky sex, which I argue is not unrealistic. In addition to this basic relationship, I estimate models in which I interact exports with existing HIV prevalence level. This analysis also strongly supports the theory, with estimated positive and significant coefficients. The range of coefficients is smaller here: the results imply that doubling exports leads to between a 30% increase and a doubling in the share involved in risky sex. Dividing the sample by region, we find this relationship holds in East Africa and Southern Africa, but not in West Africa, suggesting caution should be taken in generalizing.

Following this initial evidence of a relationship between HIV and exports, I explore whether there is evidence that this relationship is causal. Reverse causality is unlikely to be a concern because of the contemporaneous time frame: there is no reason to expect new HIV infections this year (which would be asymptomatic) to drive exports this year. In addition, I argue that it is fairly unlikely that an omitted variable (for example, government policy) is driving these results. This claim is based on two pieces of evidence. First, I find the estimates are robust to including country-specific trends in the regression. Second, I test whether exports in future years impact HIV infections in the current year and find that they do not. I argue that this evidence is suggestive of a causal relationship, although given the issues inherent in the data, causal claims should be taken with some caution.

The evidence thus far, including the evidence on causality, is silent on the mechanisms through which this effect occurs. Fully separating out these mechanisms is beyond the scope of this paper. However, Section 5 provides some preliminary evidence on the role of two factors. The first is income. Since at least some data suggests that sex is a normal good, the increased income provided by increases in exports could drive more risky sex (Ahlberg and Jensen, 1998). I explore the

contribution of income to the results by adding a control for GDP to the regressions. To the extent the income channel is responsible for the effects we see, adding a control for income will decrease the coefficient on exports. I find some evidence for the income channel, but it is limited. Although GDP does seem to impact HIV incidence (especially when estimated without the export control), adding GDP to the primary regressions suggests that only 4% to 20% of the export impact is driven by income changes.

The second mechanism I consider is trucking, or people-movement more generally. I present several pieces of evidence that support a role for this factor in driving the results. First, I demonstrate that the import of trucks increases with exports, which I interpret as indicating an increase in truck usage when exports go up. This fact, in combination with existing evidence that truckers have more high risk sex, supports the trucking mechanism. Second, I show that in countries with higher road density the export-HIV relationship is much stronger. Finally, I show cross-sectional evidence that HIV prevalence is higher in areas closer to roads; although this does not link directly to exports, it does point to the importance of transit in driving the epidemic (this analysis is similar to Djemai, 2009).

This paper adds to the literature on the relationship between income and health, as well as the medical literature on the impact of high HIV rates among high risk groups in Africa. In addition, I add to a small literature within economics on the HIV epidemic in Africa. This literature has focused largely on the effect of HIV on behavior, including sexual behavior (Oster, 2009; Thornton, 2008), fertility (Fortson, 2009; Kalem-Ozcan, 2006) and human capital (Fortson, forthcoming). In contrast to those papers, in this paper I do not explicitly consider individual behavior, although obviously individual behavior is in the background of the effects I find. Further, the effects I identify may well result from changes in individual *opportunities* for risky behavior, rather than changes in individual *choices* about risky behavior, which is what the previous literature has focused on.

As I discuss in more detail in the conclusion, it is important to keep in mind that what is estimated in this paper is the short-term effect of changes in exports on HIV incidence, within a given country. The results point to exports as a factor which drives changes in HIV incidence at this level. However, the evidence here does not indicate that variations in exports are responsible for some countries having higher HIV rates than others; this is not ruled out by the evidence, but we simply cannot speak to that variation with these data.

The rest of the paper is organized as follows. Section 2 outlines a very simple model of the epidemic and presents the estimation strategy; Section 3 describes the data used. Section 4 estimates

the relationship between HIV incidence and exports and Section 5 discusses evidence on the mechanisms. Section 6 concludes.

## 2 Epidemic Model and Estimation Framework

This section outlines a relatively simple model of the HIV epidemic and describes how changes in economic activity could impact epidemic growth. Although the model is simple, it will both suggest a particular functional form for estimation and provide us a way to connect the estimates to parameters in the real world, and thus calibrate whether magnitudes are reasonable.

### 2.1 Sexual Behavior and Economic Activity

I begin by outlining a proposed connection between sexual behavior and economic activity. Assume without loss of generality that we have a population of size 1, of which a share  $\alpha$  are high risk individuals, meaning they have more sexual partners, and this share is a function of exports. For example, high risk individuals could be truckers, and increasing exports would increase the number of truck drivers. As another example, in a place with significant natural resource mining (e.g., Zambia), miners may also be classified as high risk. More exports imply more use of miners, which increases the share of individuals in the high risk/high sex group. It is also possible to interpret this relationship as operating through income. That is, high risk individuals could be individuals with relatively high income, with increased exports increasing income.

To make progress, it is necessary to specify a model of the relationship between  $\alpha$  and exports. At first glance, a linear model might seem obvious – for example, if we assume this relationship is driven largely by truckers, then we might imagine that one extra truckload requires an constant extra fraction of a trucker. However, taking this logic to the extreme it becomes clear this is not an appropriate model. France and the Democratic Republic of the Congo have very similar populations, but in 2007 France exported 300 times as much by value as the Democratic Republic of the Congo did; it seems implausible that the number of people employed in trucking is 300 times as large in France. Clearly, countries differ in the technologies they use. By extension, increasing exports by a fixed dollar amount will have a larger impact in some places than in others; a similar argument can be made if we think this impact is through income.

A more sensible model would seem to be one in which the share of high risk individuals adjusts with exports in proportion to a baseline export level. Assume that a country has an average

export level of  $X_{ave}$  and at this average level the share of high risk individuals is  $\alpha_{ave}$ . As exports increase, the share of individuals involved in exports increases in proportion to this increase. To be more specific, denoting exports in year  $t$  as  $X_t$  and the share of high risk individuals in year  $t$  as  $\alpha_t$ , I specify the relationship  $\frac{\alpha_t}{\alpha_{ave}} = \gamma \frac{X_t}{X_{ave}}$ . To express  $\alpha_t$  as a function of  $X_t$ , I can take logs and rearrange, yielding Equation (1) below.

$$\ln(\alpha_t) = \gamma \ln(X_t) + [Constant] \quad (1)$$

That is, under this model the log share of high risk individuals and the log of exports are linearly related, with a multiplier of  $\gamma$ , which defines the magnitude of this relationship. For example,  $\gamma = 1$  implies that if exports double within a country the share of individuals who are in the high risk group doubles. When I turn to discuss magnitudes, the key question will be what the results imply about  $\gamma$  and whether those values are reasonable.

## 2.2 Sexual Behavior and Spread of HIV

The second element of the model is the relationship between the share of high risk individuals and the spread of HIV. This epidemic model will be quite simplistic, but will capture many of the important epidemic features and is very tractable.

I assume there are just the two types of individuals – high risk and low risk.<sup>2</sup> Low risk individuals have  $r$  sexual partners drawn from the general population. High risk individuals have  $r$  partners drawn from the general population and  $s$  partners drawn from some outside population (i.e., prostitutes).<sup>3</sup> The general population has an HIV rate of  $h_t$  at the start of year  $t$ , and the outside population has an HIV rate of  $\rho h_t$  where  $\rho > 1$ . Denote the chance of infection per partnership as  $\beta$ .

The chance of new infection for the two groups are given in Equations (2) and (3) below.

$$\text{Low Type: } \beta h_t r \quad (2)$$

$$\text{High Type: } \beta h_t r + \beta \rho h_t s \quad (3)$$

Combining these changes, and weighting by the shares, Equation (4) shows the probability of new

---

<sup>2</sup>We could add a second group of high risk individuals whose behavior is the same as the export group but whose share does not vary with exports (i.e., IV drug users). This will not impact our conclusions, given the additive separability. A version of this model with a third group is available from the author.

<sup>3</sup>To close the model, I assume some population of prostitutes who do not mix in the general population and who have a large number of sexual partners. I assume this population is sufficiently small in terms of numbers that they do not significantly impact the HIV rate. This seems like a reasonable assumption given the setting.

infection in the overall population in year  $t$ , which I denote  $\pi_t$ .<sup>4</sup>

$$\pi_t = \beta h_t r + \alpha_t \beta \rho h_t s \quad (4)$$

In this setup, changes in the share of individuals who are the high type will impact current HIV incidence only for the high type. However, since this ultimately impacts HIV prevalence, the future low-type incidence will also be impacted by current exports through the change in prevalence. In our specifications we attempt to control for this channel by including lagged HIV prevalence in the regressions.

### 2.3 Sexual Behavior and Exports

The goal is to estimate a relationship between sexual behavior and exports – that is, to estimate the change in infection rate ( $\pi$ ) for a change in exports  $X_t$ . Since the relationship between exports and sexual behavior is linear in logs, I will actually estimate the relationship between infection rate and log exports. To do this, I note the following identity:

$$\frac{\partial \pi_t}{\partial \ln(X_t)} = \frac{\partial \pi_t}{\partial \ln(\alpha_t)} \frac{\partial \ln(\alpha_t)}{\partial \ln(X_t)}$$

From Equation (1) we observe that  $\frac{\partial \ln(\alpha_t)}{\partial \ln(X_t)} = \gamma$ . From Equation (4), substituting  $\exp(\ln(\alpha_t))$  for  $\alpha_t$ , we find that  $\frac{\partial \pi_t}{\partial \ln(\alpha_t)} = \alpha_t \beta \rho h_t s$ . Combining these, Equation (5) gives the relationship between new infection rate and log exports.

$$\frac{\partial \pi_t}{\partial \ln(X_t)} = \gamma \alpha_t \beta \rho h_t s \quad (5)$$

It should be clear from Equation (5) that the relationship between incidence and exports is affected by the current HIV prevalence rate. For a fixed value of  $\gamma$ , the relationship should be larger when HIV prevalence is higher. We can further differentiate with respect to  $h_t$  to yield a comparative static which is not dependent on  $h_t$ , as is done in Equation (6). We will estimate both of these comparative statics in our empirical work.

$$\frac{\frac{\partial \pi_t}{\partial \ln(X_t)}}{\partial h_t} = \gamma \alpha_t \beta \rho s \quad (6)$$

---

<sup>4</sup>The one element I ignore here is that this chance of new infection applies only to uninfected individuals. In fact, I should adjust by  $(1 - h_t)$ . Because in nearly all cases  $h_t$  is close to zero, this makes only a very tiny difference, and I ignore it for simplicity. I have run the estimation adjusting for this (calculating  $\hat{\pi}_t$  as  $\frac{\pi_t}{1-h_t}$ ) and using that as the dependent variable and the results are extremely similar.

## 2.4 Estimation and Calibration

Motivated by the model above, I focus on estimating two primary regressions. I first estimate Equation (7) below, the baseline relationship between new infections and log exports. Denote  $i$  as the indicator for country and, as above,  $t$  as the indicator for year. In addition, as above, denote new infection rate in country  $i$  in year  $t$  as  $\pi_{it}$  and exports in that country-year as  $X_{it}$ .

$$\pi_{it} = \psi_0 + \psi_1(\ln(X_{it})) + \tau_i + \mu_t + \mathbf{\Lambda}\mathbf{Z}_{it} + \epsilon_{it} \quad (7)$$

where  $\tau_i$  is a fixed effect for country,  $\mu_t$  is a fixed effect for year and  $\mathbf{Z}_{it}$  is a vector of country-year specific controls. The last includes the HIV rate prevailing in the country in that year.<sup>5</sup>

The coefficient of interest from this regression is  $\psi_1$ . Based on the model above,  $\psi_1 = \gamma\alpha_t\beta\rho h_t s$ . This relationship allows us to calibrate whether the magnitude of the coefficient estimate is reasonable. In particular, using estimates of  $\beta$ ,  $s$  and  $\rho$  drawn from existing literature, along with an estimate of the average  $\alpha_t$  and  $h_t$ , the observed coefficient implies a value for  $\gamma$ . This  $\gamma$  is the relationship between exports and risky sexual behavior. Although the magnitude of this relationship is not known, we can use intuition to judge whether the magnitude we observe in the data is reasonable.

As noted above, the value of  $\gamma$  we calculate will depend on the HIV rate. Our primary coefficients are interpretable using the average HIV rate. However, when we explore heterogeneity across, for example, regions of Africa with varying HIV prevalence rates, the primary coefficient estimates are not comparable due to varying HIV prevalence. To generate comparable coefficients, we need to directly test Equation (6) above, which suggests the relationship between incidence and exports is mitigated by prevalence. We estimate Equation (8) below.

$$\pi_{it} = \delta_0 + \delta_2(\ln(X_{it}) \times HIV_{it}) + \tau_i + \mu_t + \mathbf{\Lambda}\mathbf{Z}_{it} + \epsilon_{it} \quad (8)$$

Again, among other controls  $\mathbf{Z}_{it}$  includes the control for HIV level. In this case, the coefficient of interest is  $\delta_2$ . Based on the model above,  $\delta_2 = \gamma\alpha_t\beta\rho s$ . This gives us another way to calibrate the magnitude of  $\gamma$ . Consistent with the theory, in this regression we do not include the level effect of exports, which should not enter here; in Online Appendix Table 1 we will show our primary results with this level effect included and demonstrate that the coefficients on the level effect are zero and the interaction coefficients are unaffected.

---

<sup>5</sup>The relevant HIV rate is the HIV rate prevailing at the start of year  $t$ , since that is what matters for new incidence. In practice, in the data, we use the HIV rate at the end of year  $t - 1$ , which should be the same as at the start of year  $t$ .

To provide estimates of  $\gamma$  based on the coefficient estimates, it is necessary to have an estimate of the values of  $\beta$ ,  $\rho$ ,  $s$  and an average level of  $\alpha_t$ .  $\beta$  is the per-partnership HIV transmission rate. Estimates of this vary widely (from as low as 0.1% to as high as 12%), and depend in part on the length of the partnership (for a range of estimates, see Oster, 2005). As the benchmark value, I use  $\beta = .05$  which reflects the fact that these partnerships are fairly short, but are likely to contain relatively high risk sexual acts and likely to reflect concurrent partnerships, which are higher risk (Wawer et al., 2005).  $\rho$  reflects the fact that the HIV rate among non-regular partners (i.e., prostitutes, bar girls) is higher than the overall HIV rate in the population. I estimate this multiplier factor based on data from the U.S. Census HIV/AIDS Surveillance database, which provides estimates of HIV prevalence for different groups. On average, in those data, the infection rate for prostitutes is 12 times the rate for pregnant women (the most representative group contained in those data); I therefore use  $\rho = 12$ .

The third element,  $s$ , measures the number of risky partners that individuals in the high risk group has (this is over and above any regular partners – wives or girlfriends – they have). Again, estimates of this parameter vary widely. I focus in particular on truck drivers and a study from Nigeria: in a sample of truck drivers the average number of partners in the last year is 12, and the average number of current partnerships is 6.3 (Orubuloye et al., 1993). This leads me to assume that  $s = 6$ . Fourth, I need a measure of the average value of  $\alpha_t$ . Again, I focus on truck drivers and attempt to calibrate  $\alpha_t$  based on the share of individuals who report trucking as their occupation in the Demographic and Health Surveys from several countries in Africa; these data suggest a value of around 2.5% for  $\alpha_t$ .<sup>6</sup> When estimating  $\gamma$  from regressions that *do not* include the interaction with HIV rate, I will also need to adjust for the average HIV rate, which is drawn directly from the data.

It is important to note that the estimates of  $\gamma$  that are provided are clearly sensitive to the assumptions made about  $\rho$ ,  $\beta$ ,  $s$  and  $\alpha$ . The choices used here represent my best guess as to values of these parameters. However, these do not play a direct role in the estimation, and if the reader disagrees with these choices it is easy to substitute other values and generate a different set of estimates of  $\gamma$ . It should also be noted that it is possible these values differ across areas or over time, which would drive variation in the coefficients.

As a final point, I note that both due to the methodology used to generate the estimates of the HIV rates (discussed in more detail below) and due to the nature of the epidemic, it is very likely

---

<sup>6</sup>These data come from countries Benin, Burkina Faso, Cameroon, Ethiopia, Kenya, Mali, Zambia and Zimbabwe. In all cases, responses are drawn from a survey question asking individuals their occupation.

the errors in the regressions are serially correlated. To address this, I run all regressions with adjustments for serial correlation. In addition, standard errors are adjusted for heteroscedasticity.

### 3 Data

The analysis described above requires data on new HIV infections, as well as data on exports and country-level controls. These are discussed in turn below.

#### 3.1 Data on HIV Infections

As discussed in the introduction, lack of good data on HIV incidence over time in Africa makes this analysis difficult. 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. 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, make it almost impossible to infer anything reliable about the new infection rate each year.

In recent years, population-based testing from the Demographic and Health Surveys (DHS) have provided much better prevalence estimates (ORC Macro, 2006). However, these estimates – even in the few cases in which we observe multiple years – are not sufficient to calculate incidence. To see this, denote incidence in country  $i$  in year  $t$  as  $\pi_{it}$  and prevalence as  $h_{it}$ . In addition, denote the number of deaths from HIV in country  $i$  in year  $t$  as  $d_{it}$ . The relationship between incidence, prevalence and deaths is given below.

$$\pi_{it} = h_{it} - h_{i,t-1} + d_{it} \tag{9}$$

It is clear from this equation that to calculate incidence, we must observe prevalence in the current year and in the past year, *and* observe deaths in the current year. The number of deaths are, in turn, a function of infection rates in the past. Simply observing two years of prevalence is not sufficient; we must observe the whole time pattern of prevalence, over the course of the epidemic. Given this constraint, I therefore use two data sources which provide a full time series of HIV prevalence over the course of the epidemic and allow me to calculate incidence.<sup>7</sup>

---

<sup>7</sup>This formulation ignores in and out migration. This could be an issue if migration is higher in periods of higher

## UNAIDS Data

The first dataset is from UNAIDS. UNAIDS is the United Nations organization responsible for reporting on and addressing the global HIV epidemic. UNAIDS has, for a number of years, put out estimates of HIV prevalence. However, they typically caution that these estimates are not comparable across reports (UNAIDS, 2008), making them difficult to use for calculating incidence. In 2008, UNAIDS announced a dramatic revision in their overall estimates of the magnitude of the HIV epidemic in Africa. Along with this revision, they released a full set of country-level HIV prevalence estimates, by year, for 1990-2007. These estimates are based on their best knowledge – from population-based testing, epidemic modeling and older antenatal clinic surveillance estimates – of levels and trends in the epidemic over time, and they provide the basis for the estimation of incidence.<sup>8</sup>

As noted above, to calculate incidence – even in recent years – it is necessary to know prevalence back to the start of the epidemic. For this reason, the UNAIDS data is not sufficient, since the epidemic was well underway by 1990. To fill in the gap between the apparent start of the epidemic in the mid-1970s and the start of the data in 1990, I use a combination of earlier trend data from UNAIDS and linear interpolation. More specifically, earlier UNAIDS reports from the countries in the sample typically included some time series information on infection rates at antenatal clinics through the 1980s. I use these data to calculate the trend in infection rates from the mid-1980s through 1990; the level is calibrated to match the 1990 data.<sup>9</sup> This data goes back to the early to mid-1980s; prior to that there was no test for the disease. To estimate rates in earlier years, I therefore calculate what epidemic growth rate would be necessary for the prevalence to increase from an assumed rate of zero in 1976 to the first observed year of data in each country. I use this growth rate to estimate prevalence.

The second step is calculation of incidence. As in Section 2, denote HIV prevalence in year  $t$  as  $h_t$  and incidence as  $\pi_t$ . Further, denote the chance of dying from HIV after  $\tau$  years of infection as

---

exports; these migrants are high-HIV individuals *and* they are counted in the receiving country prevalence. In practice, this is unlikely to be very important for the results for two reasons. First, the methods of estimating prevalence will tend not to capture migrants. Second, in the data, exports are uncorrelated with net migration (results available from the author).

<sup>8</sup>Another source for HIV data is the U.S. Census HIV/AIDS Surveillance Database, which reports testing data from antenatal clinics and other sources over time. It would, in principle, be possible to put these data together and (with some smoothing assumptions) generate consistent estimates of prevalence. This would require a number of additional assumptions, however. Moreover, the data from this testing should be incorporated in the UNAIDS estimates, which take into account all testing done in the country.

<sup>9</sup>Typically the infection level in earlier reports is higher than in the new UNAIDS data. To address this, I calibrate the trend to the 1990 level. For example, if the earlier reports have a rate of 1% in 1982, 5% in 1985 and 10% in 1990, and the new data reports a rate of 5% in 1990, I assume the rate was 5% in 1990, 2.5% in 1985 and 0.5% in 1982.

$c_t$ . Given this, it is possible to express incidence in year  $t$  as

$$\pi_t = h_t - h_{t-1} + \sum_{j=1}^{t-1} c_{t-j}\pi_j \quad (10)$$

That is, incidence is the difference between the stock of infections in year  $t$  and the stock in year  $t - 1$ , with an adjustment for deaths. The number of people who die in year  $t$  is equal to the number of people infected in year  $t - 1$ , multiplied by the chance of dying after one year, plus the number of people infected in year  $t - 2$ , multiplied by the chance of dying after two years, and so on. This makes it clear that the other important input to calculating incidence is time from HIV infection to death from AIDS.

Detailed data on time to death are difficult to generate, particularly in developing countries, since it requires knowing (roughly) the time of infection. For this reason, the best available data are drawn from developed countries, from time periods before HIV treatment was available (Collaborative Group on AIDS Incubation and HIV Survival, 2000). Researchers modeling the epidemic in Africa (Stover, 2003; Statistics South Africa, 2004), have used the “fast” version of these paths, which assumes that the trend in time to death is similar to the developed world, but the average time to deaths is faster (due to lower nutrition, higher disease burden, etc.). Figure 1 shows the time path to death used for this analysis; the values of  $c_t$  used in the calculations are drawn from the data underlying this figure.<sup>10</sup> Of course, these data on time to death are an approximation, and in the robustness section I will explore how the results change if time to death is faster or shaped differently.

The final output of this calculation is data on HIV incidence, by year, in 35 countries. The countries are listed in Appendix Table 1, which also reports summary statistics on average prevalence and incidence in each country. It is important to note that in this dataset estimates of prevalence for years after 1985 generally come directly from UNAIDS (either from their newly released estimates or from earlier trend data). Prior to the mid-1980s, the estimates come from linear interpolation. Although having these earlier years is necessary for calculating incidence, I do not want to rely heavily on them. I therefore limit the analysis to years after 1985.

---

<sup>10</sup>These data on time to death are estimates for a period *without* HIV treatment. I argue this fits the data, at least for a very large share of the period here. UNAIDS (2008) estimates that as late as 2002, only 50,000 people in Sub-Saharan Africa were receiving anti-retrovirals, which is about 0.2% of the infected population in that period. By 2007, which is the end of the period here, this was up to 2.1 million, which is closer to 10% of the population, and could make more of a difference. In the robustness section I will also explore how the results change if I allow for changes in the time to death in this later period.

## Mortality-Based Data

The second dataset includes estimates of incidence and prevalence based on inference from mortality data. The methodology and estimates are reported in detail in a companion paper (Oster, forthcoming). Here, I briefly outline how the estimates are produced.

The basic methodology is straightforward. Assume I observe deaths from HIV in a given year, denoted  $d_t$ . Retaining the notation from above, with  $\pi_t$  as new infections in year  $t$  and  $c_\tau$  as the chance of dying  $\tau$  years after infection, I can express  $d_t$  as

$$d_t = c_1\pi_{t-1} + c_2\pi_{t-2} + \dots + c_{20}\pi_{t-20} \quad (11)$$

Assuming that I observe  $d_t$  and the elements of the  $\mathbf{c}$  vector, the only unknowns in this equation are incidence values. Of course, with a single year of data on deaths I will not be able to solve for incidence in all years.

In order to solve for the full set of incidence data I again need to see data back to the start of the epidemic. Assume I observe some deaths in a year  $t - x$  such that there is no HIV infection before year  $t - x - 1$ . In that case,

$$d_{t-x} = c_1\pi_{t-x-1} \quad (12)$$

and from this it is possible to solve for  $\pi_{t-x-1}$ . If I then observe deaths in year  $t - x + 1$ , I can solve for  $\pi_{t-x}$ , since  $\pi_{t-x-1}$  is known:

$$d_{t-x+1} = c_1\pi_{t-x} + c_2\pi_{t-x-1} \quad (13)$$

Further backward induction yields the entire vector of incidence estimates.

These calculations require two pieces of data: estimates of time to death from HIV, which are again drawn from Figure 1, and estimates of death rates. Details on the calculation of death rates appear in Oster (forthcoming), but deserve some brief discussion here. Unfortunately, data on number of deaths from HIV is not widely available. To generate estimates of this parameter, I use information on overall total death rates alongside estimates of expected non-HIV death rates. The former data, on overall death rates, are drawn from the Demographic and Health Surveys sibling mortality histories. These histories have been used by other researchers to calculate death rates; details on their reliability (both in theory and practice) are given in Oster (forthcoming).

Expected non-HIV mortality rates are estimated from the United Nations Demographic Yearbook Historical Supplement. Obviously the ideal would be to observe a set of countries just like the countries in the sample but unaffected by HIV. This is not possible. Instead, I use data on death

rates from (mostly) non-African countries represented in the Demographic Yearbook Historical Supplement. To do this, I take advantage of the fact that there are some (limited) age groups where death rates should be largely unaffected by HIV. In particular, I argue that HIV-related deaths will be limited among children between the ages of 5 and 14; children infected at birth have largely died by their fifth birthday (without treatment) and deaths from sexually-transmitted HIV are unlikely to occur in this age group. For each country in the sample I use the DHS data to estimate death rates among children in this age group and merge these data with the Demographic Yearbook data. I select the comparison observations from the Demographic Yearbook as the 10% of the Demographic Yearbook country-years for which mortality in the 5-14 age group most closely matches the data from the HIV-affected countries. I then use these country-years to estimate counter-factual mortality among adults. Subtracting this expected mortality from the total mortality gives an estimate of mortality from HIV.

The mortality-based prevalence and incidence data is available for 11 countries – the data are limited to countries with sibling mortality histories in the Demographic and Health Surveys. In Oster (forthcoming) I present evidence suggesting that these data match well to population-based testing data, where the latter is available. These data are also highly correlated with the UNAIDS prevalence estimates: the correlation coefficient is 0.75 and the relationship between the two series is highly significant even in a regression with country and year fixed effects. The countries in these data are listed in Appendix Table 1, with summary statistics on incidence and prevalence. As with the UNAIDS data, I limit the analysis to 1985 and after. Because the death rates rely on retrospective reporting, data from further back in time will be less reliable.

Both the UNAIDS data and the mortality-based estimates have drawbacks. In the case of the UNAIDS data, the exact methodology used to generate the prevalence estimates is somewhat opaque, and some assumptions are necessary to generate incidence data from the prevalence measures. The methodology generating the mortality-based estimates is more transparent, but has more significant data requirements and can be used only for a smaller sample of countries. The hope is that by using both datasets and exploring, to the extent possible, robustness with respect to the assumptions generating the HIV incidence values, I can avoid the concern that the results are driven by some feature of the data generating process.

### **3.2 Data on Exports**

The other main data requirement is data on exports at the country-year level. I use three measures of exports. The first is total export value from the World Development Indicators (WDI). This is available directly from the WDI and represents the value of all exports; this measure has the advantage of being comprehensive.

In addition to this, I use data from the NBER-United Nations Trade Data described in Feenstra et al. (2004). This source provides detailed data on exports and imports by SITC code. This data is available only through 2000, but I collected data directly from Comtrade using the Feenstra et al. (2004) methodology to extend the data through 2007. There are at least three advantages of these data over the WDI. First, the data is based on information on imports from countries with reliable reporting, which addresses many of the concerns one might have about noisy reporting among African countries. Second, because exports are broken down by SITC codes it is possible to focus on only major exports, which is again useful in limiting noise. Third, they provide information on volume as well as value. Following the methodology used in Jones and Olken (2010), when calculating exports from these data I focus on SITC-4 codes that are observed continuously for the entire period from 1985 through 2007. This limits the data to the major exports, which for most of the countries in the sample are virtually the only exports.

Summary statistics on the three trade measures are shown in Panel A of Table 1.

### **3.3 Data for Mechanisms**

In Section 5 of the paper I discuss the mechanisms behind the export-HIV relationship and use a number of additional pieces of data. The first is data on GDP per capita, which I draw from the Penn World Tables (version 6.3). In addition, I use several pieces of data related to trucking. The first is the volume of truck imports by country, drawn from the NBER-United Nations data described above. The second is kilometers of paved roads by country, drawn from the World Development Indicators. Finally, I use data on HIV prevalence and distance to road by cluster in the Demographic and Health Surveys. These data are all summarized in Panel B of Table 1.

## **4 Results: HIV Incidence and Exports**

This section presents the primary results in the paper, estimating the relationship between exports and new HIV infections. Subsection 4.1 below presents the basic results on HIV and exports, and

Subsection 4.2 discusses issues of causality. Subsection 4.3 discusses the robustness of these basic results to the assumptions used to generate HIV incidence.

## 4.1 Primary Results

Panel A of Table 2 shows the initial estimates of the impact of exports on HIV incidence using the basic controls: country and year fixed effects and lagged HIV prevalence rate. Columns 1-3 focus on the UNAIDS data, whereas Columns 4-6 show results for the mortality-based estimates. The coefficient on exports is significant and positive in all but one specification. For the UNAIDS data, where I have the largest sample size, I find significant impacts regardless of which measure of exports I use. For the mortality-based data, the estimates are significant for the two export value measures, but not for the export volume measure. This last coefficient is also positive, but the p-value is 0.26. Despite this exception, the overall evidence suggests a strong positive relationship between exports and HIV incidence.<sup>11</sup>

There are two ways to interpret magnitudes here. First, I can interpret the magnitudes directly from the coefficient estimates. In Column 1, the results suggest that a doubling of export value leads to an increase of 0.07 percentage points in the incidence rate; this is on a base of 0.7 percentage points, so this represents about a 10% increase in incidence. Second, as described in Section 2, it is possible to combine the coefficient estimates with the assumed model parameters and the average HIV prevalence rate to estimate a value for the  $\gamma$  parameter. This value is reported for each coefficient at the bottom of the panel. These estimates of  $\gamma$  range from 0.10 to 1.00, suggesting that a doubling of exports leads to somewhere between a 10% increase and a doubling of the share of people in the high risk group. These values do not seem unrealistic. A doubling of exports leading to a doubling in the share of people in the high risk group suggests there are limited economies of scale: if you export twice as much, you need twice as many truckers, miners, etc. It seems likely there are *some* economies of scale, which is consistent with the fact that most of the estimates of  $\gamma$  are significantly less than 1.

Comparing the two HIV measures, we see that they differ somewhat. The estimates with the mortality-based data are larger, but less consistent over the three export measures. This may be

---

<sup>11</sup>The coefficient on lagged HIV prevalence in these regressions is roughly zero; positive in the case of the UNAIDS data and negative for the mortality-based data. The fact that this is not positive and significant seems puzzling. In practice, the lack of a positive and significant coefficient here appears to be due to the serial correlation correction, which particularly impacts this coefficient because of the high degree of serial correlation in this variable. A simpler, non-adjusted regression does typically yield a positive coefficient. It is also worth noting it is somewhat difficult to interpret these coefficients once I adjust for country and year fixed effects, since so much of the variation is soaked up in that case.

due to the mortality-based data having more year-on-year variation than the UNAIDS data, which results in higher coefficient estimates. It is also possible this difference reflects the differences in the HIV prevalence rate in the two samples, since the mortality-based data focuses on more heavily infected countries. As discussed in Section 2, this is an inherent issue with comparing the simple export coefficients across samples.

To address this, Panel B of Table 2 shows the estimates of Equation (8); in this specification, I interact lagged HIV prevalence with exports. As discussed above, we do not include the level effect of exports, which is assumed to be zero; Online Appendix Table 1 demonstrates that this assumption is appropriate. The interaction results strongly support the theory: the coefficients are consistently positive and significant in all but one case. Further, the one insignificant coefficient is approaching significance in this regression (p-value of 0.15). At the bottom of the table we report estimates of the implied  $\gamma$ . These still have a fairly large range (0.3 to 1.0), but smaller than the range we observed in the level effects.<sup>12</sup>

One issue which deserves mention is the fact that the one case in which we do not see a significant relationship in Table 2 is the regression of mortality-based incidence on export volume. To the extent that the theory is that exports increase the share of high risk people (truckers, miners) we might expect export volume to be more predictive than export value. Consistent with this, with the UNAIDS data we find (in Panel B) that the precision is higher when we look at export volume. For the mortality-based data this is not the case. It is possible – even likely – that this is due to the fairly small sample size and the fact that the data on export volume is noisier than the value data. A second possibility is that export value may be more predictive than export volume because it is more closely linked to income; if part of the mechanism by which exports increase risky behavior is through income, this would predict value would matter more. I discuss this mechanism in more detail in Section 5.

The evidence in Table 2 points to a positive relationship between exports and HIV; encouragingly, the evidence in the interaction regressions (Panel B) is supportive of the implication of the model I outline in Section 2. Further, the implied values for  $\gamma$  indicate at least broadly plausible magnitudes of the relationship between exports and the share of high risk individuals.

---

<sup>12</sup>As noted, our estimates are adjusted for serial correlation and heteroscedasticity. We do not allow for more general clustering by country, since we argue that the primary issue with the standard errors is serial correlation. However, in Online Appendix Table 2 we present our primary estimates with clustered standard errors calculated based on the bootstrap procedures described in Cameron, Gelbach and Miller (2008). This procedure is appropriate when the number of clusters is small, although it is unclear how appropriate this is for serial correlation corrected estimates, which is why we do not use it in the primary analysis. The standard errors are similar, although obviously larger given the loss of efficiency in the broader clustering.

Before turning to a discussion of identification and robustness, it seems worthwhile to briefly explore heterogeneity in these coefficients across space and time. To give a visual sense of the data, Figure 2 shows graphs of exports and incidence, over time, for the countries in the sample (this graph uses the export value measure from the NBER).<sup>13</sup> In some cases we can see both UNAIDS incidence and mortality-based incidence; in others only the UNAIDS data is observed. In general, these figures show evidence consistent with what we see in Table 2 – exports and incidence are moving together. This is, of course, not universally true, but does seem to be evidenced in a large number of countries.

Table 3 shows the central coefficients in the paper estimated separately for three regions of Africa: East Africa, Southern Africa and West Africa (countries in each region are listed at the bottom of the table). This is only possible to do for the UNAIDS data, where we have more countries. The two panels of this table are the same as Table 2: Panel A estimates the level effects of exports and Panel B estimates the interactions with HIV prevalence. Based on the level effects, we observe East Africa, in particular, and Southern Africa to some extent have very large impacts. There is no evidence of an effect of exports on HIV incidence in West Africa. However, in comparing coefficients across regions we again come up against the issue of the interaction with HIV rate; there are large differences in HIV prevalence across regions, making the coefficients not directly comparable.

The cross-region comparisons in Panel B of Table 3 are much more appropriate, as the differences in prevalence are adjusted for. Consistent with this, we find much more comparable coefficients. The relationship is still strongest in the East African countries, but the difference with the baseline coefficient is smaller. In these regressions we also see more consistency across East Africa and Southern Africa. However, the relationship still does not seem to hold in West Africa. The lack of a relationship in this region may be due to the more limited extent of the epidemic there. This difference may also be due to the fact that there is more focus on exports that are less associated with trucking, such as oil. These results do suggest that this export-HIV relationship is context specific, which is important to keep in mind when thinking about generalizability.

We can also explore heterogeneity in the results over time. The overall sample runs from 1985 through 2007. However, for the UNAIDS data in particular, the period from 1990 onwards represents a slightly different method of data construction (i.e., not relying on the older reports). This also represents a period in which the epidemic was growing more slowly and leveling off in a

---

<sup>13</sup>For space reasons, and because they are much less informative, this graph leaves out a few countries where I have data on exports only after 2000.

number of countries. The final row in each panel of Table 3 shows coefficient estimates when the data is limited to the period from 1990 onwards. The length of time before 1990 (1985-1989) is too short to estimate the relationship for that period separately.

For the mortality-based data, the post-1990 estimates look very similar to the overall estimates regardless of how we estimate the relationship. For the UNAIDS data, the coefficient estimates are smaller in the period after 1990. This is particularly true in Panel A; when we estimate the coefficient on interactions in Panel B we see estimates much closer to the overall effect and with similar significance. The lower level of significance and smaller coefficients in this latter period may reflect the lack of variation in incidence after 1990, since the UNAIDS data in this period is more extensively smoothed. It may also reflect the fact that there is more limited variation in incidence in this period simply due to the lower growth of the epidemic. Despite this, the consistency of results in Panel B is encouraging.

Overall, the results in these two primary tables suggest there is a positive relationship between HIV and exports and, by extension, between risky sex and exports. At the moment we are agonistic about what specific mechanism drives this relationship; possibilities include export-induced changes in income, and export-driven changes in truckers or migrants. Before going into more detail about the role of these two mechanisms, I first attempt to address at least some of the concerns about whether the HIV-export relationship is causal.

## 4.2 Identification

The regressions shown thus far demonstrate a relationship between HIV incidence and exports controlling for country and year fixed effects. This means that the results are not driven by any variables which differ across countries, or any overall time trends. In this subsection I provide some additional evidence in favor of a causal interpretation of this relationship. It is important to note that this issue is distinct from the question of what mechanism drives this relationship, which I will discuss a bit more in the next section.

A first issue with identification is the possibility of reverse causality: higher HIV causing higher exports. This seems very unlikely. First, in the time frame I am considering – variations at the yearly level – it is hard to see how more new infections (which would be asymptomatic) could affect the economy. Second, even to the extent that HIV might affect exports in the long run (for example, by debilitating the workforce) the relationship would likely be in the other direction: more HIV would mean lower exports, not higher. For both of these reasons reverse causality should not be

a major concern.

A similar timing-related logic applies to theories that the results are driven by a relationship between some other cause of death and exports. For example, if road accidents are an important source of mortality, they might go up during periods with more exports, when there is more trucking. However, this could not drive the results, because what I am estimating is the relationship between new HIV *infections* and exports, not HIV deaths and exports. Additional mortality from other causes in any given year would be attributed to HIV infections an average of 10 years in the past and not to HIV infections in the current year. If anything, this type of story is likely to dampen the results. A final concern is that higher exports this year might lead to higher income in the long-term and result in lower mortality. This would mean high exports now would be associated with low mortality in the future; this low mortality would show up in the data as an underestimate of incidence. This would mean high exports associated with low incidence; again, biasing against finding a positive result.

A second issue lies in omitted variable bias. Again, I note that the inclusion of country and year fixed effects means that any confounding factor would have to also vary within a country over time. I can take this a step further by estimating the regressions using these fixed effects *plus* country-specific trends. In some sense using such extensive controls with a limited sample size may be a concern, and make it somewhat harder to understand where the identification is coming from. However, to the extent that the results hold qualitatively with these trends included, this excludes a large set of possible confounds. For example, one possible confounding factor is changes in government: if a new government comes in that cares a lot about the economy, but not about health, we could see an increase in exports but a decrease in HIV prevention. However, these policy changes are unlikely to happen quickly, and including country-specific trends should capture them. A similar point can be made about confounding impacts of HIV knowledge or prevention spending: although these change over time, the change is largely a smooth increase, which should be captured by the inclusion of trends below.

The first row in each panel of Table 4 repeats our baseline estimates for comparison. The second and third rows in each panel estimate the regressions from Table 2, but with controls for either country-specific linear trends or country-specific quadratic trends. We should note that the quadratic trends, in particular, are asking a lot of the data given the relatively limited sample size. Again, Panel A estimates level effects and Panel B estimates the interactions with HIV prevalence. For the mortality-based data, the estimates are similar regardless of the trend estimates used. For

the UNAIDS data, the inclusion of these trends (particularly the quadratic trends) has some impact on the size and significance of the level effects (Panel A). However, the interaction effects remain very similar regardless of which trends we include.

As a second type of test, I can explore the causality issues broadly by looking at whether *future* exports drive current HIV rate. If true, this would suggest either some omitted variable influence or, in general, call into question the specification and data. To explore this, I regress each measure of HIV on current exports and future exports, with leads of different lengths (2-10 years).<sup>14</sup> Figures 3a and 3b graph coefficients on present and future exports from these regressions. These figures use the measure of export value from the NBER; figures for the other measures are similar and are shown in Online Appendix Figures 1-4. For both HIV measures the coefficient on current exports is consistently positive and significant in all specifications. I note that the coefficient varies some, particularly in the case of the UNAIDS data, simply because the sample becomes more restricted (i.e., to early years) when I include very long forward leads. In Figure 3a, focusing on the UNAIDS data, the coefficient on future exports is small and insignificant at all leads. In Figure 3b, focusing on the mortality-based data, this is largely true as well, although the coefficient 7 years out is significant. Given the lack of significance for the other leads it seems likely this just reflects noise in the data. Overall, however, these figures are strongly supportive of the results. Exports this year matter; exports in the future do not.<sup>15</sup>

These analyses provide support for the claim that the relationship between HIV and exports is causal, at least in the sense that if exports increase, we should expect to see HIV increase. The next subsection discusses the robustness of this result to the assumptions which generate the mortality-based incidence; I then move to briefly discuss what mechanisms might drive this relationship.

### 4.3 Robustness to Assumptions Generating HIV Data

One important concern with the regressions above is that the relationship between HIV and exports is being driven by the assumptions that generate incidence, rather than by the underlying relationship. For both the UNAIDS data and, especially, the mortality-based data, assumptions go into converting either the prevalence data (for UNAIDS) or data on deaths (for the mortality-based

---

<sup>14</sup>I begin with two years, rather than one year, since noise in the data is such that infections from next year might well be attributed to this year, making this a less powerful test at very short leads.

<sup>15</sup>The coefficients here come from multiple regressions (one for each lead); the results look very similar if I run a single regression with leads of different lengths included.

estimates) into incidence rates. In this section I briefly consider how the relationship changes if I vary two of the central assumptions: the time path to death from infection and the epidemic start date.

To address the issue of time to death, I first generate both the UNAIDS incidence and the mortality-based incidence using two alternative paths of time to death. One of the alternative paths features a faster time to death: this 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 1 (both alternative paths are illustrated in Online Appendix Figure 5).

In addition to these two alternatives, it may be important to consider how the availability of treatment could change these results. As noted in the data section, treatment levels before 2002 were extremely low – UNAIDS estimates that only 0.2% of infected individuals in Sub-Saharan Africa were receiving treatment by 2002. This suggests that, for the mortality-based data, which is all from the pre-2002 period, this change is not important. The UNAIDS data, however, goes through 2007, at which point 10% of HIV-positive individuals were receiving treatment (UNAIDS, 2008). To explore this, as a third alternative I assume the standard time to death prior to 2002 and then assume the death rate is cut in half after that. This is obviously an extreme change, but it should give an upper bound on how this impacts the results.

Table 4 shows the primary coefficient estimates with these varying assumptions. Making the time to death either flatter or faster (rows 4 and 5 in each panel) prompts only a limited difference in the results with either dataset. In the case of the mortality-based data, the coefficients are, on average, larger but not significantly so. Changing the assumptions on treatment (row 6) also makes only a limited difference in the coefficients.

I also consider how the results change if I change the assumed “start date” of the epidemic by four years in either direction (either assume the epidemic starts 4 years earlier or 4 years later). These estimates are shown in the rows 7 and 8 of each panel of Table 4. Again, these changes make little difference to the results, for either the UNAIDS data or the mortality-based data. Assuming a later start date increases the coefficients in both sets of regressions but, again, these are within the margin of error of the primary results.

Overall, the results in Table 4 indicate that, while the exact magnitude of the relationship between HIV and exports can be somewhat affected by the assumptions about generating incidence, the general conclusions of this section are robust to changes in the important elements of the data

generation.<sup>16</sup>

## 5 Mechanisms: Income versus Trucking and People-Movement

The results thus far suggest that changes in exports impact HIV and I have made the case that this relationship is causal. In Section 2, I suggested a broad mechanism behind this relationship: increased exports increases the share of individuals engaged in risky behavior. There are at least two more specific mechanisms that might drive this. A first possibility is that this is driven by changes in income. The causal channel in this case would be: increased exports  $\rightarrow$  increased income  $\rightarrow$  more risky sex. A second possibility is that this is driven by changes in the share of people involved in high risk export-related activities like trucking and mining. In this case, the causal channel would be: increased exports  $\rightarrow$  more truckers, miners, etc.  $\rightarrow$  more risky sex. It is beyond the scope of this paper to fully separate out these mechanisms and allocate shares to them. However, in this section I provide some suggestive evidence on the role that each plays.

### 5.1 Income

I begin with income. It is generally thought that sex is a normal good; people with higher income have more sex. By extension, increases in income should increase risky behavior and HIV (this could also go through a sex worker channel, if consumption of sex work increases when people are richer). We can see some evidence of this relationship without any reference to exports in Online Appendix Table 3, which runs regressions of the form reported in Panel A of Table 2 but replaces log exports with log GDP per capita (from the Penn World Tables). In general, income is positively related to HIV, although it is significant only in the UNAIDS data.

From the perspective of the export argument, the key issue is whether including a control for income in the main regressions mitigates the impact of exports and, if so, by how much. If the entire impact of exports is driven by changes in income, then we would expect including an income control to eliminate the export impacts. In contrast, if the impact of exports is driven in part by other mechanisms (truckers, etc.) then the export impact should survive the GDP control.

This test is done in Table 5. In this table I report results from the regressions from Panel A of Table 2, but add the GDP control. I find that adding the GDP control has some impact on the

---

<sup>16</sup>It is worth noting that there are a number of other assumptions that go into generating the mortality-based data, which do not affect the UNAIDS data. For example, due to noise in the death data I do some smoothing over years, requiring the choice of a particular smoothing parameter. Oster (forthcoming) discusses in more detail the robustness of the prevalence estimates to these assumptions.

export results, although the effects are fairly minimal. Comparing the coefficients there to the corresponding coefficients in Table 2, I find that between 4% and 20% of the export impact seems to be explained by adding the GDP control. The impact of GDP in these regressions is positive, although typically not significant, or only marginally so. It is perhaps surprising that we do not see stronger effects of income even in these regressions, although this may simply point to the fact that GDP is measured with noise, and may not directly capture the type of income most important for driving risky sex.<sup>17</sup>

## 5.2 Trucking, People-Movement

The evidence in Table 5 suggests that income explains a small share of the export effect; the majority of the effect *cannot* be explained by increasing income. I turn next to the second channel: movement of people. More exports mean more production, which may use migrant or other temporary labor, and more trucking to move things from the place of production to cities or ports. Here, I provide three pieces of evidence in favor of this hypothesis. It is important to note that while this evidence is suggestive, it will fall short of proving this is *the* mechanism driving these results.

*Truckers and Sex; Trucks and Exports* The first argument here is simply that truckers have more sex than individuals in the general population, and more exports means more trucking. The relationship between truckers and sex is well known in existing work and, in fact, existing evidence points to a number of ways in which trucking and HIV are linked. 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; Anarfi et al., 1997; Anarfi, 1993; Orubuloye et al., 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, they 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). The level of risky behavior is significant; in a survey in Nigeria, Orubuloye et al. (1993) find that the truck drivers in the sample have an average of 12 non-marital partners a year.

This existing evidence suggests a link between trucking and sex. Panel A of Table 6

---

<sup>17</sup>It is difficult to do a parallel analysis for the interaction regressions. Simply controlling for GDP is not helpful since the coefficients are interacted; but including interactions between GDP and HIV prevalence will overstate the importance of GDP because of the correlation generated by the prevalence interaction. We therefore focus on the level effects here.

provides evidence in favor of the second step: more exports mean more trucking. Data on actual quantity of goods trucked is available for only a very limited sample of country-years, making it difficult to use in a direct analysis. As an alternative, I use the NBER-United Nations data to generate a measure of the weight of trucks imported by country and year. I argue that if increased exports leads to increased trucking, then this should be reflected in increased need for trucks (either to increase total quantity or to replace trucks that have broken down due to the additional usage). I generate a measure of the weight of trucks imported and relate this to the export measures.

Columns 1-3 of Panel A of Table 6 show basic regressions of log weight of trucks imported on log exports (for the three export measures) with country and year fixed effects included.<sup>18</sup> The coefficients are consistently positive and significant: in years with more exports, more trucks are imported. Columns 4-6 show that this does not simply reflect an increase in overall imports; these columns add log of total import value in dollars (again from the NBER, and excluding imports of trucks) and the coefficient on export remains significant.

The coefficient on exports (0.20 to 0.40) suggests that a doubling of exports leads to between a 20% and 40% increase in the number of trucks imported. It is worth noting that this lines up fairly closely with the average value of  $\gamma$  estimated above (about 0.30). The estimates of  $\gamma$  were interpreted as indicating that a doubling of exports leads to about a 30% increase in the share of individuals engaged in risky sex. If I think these are primarily truckers, this is broadly consistent with the 30% increase in trucking imports. Obviously these numbers are not directly linked, but they do suggest some consistency in magnitudes.

All of this evidence – the evidence that truckers are high risk and the link between exports and truck imports – suggest a role for trucking in mediating the export-HIV relationship. If having more exports means there are more trucks, and truckers have more sex when they are away from home than not, these together imply that an increase in exports should increase risky sex. The fact that the magnitude of the export-truck import relationship is consistent with the magnitude of the HIV-exports relationship is also supportive.

*Variations across Countries: Roads* I also present two analysis which look more directly at the role of roads in mediating this relationship. First, I take advantage of variation in road availability across countries in the sample. To the extent that an important mechanism here is

---

<sup>18</sup>The sample size is smaller in these regressions because I have these data only for the years and countries covered by the NBER-United Nations data. I do not have this for the Comtrade extension of the data. This means I cover only through 2000, and do not have data for the following countries: Botswana, Equatorial Guinea, Eritrea, Lesotho, Namibia and Swaziland.

trucking or movement of people, we would expect the effect of exports to be stronger in areas where increases in economic activity will have a larger effect on these variables. This is likely to be more true in countries with more roads. Put differently, if we see an equally strong relationship between exports and HIV in countries with no roads, or very limited roads, this would suggest that trucking is *not* an important mechanisms. Conversely, if we see a stronger relationship in countries with extensive road networks, this could point toward the importance of transit and movement in mediating the export-HIV relationship.

I use data on total kilometers of paved roads by country from the World Development Indicators (using data on all roads yields similar results). This measure does not vary much over time, so I use a constant value per country, representing the average over the sample period. I estimate the primary specifications from Panel A of Table 2, but include an interaction between the road measures and exports. I also include an interaction between total land area and exports, to ensure the results are not driven by differences by country size.<sup>19</sup>

These regressions are shown Panel B of Table 6, with Columns 1-3 showing the results for the UNAIDS data, and Columns 4-6 showing the results for the mortality-based data. In all but the first column, the coefficient on the interaction is positive and significant. In other words, the export-HIV relationship is larger in areas where there are more roads, pointing to the importance of transit in this relationship.

*HIV and Roads* As a final piece of evidence on the trucking and people-movement mechanism, I turn to a slightly different type of test. To the extent that truckers are an important part of the epidemic, and that increasing their numbers increases HIV, one clear prediction is that HIV rates should be higher in areas closer to major roads. It is possible to test this using data from a number of recent DHS which provide information on (a) HIV prevalence and (b) GPS location of the survey cluster. These data, along with GIS data on road networks in Africa, allow me to test whether HIV prevalence is higher in areas closer to roads. It is important to be clear that this test does not directly reference exports. For almost all areas, the DHS data on HIV rates is available for only one point in time, so it is not possible to perform this test to directly relate incidence and exports. What I *can* do, however, is explore how distance from the road is related to HIV prevalence. This test is similar to that implemented in Djemai (2009).

To do this, I generate two measures of HIV prevalence by survey cluster for the available

---

<sup>19</sup>It is actually not obvious whether it is appropriate to control for country size or not, since the theory is really just that more roads lead to more infections, and not that they do so adjusting for country size. The control seems in order, however, given the possibility of other characteristics which differ by country size.

countries in the DHS (countries are listed in the table notes): log HIV prevalence and whether any individuals in the cluster tested positive for HIV (a cluster has an average of 30 people). I regress these measures on distance from the cluster to the nearest road, restricting the data to non-urban clusters. I include fixed effect for district (smaller than a country), as well as controls for cluster demographics (age, education, employment, religion). These regressions are shown in Table 7. For either measure of HIV, using either a continuous or discrete measures of distance to the road, we see a significant relationship between the two. Areas that are closer to the road have higher HIV prevalence. These effects are actually fairly large.

This relationship demonstrates a direct link between roads and HIV prevalence. Again, it is important to emphasize that this *does not* directly relate to exports, but it does provide evidence that transport and HIV seem to be closely related.

## 6 Conclusion

This paper addresses the connection between HIV and economic activity, arguing that increases in economic activity 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 important to note in interpreting all of the results in the paper 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, 2009)). 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 that drives HIV, and is not captured here. Although the estimates here may point to an important role for exports in explaining within country variation they by no means explain all of the cross-country or cross-region variation.

## References

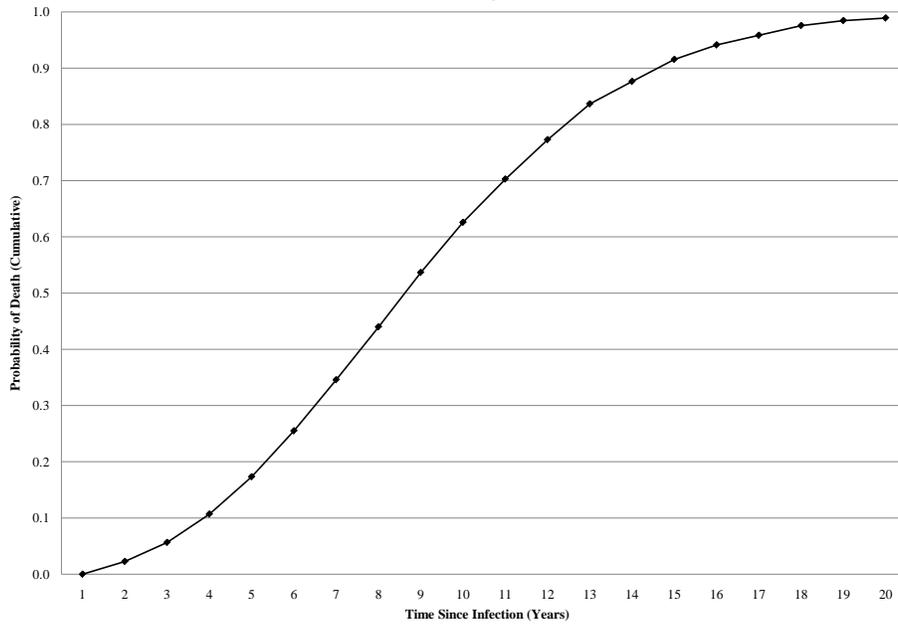
- Ahlberg, Dennis and Eric Jensen**, *Confronting AIDS: Evidence from the Developing World*, European Commission of World Bank, 1998.
- Anarfi, John**, “Sexuality, migration and AIDS in Ghana – A socio-behavioral study,” *Health Transition Review*, 1993, 4, 273–295.
- , **Ernest Appiaha, and Kofi Awusabo-Asareb**, “Livelihood and the risk of HIV/AIDS infection in Ghana: the case of female itinerant traders,” *Health Transition Review*, 1997, 7-8.
- Arbache, Jorge, Delfin Go, and John Page**, “Is Africa’s Economy at a Turning Point?,” *World Bank Policy Research Working Paper No. 4519*, 2008.
- 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.
- Cameron, Colin, Jonah Gelbach, and Douglas Miller**, “Bootstrap-Based Improvements for Inference with Clustered Standard Errors,” *Review of Economics and Statistics*, 2008, 90 (3), 414–427.
- Case, Anne and Angus Deaton**, *Health and Wellbeing in Udaipur and South Africa*, University of Chicago Press, 2006.
- 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.
- Djemai, Elodie**, “How do roads spread AIDS in Africa? A Critique of Received Policy Wisdom,” *Mimeo, Toulouse School of Economics*, 2009.
- Dollar, David**, “Is Globalization Good for your Health?,” *Bulletin of the World Health Organization*, 2001, 79 (9), 827–833.
- , “Lessons from China for Africa,” *World Bank Policy Research Working Paper No. 4531*, 2008.
- Feenstra, Robert, Robert Lipsey, Haiyan Deng, Alyson Ma, and Hengyong Mo**, “World Trade Flows: 1962-2000,” *NBER Working Paper No. 11040*, 2004.
- Fortson, Jane**, “HIV/AIDS and Fertility,” *American Economic Journal: Applied Economics*, 2009, 1 (3), 170–194.
- , “Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa,” *Review of Economics and Statistics*, forthcoming.
- Halperin, Daniel and Glenn Post**, “Global HIV Prevalence: The Good News Might be Even Better,” *The Lancet*, 2004, 364 (9439), 1035–1036.
- Jones, Ben and Ben Olken**, “Climate Shocks and Exports,” *NBER Working Paper No. 15711*, 2010.

- Kalemi-Ozcan, Sebnem**, “AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa,” *NBER Working Paper No. 12181*, 2006.
- Levine, David and Dov Rothman**, “Does Trade Affect Child Health,” *Journal of Health Economics*, 2006, *25*, 538–554.
- 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.
- McNeill, William**, *Plagues and Peoples*, Garden City, NY: Anchor Press, 1976.
- 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**, “Sexually Transmitted Infections, Sexual Behavior and the HIV/AIDS Epidemic,” *Quarterly Journal of Economics*, 2005, *120* (2), 467–515.
- , “HIV and Behavior Change: Why Not Africa?,” *NBER Working Paper No. 13049*, 2009.
- , “Estimating HIV Prevalence and Incidence in Africa From Mortality Data,” *BE Press: Topics in Economic Analysis and Policy*, forthcoming.
- Owen, Ann and Stephen Wu**, “Is Trade Good for your Health?,” *Review of International Economics*, 2007, *15* (4), 660–682.
- Pritchett, Lant and Lawrence Summers**, “Wealthier is Healthier,” *Journal of Human Resources*, 1996, *31* (4), 841–868.
- Ruhm, Christopher**, “A Healthy Economy Can Break Your Heart,” *Demography*, 2007, *44* (4), 829–848.
- Statistics South Africa**, “Mid-year population estimates, South Africa,” Technical Report, Statistics South Africa 2004.
- 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.
- Thornton, Rebecca**, “The Demand for, and Impact of, Learning HIV Status,” *American Economic Review*, 2008, *98* (5), 1829–1863.
- UNAIDS**, *2008 Report on the global AIDS epidemic*, Joint United Nations Program on HIV/AIDS, 2008.

**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.

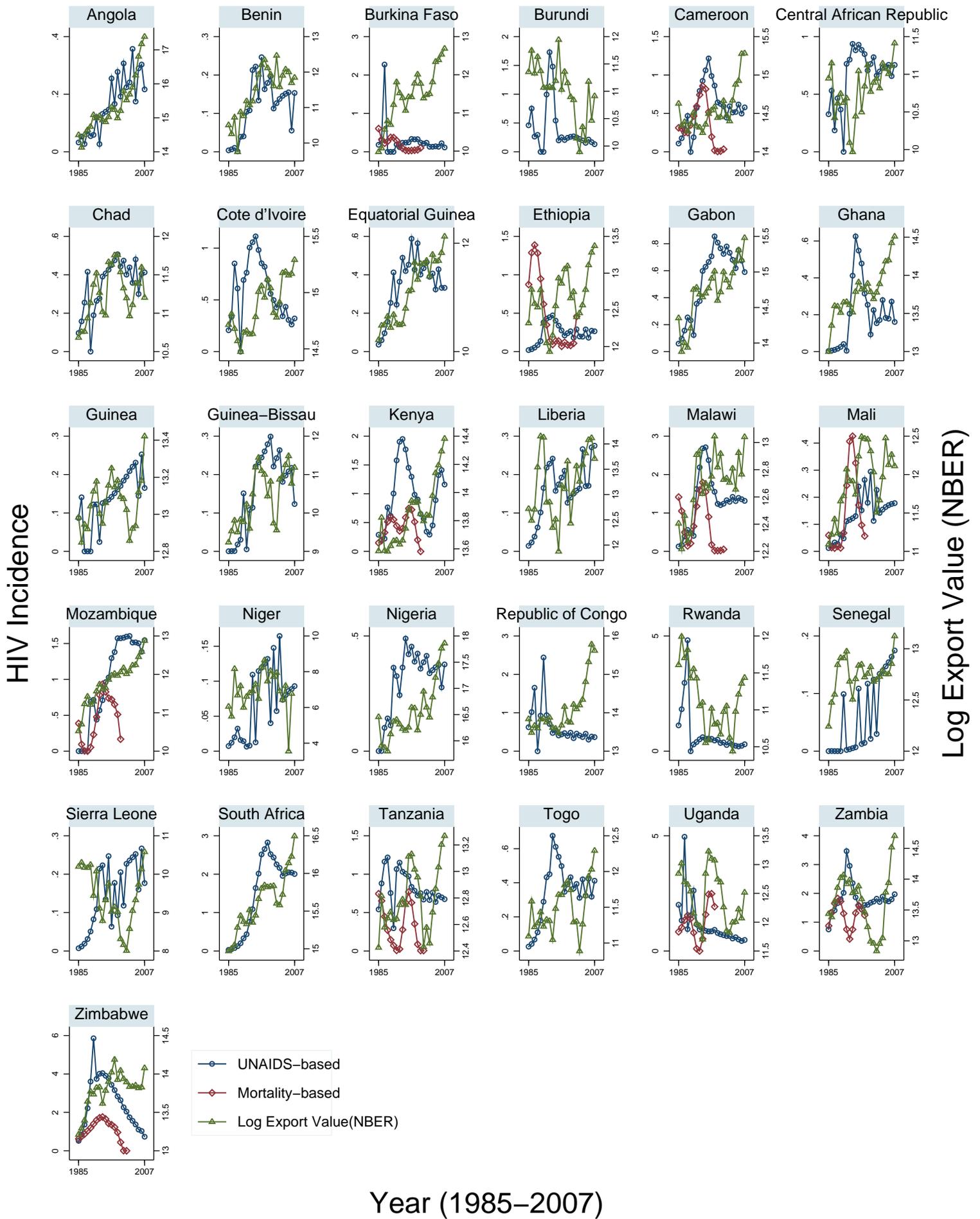
**Wilson, Mary**, “Travel and the Emergence of Infectious Disease,” *Emerging Infectious Diseases*, 1995, *1* (2), 39–46.

**Figure 1:  
Cumulative Probability of Death from HIV**

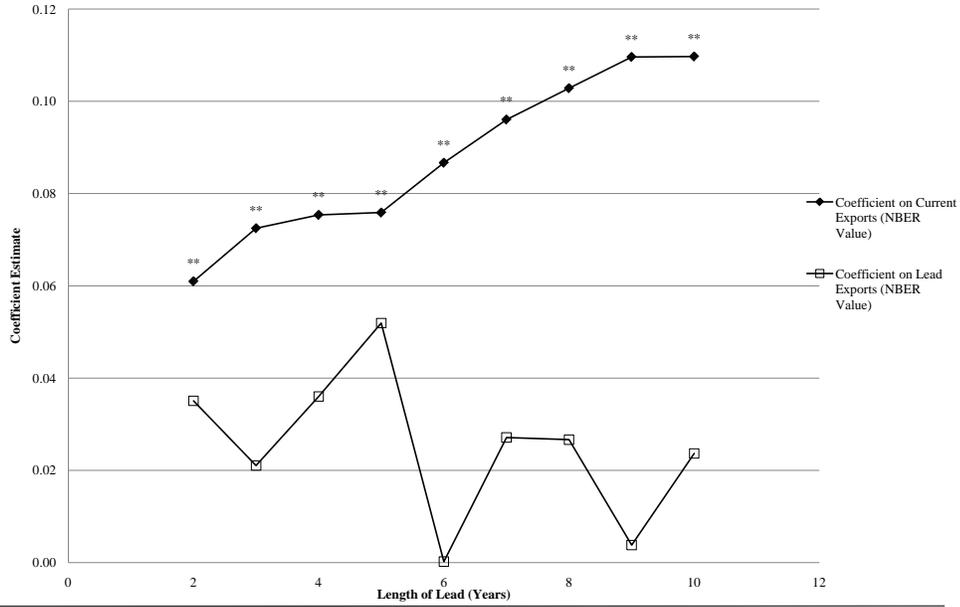


Notes: Figure presents the cumulative probability of death as a function of years from infection with HIV. The figure is based on data from Collaborative Group on AIDS Incubation and HIV Survival, 2000.

# Figure 2: HIV Infections and Exports, by Country

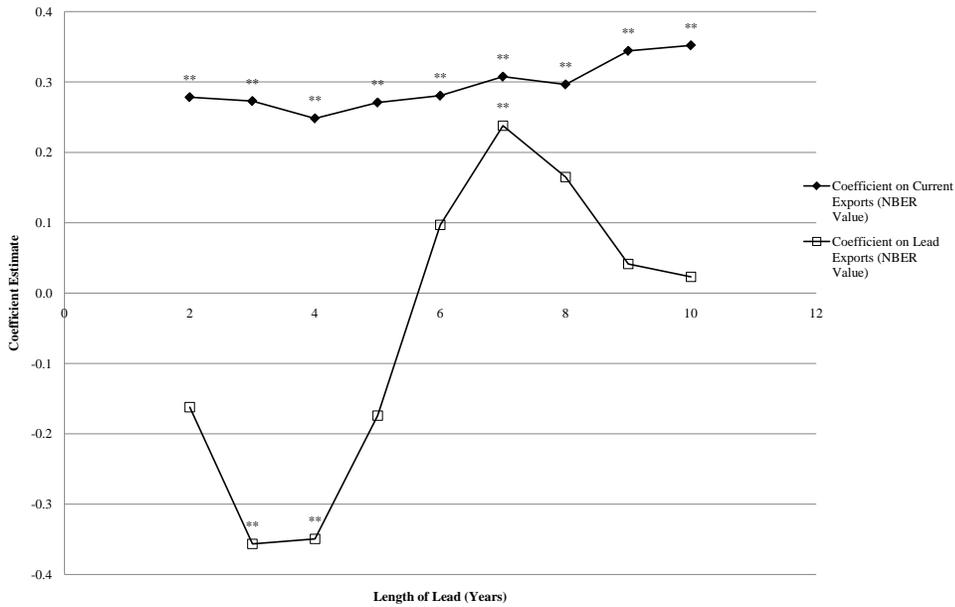


**Figure 3a:**  
**Effect of Current and Future Exports on HIV Incidence, UNAIDS Data**



Notes: This figure shows coefficients on current and future exports from regressions of the form in Table 2. The coefficients come from a total of nine regressions, each of which includes current exports and exports at a given lead length. \*\* significant at 5% level. The export measure used is the NBER Export Value. The Online Appendix shows similar graphs for the other export measures.

**Figure 3b:**  
**Effect of Current and Future Exports on HIV Incidence, Mortality-Based Data**



Notes: This figure shows coefficients on current and future exports from regressions of the form in Table 2. The coefficients come from a total of nine regressions, each of which includes current exports and exports at a given lead length. \*\* significant at 5% level. The export measure used is the NBER Export Value. The Online Appendix shows similar graphs for the other export measures.

**Table 1. Summary Statistics on Trade, Mechanism Analysis Data**

<b>Panel A: Trade Measures</b>			
	<b>Mean</b>	<b>Std. Dev.</b>	<b>Number of Obs.</b>
Export Value (in millions), WDI	\$3098	\$8448	793
Export Value (in millions), NBER/Comtrade	\$1556	\$4505	747
Export Volume (in mill. of kg), NBER/Comtrade	6115	17,503	747
<b>Panel B: Income, Road Density, HIV and Road Distance</b>			
	<b>Mean</b>	<b>Std. Dev.</b>	<b>Number of Obs.</b>
GDP per Capita	\$1921	\$2246	747
Truck Imports (in thousands of kg)	5884	7704	592
Paved Roads (in thousands of km)	7.62	14.73	724
<b>DHS Data</b>			
Cluster HIV Prev. (1-100)	7.11	11.42	3672
Any HIV in Cluster	.451	.497	3672
Distance to Road (km)	4.46	6.41	3672

Notes: This table shows summary statistics on the trade measures and variables used in the mechanism section. Summary statistics on HIV measures, by country, are in Appendix Table 1.

**Table 2. HIV Incidence and Exports**

<i>Dependent Variable:</i>	<i>Incidence Rate (0-100)</i> <i>UNAIDS Data</i>			<i>Incidence Rate (0-100)</i> <i>Mortality-Based Data</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Level Effects</b>						
Explanatory Variables:						
Log Export Value (WDI)	.073*** (.027)			.474*** (.176)		
Log Export Value (NBER)		.072** (.031)			.278** (.120)	
Log Export Volume (NBER)			.048** (.025)			.091 (.081)
HIV Prev., t-1	.023 (.028)	.022 (.028)	.023 (.029)	-.063* (.036)	-.056 (.035)	-.072* (.037)
<i>Controls in all Columns: Country and Year Fixed Effects</i>						
<b>Implied <math>\gamma</math></b>	<b>.156</b>	<b>.152</b>	<b>.102</b>	<b>1.00</b>	<b>.576</b>	<b>.192</b>
Number of Observations	720	747	747	161	166	166
<b>Panel B: Lagged HIV-Export Interactions</b>						
Explanatory Variables:						
Value, WDI $\times$ HIV Prev., t-1	.026** (.010)			.092*** (.025)		
Value, NBER $\times$ HIV Prev., t-1		.026** (.012)			.062*** (.017)	
Vol., NBER $\times$ HIV Prev., t-1			.020** (.008)			.025 (.017)
HIV Prev., t-1	-.541** (.259)	-.342* (.194)	-.256* (.144)	-1.97*** (.532)	-.868*** (.223)	-.388 (.221)
<i>Controls in all Columns: Country and Year Fixed Effects</i>						
<b>Implied <math>\gamma</math></b>	<b>.288</b>	<b>.288</b>	<b>.222</b>	<b>1.02</b>	<b>.688</b>	<b>.278</b>
Number of Observations	720	747	747	161	166	166
Notes: This table shows the relationship between HIV incidence and exports. Columns 1-3 use data on HIV incidence rate derived from UNAIDS data; Columns 4-6 use estimates based on inference from mortality data (Oster, forthcoming). Countries in each dataset are listed in Appendix Table 1. The export measures are log total exports; incidence rate is on a scale from 1 to 100. All columns include controls for country and year fixed effects. Lagged HIV prevalence is from the appropriate HIV series. Implied values of gamma are calculated using calibration parameters discussed in Section 2. Standard errors in parentheses. Regressions are run adjusting for serial correlation using a Prais-Winsten regression with robust standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.						

**Table 3. Heterogeneity: Variation in Impact Across Areas, Time**

<i>Exp. Measure</i>	<i>UNAIDS Data</i>			<i>Mortality-Based Data</i>		
	<i>WDI, Value</i>	<i>NBER, Value</i>	<i>NBER, Vol.</i>	<i>WDI, Value</i>	<i>NBER, Value</i>	<i>NBER, Vol.</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Level Effects</b>						
<i>Variation to Analysis:</i>						
<b>Baseline</b>	.073*** (.027)	.072*** (.031)	.048** (.025)	.474*** (.176)	.278** (.120)	.091 (.081)
<b>East Africa</b>	.370** (.149)	.406*** (.149)	.394** (.165)			
<b>Southern Africa</b>	.257** (.131)	-.035 (.133)	.115 (.106)			
<b>West Africa</b>	.006 (.012)	-.014 (.011)	-.007 (.010)			
<b>1990 and After</b>	.035 (.029)	.018 (.015)	.032** (.013)	.628*** (.237)	.392** (.156)	.114 (.104)
<b>Panel B: Lagged HIV-Export Interactions</b>						
<i>Variation to Analysis:</i>						
<b>Baseline</b>	.026** (.010)	.026** (.012)	.020** (.008)	.092*** (.025)	.062*** (.017)	.025 (.017)
<b>East Africa</b>	.040** (.017)	.058** (.023)	.063** (.025)			
<b>Southern Africa</b>	.023*** (.007)	.007 (.006)	.013** (.005)			
<b>West Africa</b>	.0001 (.006)	.0004 (.005)	.0004 (.005)			
<b>1990 and After</b>	.015* (.009)	.008 (.007)	.013* (.007)	.086** (.037)	.064*** (.021)	.011 (.021)

Notes: This table explores heterogeneity in the estimates from Table 2 across region and time period. Panel A estimates the level effect of exports. Panel B shows the coefficient on the interaction between exports and lagged HIV prevalence (as in Panel B of Table 2). We report here only the relevant coefficients; other controls include country fixed effects, year fixed effects and lagged HIV prevalence. East Africa includes Burundi, Ethiopia, Eritrea, Kenya, Malawi, Rwanda, Tanzania and Uganda. Southern Africa includes Angola, Botswana, Mozambique, South Africa, Lesotho, Namibia, Swaziland, Zambia and Zimbabwe. The remainder are coded as West Africa. Due to sample size issues, we cannot run the regional analysis using the mortality-based data. Standard errors in parentheses. Regressions are run adjusting for serial correlation using a Prais-Winsten regression with robust standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Table 4. Robustness: Country-Specific Trends, Assumptions Generating Incidence**

<i>Exp. Measure</i>	<i>UNAIDS Data</i>			<i>Mortality-Based Data</i>		
	<i>WDI, Value</i>	<i>NBER, Value</i>	<i>NBER, Vol.</i>	<i>WDI, Value</i>	<i>NBER, Value</i>	<i>NBER, Vol.</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Level Effects</b>						
<i>Variation to Analysis:</i>						
<b>Baseline</b>	.073*** (.027)	.072*** (.031)	.048** (.025)	.474*** (.176)	.278** (.120)	.091 (.081)
<b>Linear Trends</b>	.092** (.040)	.045 (.029)	.034 (.023)	.533*** (.203)	.330*** (.118)	.127 (.080)
<b>Quadratic Trends</b>	.008 (.032)	.030 (.029)	.041* (.023)	.478** (.203)	.353*** (.109)	.147* (.078)
<b>Faster Death</b>	.075*** (.027)	.075*** (.031)	.050** (.024)	.546** (.234)	.388** (.159)	.112 (.097)
<b>Flatter Death</b>	.070*** (.027)	.070** (.030)	.047** (.024)	.510** (.240)	.302** (.146)	.082 (.086)
<b>“Treatment”</b>	.076** (.030)	.058* (.031)	.042* (.025)			
<b>Earlier Start Date</b>	.068*** (.026)	.056** (.026)	.038* (.021)	.530** (.232)	.354** (.159)	.128 (.091)
<b>Later Start Date</b>	.088*** (.029)	.110** (.044)	.071** (.035)	.684** (.278)	.438** (.178)	.129 (.108)
<b>Panel B: Lagged HIV-Export Interactions</b>						
<i>Variation to Analysis:</i>						
<b>Baseline</b>	.026** (.010)	.026** (.012)	.020** (.008)	.092*** (.025)	.062*** (.017)	.025 (.017)
<b>Linear Trends</b>	.033** (.014)	.032** (.014)	.029** (.014)	.095*** (.027)	.068*** (.018)	.026 (.019)
<b>Quadratic Trends</b>	.017 (.014)	.022** (.011)	.024** (.011)	.085*** (.031)	.068*** (.018)	.020 (.021)
<b>Faster Death</b>	.026** (.010)	.026** (.012)	.020** (.009)	.069*** (.021)	.062*** (.014)	.031 (.019)
<b>Flatter Death</b>	.026** (.010)	.026** (.012)	.020** (.008)	.061*** (.020)	.048*** (.012)	.028* (.016)
<b>“Treatment”</b>	.019 (.013)	.020 (.015)	.020* (.011)			
<b>Earlier Start Date</b>	.023** (.009)	.024** (.011)	.019** (.007)	.072*** (.021)	.047*** (.013)	.031* (.018)
<b>Later Start Date</b>	.028*** (.010)	.029** (.012)	.022** (.008)	.073*** (.026)	.065*** (.017)	.032 (.021)

Notes: This table replicates Table 2 using varying assumptions in the analysis. The first two variations show the impact of adding country-specific liner trends and country-specific quadratic trends to the regression. The remainder test alternative assumptions in the generation of HIV incidence. Panel A estimates the level effect of exports. Panel B shows the coefficient on the interaction between exports and lagged HIV prevalence (as in Panel B of Table 2). Other controls include country fixed effects, year fixed effects and lagged HIV prevalence. The assumptions about time to death can be seen visually in Online Appendix Figure 5. “Treatment” incorporates the assumption that deaths are cut in half after 2002. Earlier start date is four years before the primary start date; later start date is four years later. Standard errors in parentheses. Regressions are run adjusting for serial correlation using a Prais-Winsten regression with robust standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Table 5. Income Mechanism: HIV and Exports with GDP Control**

<i>Dependent Variable:</i>	<i>Incidence Rate (0-100)</i> <i>UNAIDS Data</i>			<i>Incidence Rate (0-100)</i> <i>Mortality-Based Data</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
Explanatory Variables:						
Log Export Value (WDI)	.0591* (.035)			.4453** (.181)		
Log Export Value (NBER)		.0628** (.031)			.2629** (.116)	
Log Export Volume (NBER)			.046* (.025)			.0761 (.083)
Log GDP	.055 (.086)	.1014* (.062)	.1258** (.06)	.2625 (.345)	.3668 (.313)	.4446 (.342)
<i>Controls in all Columns: Country and Year Fixed Effects, lagged HIV Prevalence</i>						
<b>% Exp. Effect Explained</b>	<b>19.1%</b>	<b>12.8%</b>	<b>4.1%</b>	<b>6.1%</b>	<b>5.4%</b>	<b>16.4%</b>
Number of Observations	720	747	747	161	166	166

Notes: This table shows the relationship between HIV incidence and exports controlling for GDP. Columns 1-3 use data on HIV incidence rate derived from UNAIDS data; Columns 4-6 use estimates based on inference from mortality-based data (Oster, forthcoming). Countries in each dataset are listed in Appendix Table 1. The export measures are log total exports; incidence rate is on a scale from 1 to 100. All columns include controls for country and year fixed effects. Lagged HIV prevalence is from the appropriate HIV series. The % of the export effect explained is calculated based on the comparison with coefficients in Table 2. Standard errors in parentheses. Regressions are run adjusting for serial correlation using a Prais-Winsten regression with robust standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Table 6. Mechanisms: Truck Imports, Exports and Road Density**

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Truck Imports and Exports</b>						
<i>Dependent Variable:</i>	<i>Log Weight of Trucks Imported</i>					
Explanatory Variables:						
Log Export Value (WDI)	.380*** (.097)			.255** (.106)		
Log Export Value (NBER)		.364*** (.071)			.282*** (.069)	
Log Export Volume (NBER)			.230*** (.074)			.217*** (.075)
Log Import Value (NBER)				.778*** (.132)	.724*** (.106)	.800*** (.105)
<i>Controls in all Columns: Country and Year Fixed Effects</i>						
Number of Observations	449	571	571	449	571	571
<b>Panel B: Exports and HIV By Road Km</b>						
<i>Dependent Variable:</i>	<i>Incidence Rate (0-100) UNAIDS Data</i>			<i>Incidence Rate (0-100) Mortality-Based Data</i>		
Explanatory Variables:						
Log Export Value (WDI)	.073 (.052)			-.758** (.353)		
Log Export Value (NBER)		.096** (.047)			-.169 (.227)	
Log Export Volume (NBER)			.056* (.034)			-.116 (.286)
Value, WDI × Road Km.	.0038 (.0029)			.143*** (.024)		
Value, NBER × Road Km.		.011*** (.004)			.069*** (.019)	
Vol., NBER × Road Km.			.019*** (.006)			.058* (.033)
<i>Controls in all Columns: Country and Year Fixed Effects, lagged HIV Prevalence, country land area interacted with exports</i>						
Number of Observations	698	724	724	161	166	166
Notes: Panel A estimates the relationship between volume of trucks imported to the country and exports. Truck import volume comes from the NBER trade data. Total import value (Columns 4-6) excludes trucks. Panel B estimates the relationship between HIV and exports, controlling for interactions with kilometers of roads in the country. These regressions also include a measure of total land area in the country interacted with exports. In Panel B, the level effect of roads and land area is captured by the country fixed effects. Lagged HIV prevalence is from the appropriate HIV series. Standard errors in parentheses. Regressions are run adjusting for serial correlation using a Prais-Winsten regression with robust standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.						

**Table 7. Mechanisms: HIV and Roads**

<i>Dependent Variable</i>	<i>Log HIV Prevalence</i>		<i>Any HIV in Cluster</i>	
Explanatory Variables:				
Distance to Road (km)	-.025*		-.0021*	
	(.014)		(.001)	
Distance 2km-10km		-.269*		-.021
		(.156)		(.013)
Distance > 10km		-.602**		-.055**
		(.292)		(.026)
<i>Controls in all Columns: Region fixed effects, cluster averages for: age, age-squared, education, working, # of children (total and at home), marriage, durable good ownership, Muslim.</i>				
Number of Observations	3629	3629	3629	3629
Notes: This table uses data from the Demographic and Health Surveys to estimate the relationship between HIV prevalence in a cluster and the distance to road. Data on road locations are taken from GIS maps of Africa; clusters are identified in the survey with their GPS coordinates. All regressions are limited to rural areas. In Columns 2 and 4 the omitted category is that the road is closer than 2km away. Region fixed effects are included with the controls; a region is a sub-area within a country. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.				

## Appendix Figures and Tables

Appendix Table 1. *Summary Statistics on HIV Measures*

Country	UNAIDS Avg. Prev. (0-100)	UNAIDS Avg. Incid. (0-100)	Mortality-Based Avg. Prev. (0-100)	Mortality-Based Avg. Incid. (0-100)
Angola	0.997	0.168		
Benin	0.762	0.118		
Botswana	25.300	2.627		
Burkina Faso	1.868	0.276	3.248 [1985-1999]	0.202
Burundi	3.212	0.404		
Cameroon	3.787	0.577	2.802 [1985-2000]	0.357
Central African Republic	4.428	0.661		
Chad	2.144	0.349		
Congo, Rep.	4.373	0.636		
Cote d'Ivoire	4.241	0.591		
Equatorial Guinea	2.324	0.351		
Eritrea	1.229	0.154		
Ethiopia	1.669	0.246	4.413 [1985-2001]	0.481
Gabon	3.343	0.532		
Ghana	1.350	0.203		
Guinea	0.777	0.135		
Guinea-Bissau	0.981	0.157		
Kenya	6.567	0.958	3.116 [1985-1999]	0.420
Lesotho	23.557	2.592		
Liberia	1.008	0.162		
Malawi	8.790	1.311	7.539 [1985-2000]	0.705
Mali	0.858	0.136	2.283 [1985-1997]	0.147
Mozambique	5.945	1.009	3.848 [1985-1999]	0.439
Namibia	15.129	1.746	3.923 [1985-1996]	0.546
Niger	0.398	0.067		
Nigeria	1.983	0.307		
Rwanda	5.937	0.772		
Senegal	0.305	0.064		
Sierra Leone	0.878	0.146		
South Africa	8.946	1.487		
Swaziland	26.371	2.889		
Tanzania	5.918	0.816	3.185 [1985-2000]	0.320
Togo	2.323	0.351		
Uganda	9.466	1.132	9.947 [1985-1997]	1.232
Zambia	12.625	1.820	9.952 [1985-1997]	1.248
Zimbabwe	18.948	2.514	7.705 [1985-2001]	1.060

Notes: This table reports summary statistics on average prevalence and incidence for each country in the sample, over the period covered in the data used for analysis. All countries have UNAIDS data for 1985 through 2007. The mortality-based data is available for a varying number of years; years for each country are in square brackets.