

NBER WORKING PAPER SERIES

HIV BREAKTHROUGHS AND RISKY SEXUAL BEHAVIOR

Dana Goldman
Darius Lakdawalla
Neeraj Sood

Working Paper 10516
<http://www.nber.org/papers/w10516>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2004

We are grateful to Jean-Jacques Laffont, Will Manning, Vazha Nadareishvili, and participants in the 2004 Annual Health Economics Conference at the University of Alabama-Birmingham, for helpful comments and discussion. Abby Alpert provided exceptional research assistance. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Dana Goldman, Darius Lakdawalla, and Neeraj Sood. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

HIV Breakthroughs and Risky Sexual Behavior
Dana Goldman, Darius Lakdawalla, and Neeraj Sood
NBER Working Paper No. 10516
May 2004, Revised November 2005
JEL No. I1

ABSTRACT

Recent breakthroughs in the treatment of HIV have coincided with an increase in infection rates and an eventual slowing of reductions in HIV mortality. These trends may be causally related, if treatment improves the health and functional status of HIV+ individuals and allows them to engage in more sexual risk-taking. We examine this hypothesis empirically using access to health insurance as an instrument for treatment status. We find that treatment results in more sexual risk-taking by HIV+ adults, and possibly more of other risky behaviors like drug abuse. This relationship implies that breakthroughs in treating an incurable disease like HIV can increase precautionary behavior by the uninfected and thus reduce welfare. We also show that, in the presence of this effect, treatment and prevention are social complements for incurable diseases, even though they are substitutes for curable ones. Finally, there is less under-provision of treatment for an incurable disease than a curable one, because of the negative externalities associated with treating an incurable disease.

Dana Goldman
RAND
1700 Main Street
Santa Monica, CA 90407
and NBER
dgoldman@rand.org

Neeraj Sood
RAND
1700 Main Street
Santa Monica, CA 90407
sood@rand.org

Darius Lakdawalla
RAND
1700 Main Street
Santa Monica, CA 90407
and NBER
darius@rand.org

I Introduction

Over the past decade, HIV in developed countries has been transformed from a death sentence into a chronic, manageable disease. Since 1995, the overall US AIDS death rate has fallen by almost 70 percent. This remarkable turnaround was driven in large part by breakthrough drugs. Perhaps most important was the introduction of highly active antiretroviral therapy (HAART), the currently recommended treatment regimen for HIV [Carpenter et al., 2000].

However, Figure I demonstrates that progress against HIV and AIDS has stalled in the US. AIDS death rates reached a plateau in 1998, at around 5 deaths per 100,000. Even more troubling has been the rise in the HIV infection rate, which climbed by over forty percent from 1998 to 2001.¹ Moreover, the increased rate of infection has occurred over a period during which sexually active Americans have become more *vigilant* about safe sex, rather than more complacent.² Are these trends in HIV infection and treatment breakthroughs linked? And how have infection rates risen in spite of increasing precaution in the population at large?

This paper posits that increases in HIV infection rates might themselves have resulted from improvements in HIV treatment and the accompanying declines in the deadliness of HIV. In particular, we argue that advances in treatment have improved the health of the infected, reinvigorated their sexual activity, and thus rekindled the spread of HIV.³ While risky behavior and sexual promiscuity among the relatively small number of HIV+ people will not show up in nationwide prevention statistics, it is of first-order importance for HIV

¹The increased infection rate is almost certainly not due to better diagnosis, as the rate of HIV testing was stable over this period (based on the General Social Surveys, 1998-2002). Similar growth in incidence after the diffusion of HAART treatment has been observed for gay men in San Francisco [Katz et al., 2002], heterosexuals in France [Gremy and Beltzer, 2004], and gay men in Amsterdam [Dukers et al., 2002].

²From 1998 to 2002, the rate of condom usage among unmarried Americans rose consistently, and by 17% over the entire period; this is statistically different from zero. These statistics are compiled from the 1998 to 2002 General Social Surveys, which asked respondents if they used a condom the last time they had sex. The sample is unmarried adults who had sex within the last year. Similar trends of flat or increasing precaution are apparent along various complementary dimensions, such as the number of different partners, frequency of casual sex, sex outside a committed relationship, and sex with a prostitute. (These results are based on the authors' calculations using the General Social Surveys, which are described later.)

³Both treatment and risk-taking (or lack of preventive) behavior determine rates of infection [Auld, 2003].

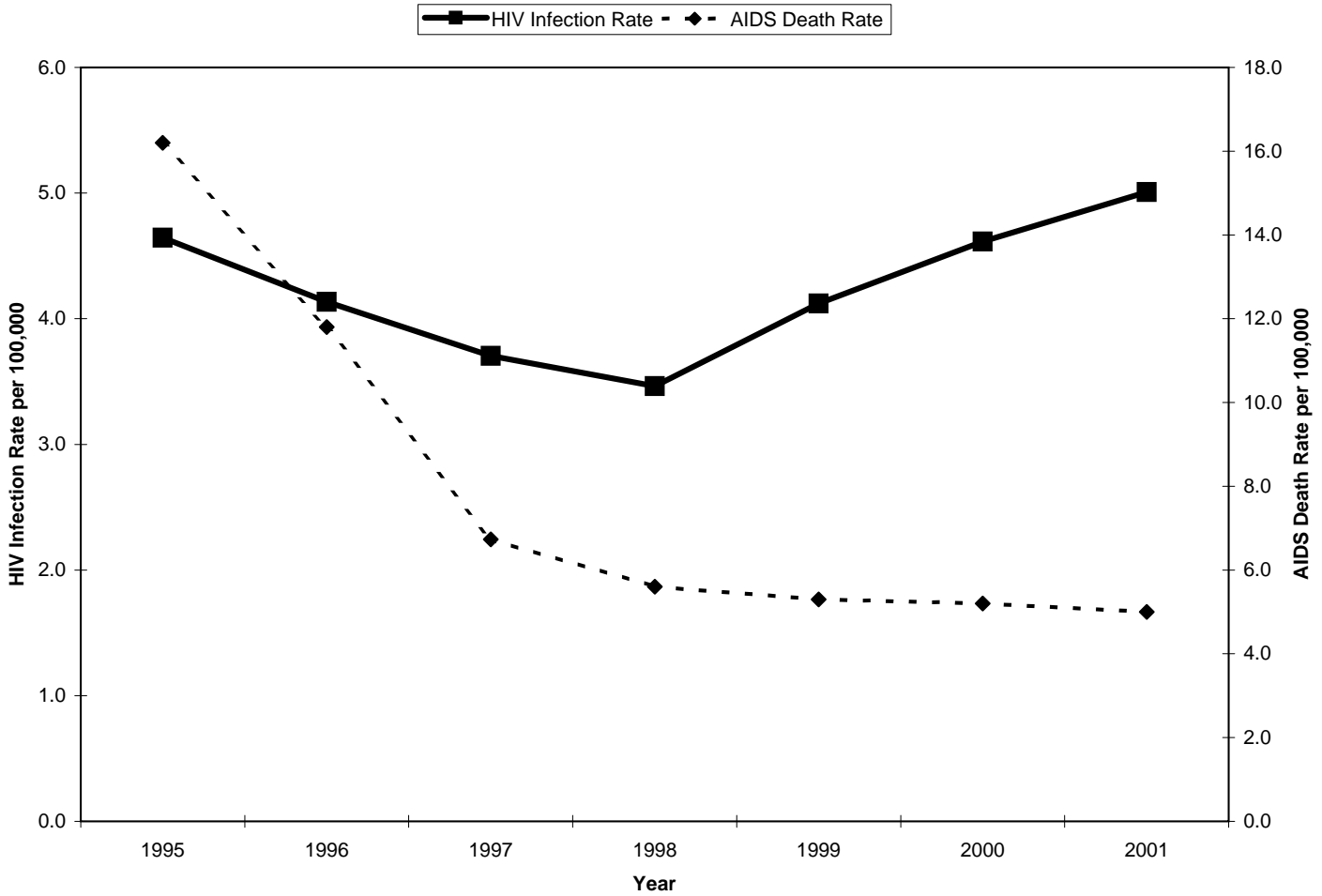


Figure I
 HIV Incidence and Deaths from AIDS in the US.

Note: Data are from CDC HIV/AIDS Surveillance Report 1993-2001 (Year-end editions), and CDC HIV Mortality L285 Slide Series 2000. Also see Centers for Disease Control [1998]. Data on HIV infection rates are from the 26 states that collected confidential data on HIV infection rates throughout this period. The death rate, on the other hand, is based on nationwide numbers.

incidence, because the HIV+ represent society’s vectors of infection.

Clinical trials and observational studies conclusively demonstrate that HAART lowers mortality and improves the physical health of HIV+ persons [Hammer et al., 1997; Staszewski et al., 1999; Jordan et al., 2002; Detels et al., 1998; Palella et al., 1998; Lucas et al., 1999; Vittinghoff et al., 1999; Lucas et al., 2003; Duggan and Evans, 2005]. We argue that these considerable health benefits of HAART have led HIV+ people to increase their sexual activity and spread the disease. Sexual activity by the HIV+ carries a significant risk of transmission, because many display a clear preference for HIV-negative sex partners: 45% of HIV+ individuals actively seek HIV-negative sexual partners at least some of the time, and 20% seek HIV-negative partners all the time.⁴ This may be due in part to a general taste for healthy, high-quality partners, and possibly also to the unique incentive of individuals with impaired immune systems to seek out “clean” partners.

The epidemiological literature has downplayed the causal link between treatment and sexual activity, because simple correlations—both in our data and elsewhere—often find that treatment and sexual activity are unrelated or negatively related [Crepaz et al., 2004; Moatti et al., 2003]. However, these correlational analyses suffer from the problem that patients on HAART are intrinsically much sicker than other patients. This difference in health status masks the causal effect of treatment on sexual behavior, because the poor health of the treated population leads to diminished sexual activity.

To overcome this problem, we exploit state-level variation in the availability of public insurance for HIV-sufferers. Since treatment with HAART is expensive—costing on average about \$13,000 per year—people with insurance are more likely to get HAART. We find that HIV-sufferers who get treatment because they live in states with more generous Medicaid eligibility rules are more likely to engage in sexually risky or promiscuous behavior. According to these estimates, plausibly exogenous treatment with HAART causes an HIV+ individual

⁴These figures are based on the authors’ calculations from the HIV Cost and Services Utilization Study (HCSUS), which is described in Section III.

to have sex with more than twice as many partners in a given 6-month period. A simple theoretical model suggests that these effects of HAART are likely to have increased the risk of infection for the uninfected by 44% or even as much as 100%.

This positive relationship between treatment and risky behavior has several important implications for welfare and behavior. Since HIV is an *incurable* infectious disease, improvements in HIV treatment may not have unambiguous benefits for the uninfected. While the cost of being infected falls, the risk of infection rises with the stock of HIV+ people, who now live longer and more sexually active lives. Breakthroughs can reduce welfare among the uninfected if the cost of higher infection risk more than offsets the welfare gain from the reduced cost of infection. Calibrations of our theoretical model suggests that welfare losses may indeed have occurred among uninfected Americans.

An appropriate policy response to these adverse welfare effects may be an increase in public subsidies for HIV-prevention. Treatment breakthroughs can raise the returns to HIV-prevention by increasing the external costs of risky behavior. This occurs because: first, the increased disease prevalence means that risky behavior is more likely to result in an infection; and second, each new infection is likely to result in more secondary infections due to the increased longevity of the infected. Given these effects, we explore the conditions under which treatment breakthroughs call for greater prevention subsidies.

We begin by laying out the positive implications of HIV treatment breakthroughs for behavior and the prevalence of HIV. Next, we estimate the effect of treatment on sexual activity among the HIV+ and use these estimates to quantify the impact of new treatments on infection risk in Section VI. Finally, in Section VII, we quantify the likely impact on the welfare of the uninfected and then discuss whether public prevention spending can be used to mitigate welfare losses for the uninfected.

II The Positive Implications of HIV Breakthroughs

Treating HIV expands the pool of survivors able to spread the infection, and strengthens their physical ability to engage in transmissive behavior. The theoretically predicted result is an increase in the prevalence of the disease. In response, the precautionary behavior of the uninfected may rise or fall depending on two offsetting effects: the decline in the costliness of the newly treatable disease, versus the increased risk of exposure.⁵

Consider a population of sexually active individuals, with the HIV prevalence π . All individuals value risky sexual activity. For simplicity, suppose that HIV+ people engage in as much risky behavior r^+ as their (exogenous) health allows and receive the corresponding level of per-period utility $w(r^+)$. With probability $1 - s$, they die or become too sick for sex.⁶ The lifetime utility of a sexually active infected individual is $\frac{w(r^+)}{1-s\beta}$, where β is the one-period discount factor.

Uninfected individuals derive concave, increasing utility $u(r)$ from their risky behavior r . The probability of infection ϕ rises in a person's own risky behavior r , the prevalence of HIV π , and the risky behavior of the HIV+ r^+ : $\phi(r; r^+, \pi)$. If r^* represents the equilibrium level of risky behavior among the uninfected and π' is next period's prevalence, the lifetime utility of the uninfected person can be written as:⁷

$$v(\pi) = \max_r \left\{ u(r) + \beta \phi(r; r^+, \pi) \frac{w(r^+)}{1-s\beta} + \beta(1 - \phi(r; r^+, \pi))v(\pi') \right\} \quad (1)$$

$$s.t. \pi' = \frac{\pi s + (1 - \pi)\phi(r^*; r^+, \pi)}{\pi s + (1 - \pi)} \quad (2)$$

⁵The economic theory of AIDS prevention and transmission was pioneered by Philipson and Posner [1993]; Over et al. [2004] explores the specific relationships between the new treatments, prevention investments, and risk-taking.

⁶For our purposes, it is not important to distinguish between these two outcomes, because death and illness are absorbing, zero-utility states.

⁷A single uninfected individual is assumed to have no control over r^* .

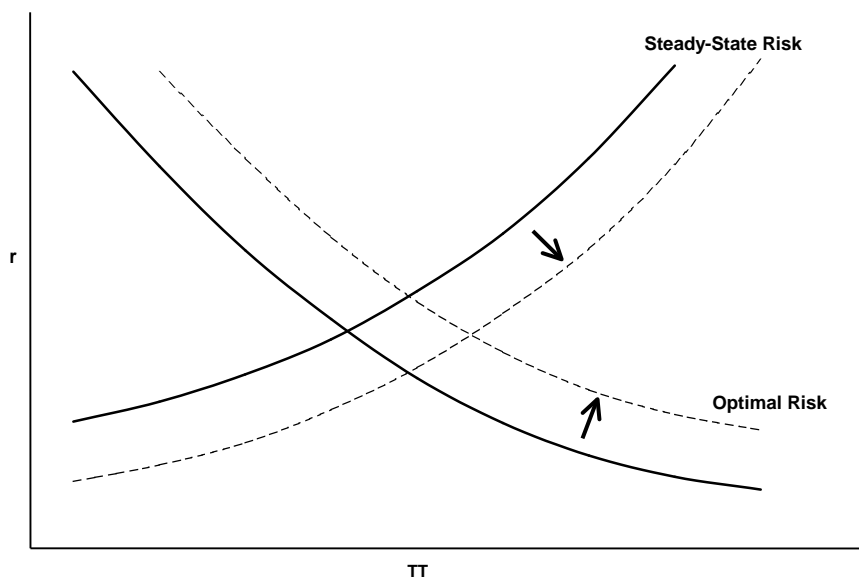


Figure II
Effect of Treatment Breakthroughs on Steady-State Risk-Taking and HIV Prevalence
Among Sexually Active.

The unique pair of steady-state risk-taking and disease prevalence levels satisfies:⁸

$$\frac{du}{dr} - \beta \frac{\partial \phi}{\partial r} \left(v^{ss} - \frac{w(r^+)}{1 - s\beta} \right) = 0 \quad (3)$$

$$\pi(1 - s) - \phi(r; r^+, \pi) = 0 \quad (4)$$

Each equation defines a relationship between r and π , indexed by the two correlates of HIV+ health, s and r^+ .

Figure II demonstrates the intuition behind the steady-state dynamics that result from the pair of equations above. Equation 3 defines the “Optimal Risk” curve, which characterizes the *behavioral* relationship between prevalence and risk-taking: uninfected individuals

⁸The unique steady-state exists under assumptions best-suited to low-prevalence developed countries, as discussed in Appendix V. Kremer [1996] analyzes the more general possibility of multiple steady-states, which are particularly important when people become “fatalistic” about the risk of the disease, so that increased prevalence primarily raises the cost of staying healthy. In the US, where prevalence is relatively low, increased risk appears to lower risk-taking [cf, Ahituv et al., 1996], at least for the average person.

are more cautious when faced with higher disease prevalence. Breakthroughs in treatment shift this curve up by improving the welfare of the HIV+ (i.e., raising (r^+, s)), which makes the uninfected more willing to risk infection at a given level of exposure risk. On the other hand, the “Steady-State Risk” curve, defined by Equation 4, characterizes the steady-state or *epidemiological* relationship between risk-taking and prevalence: more risk-taking results in higher prevalence. Breakthroughs shift this curve down, because a given level of prevalence can be supported by less risk-taking on the part of the uninfected. In other words, increases in survival and sexual activity among the HIV+ multiply the effects of risky behavior among the uninfected, so that less of it is needed to sustain the disease at a given level.

The shifts in these two curves both combine to increase prevalence:⁹ there are more sexually active HIV+ people who can spread infection; this is reinforced, because the uninfected take more risk at any given level of prevalence. However, the effects on precaution are offsetting and ambiguous: the lower cost of infection encourages risk-taking, while higher prevalence discourages it.

This reasoning explains why precaution may have risen, as suggested by Figure I, in spite of the falling cost of HIV-infection. It also illustrates the way in which new treatments can increase risk for the uninfected, and reveals the importance of quantifying the impact of treatment breakthroughs on $\phi(\cdot; r^+, \pi)$, the risk schedule faced by the uninfected. In our empirical analysis, we recover the impact of treatment on r^+ and combine this with the known clinical effects of treatment on survival (s), to arrive at estimated increases in the risk faced by the uninfected. The increased risk is then used in conjunction with the theoretical model to assess whether or not the welfare of the uninfected declined in the US.

⁹In the most general models of rational epidemics, the change in long-run prevalence may be ambiguous. In Appendix V, we provide support for the regularity conditions that lead to this unambiguous result, and we explain why they are appropriate for the recent history of HIV in the US.

III Data

We use data from a nationally representative study of HIV+ patients in care—the HIV Costs and Services Utilization Study (HCSUS). The HCSUS employed a multi-stage national probability sample design to identify HIV+ patients over 18 years old, who made at least one visit for regular care in the contiguous United States in January or February of 1996. It does not include HIV+ patients whose only contact with the health care system was through military, prison, or emergency department facilities, or who have not made contact with the health care system for their HIV. HCSUS collected data between March 1996 and January 1998 — a period when HAART entered clinical practice and disseminated widely. HCSUS is a panel data set with three waves of interviews, which we refer to as “Baseline,” “Follow-Up 1,” and “Follow-Up 2.” Questions about sexual activity were posed to a random sample of 1,794 respondents in Follow-Up 2. 1,421 of these respondents completed the interview (a completion rate of 79%, or 84% after adjusting for mortality).¹⁰ We use this subsample of HCSUS respondents for our analysis of sexual behavior.¹¹

As our outcome measure, we study the number of sex partners an individual reports over the past 6-month period. The key explanatory variable of interest is treatment with HAART. This is derived from HCSUS questions about medications taken by each respondent. HAART is defined (both clinically, and by the survey) as a combination drug therapy involving three types of drugs: nucleoside reverse transcriptase inhibitors, protease inhibitors, and non-nucleoside reverse transcriptase inhibitors. All HAART regimens require at least three drugs, and the vast majority involve at least one protease inhibitor. A respondent is classified by HCSUS as being on HAART if he/she reports taking any of the recognized HAART drug combinations. Table I presents descriptive statistics for all model variables.

¹⁰This sample is representative of the 197,063 HIV+ US adults who received care in 1996 and survived until 1998.

¹¹Of the 1,421 respondents, 1396 answered the question about “number of sex partners in the last 6 months.”

Table I
Weighted descriptive statistics by HAART status (N=1396).

Variable	No HAART ^a (N=573)	HAART ^a (N=823)	Overall (N=1396)
Demographics			
Age (years)	39	39	39
Non-white (%)	65	56	60
Female (%)	33	25	28
Education			
Less than HS degree (%)	31	25	27
High school degree (%)	31	29	30
Some college or AA Degree (%)	28	27	27
College Degree (%)	11	19	16
Lowest ever CD4 count (cells/μl)^b			
>500 (%)	11	2	6
200-499 (%)	46	33	39
50-200 (%)	26	37	33
0-50 (%)	17	27	23
AIDS (%) ^c	33	46	41
State instruments			
Medically needy threshold (% of FPL)	47	49	49
SSI threshold >65% of FPL (% respondents)	93	91	92
Number of partners in the last 12 months			
0-1 partner (%)	73	74	74
2 partners (%)	9	8	8
3 or more partners (%)	19	18	18
Number of new partners in the last 12 months			
No partners (%)	68	70	69
1 partner (%)	15	12	13
2 or more partners (%)	17	18	18

Notes: State instruments were obtained from Westmoreland (1999). All other variables come from HCSUS Follow-Up 2. All data are directly self-reported, except where noted.

^aHAART is defined as a combination drug therapy involving three types of drugs: nucleoside reverse transcriptase inhibitors, protease inhibitors, and non-nucleoside reverse transcriptase inhibitors. All regimens require at least three drugs, and the vast majority involve at least one protease inhibitor. A respondent is classified by the survey as being on HAART if he/she reported taking any one of the HAART combinations.

^bCD4 count, which is self-reported, is used to gauge the health of an HIV+ person. A higher count indicates better health.

^cThe presence of AIDS is based on self-reported CD4 counts, and self-reported opportunistic infections. The HCSUS imputes the presence of AIDS using these data, according to CDC algorithms.

To characterize the health of treated and untreated individuals, we report both the prevalence of AIDS,¹² and the distribution of lowest-ever CD4 counts. The CD4+ t-lymphocyte cell count is a critical measure of the function of a patient’s immune system. A depletion in these cells correlates strongly with the worsening of HIV disease and physical health [Fauci et al., 1998]. Lowest-ever CD4 count measures severity by gauging the farthest progression of the disease. Table I splits CD4+ counts into four categories. Patients with CD4+ lymphocyte counts below 50 have a very poor prognosis in general; while those with counts above 500 are considered reasonably healthy.

Table I corroborates the findings of other researchers in showing that, unconditionally, the treated are no more promiscuous than the untreated. However, it also demonstrates the difficulty of drawing causal inferences from this fact alone. Treated patients suffer from more advanced disease, as they are more likely to have CD4 counts that once dropped below 50, or below 200, and they are also more likely to have AIDS.

IV Empirical Model

Table I illustrates how simple comparisons of sexual activity by treatment status can be misleading. As a result, we develop an identification strategy that isolates quasi-random variation in the provision of treatment. The strategy derives its power from HAART’s high cost and from its extraordinary effectiveness. In 1998, the average HAART regimen cost Medicaid about \$13,000, but HAART cuts mortality (in clinical trials) by about 50%. Since only one-quarter of the HIV+ have private insurance,¹³ obtaining health insurance through Medicaid can mean the difference between receiving or not receiving HAART, and in turn the difference between good health and rapid physical deterioration. This suggests

¹²The occurrence of AIDS is derived from the individual’s self-reported CD4 counts, and the self-reported occurrence of opportunistic infections. The HCSUS imputes the presence of AIDS according to these data, using CDC algorithms.

¹³In HCSUS Follow-up 2, 25% have private insurance, 17% are uninsured, and 58% have public insurance. 51% of respondents have some Medicaid coverage.

the possibility of using Medicaid eligibility rules as instruments for HAART receipt that are arguably unrelated to other determinants of sexual activity.

Previous research has already documented the strong effect of public insurance on the receipt of HAART and correspondingly on health and survival. Bhattacharya et al. [2003] use Medicaid eligibility rules (the same ones proposed as instruments here) to instrument for public insurance, as 90% of the publicly insured HIV+ have Medicaid coverage. They find that Medicaid eligibility rules are extremely strong predictors of public insurance (Wald Test = 38) with the medically needy threshold having the highest predictive power. Their results show that the receipt of public insurance reduces 1-year mortality rates by 66% [also cf, Goldman et al., 2001]. They also find that controlling for receipt of HAART virtually wipes out the mortality gradient across the publicly insured and uninsured. These results suggest that access to HAART may explain a substantial proportion of the mortality benefits of public insurance for the HIV+, if not all of it.

A Treatment Effects Model of HAART and Sexual Behavior

The first step in developing our identification strategy is to build a model that allows for the joint determination of treatment status and sexual activity. A simple way to do this is to use a standard “treatment effects” model with instrumental variables. Specifically, we use a two-equation model where the first-stage models the binary receipt of HAART treatment as a probit, and the second-stage is a linear model for the number of sexual partners. As instruments for the receipt of treatment, we rely on state Medicaid policy variables that, we argue, affect the availability of treatment, but not sexual activity.

Let T_i^* represent the latent index function that measures the treatment propensity for HIV+ patient i .

$$T_i^* = \beta_1 X_i + \beta_2 Z_i - \epsilon_{T,i} \tag{5}$$

The vector X_i represents observed exogenous covariates that determine treatment propensity:

age, gender, education, along with state-level social and economic factors. (The full list of variables appears in Section V.) Z_i is our vector of state Medicaid eligibility instruments. Treatment is also assumed to depend on a random error component $\epsilon_{T,i}$ that is uncorrelated with X_i and Z_i . Define T_i as the indicator variable for whether individual i actually received HAART; it equals unity if and only if the latent index T_i^* exceeds zero.

Let S_i represent the number of sex partners for HIV+ patient i :

$$S_i = \zeta_1 T_i + \zeta_2 X_i - \epsilon_{S,i} \tag{6}$$

For simplicity, we assume S_i depends linearly on X_i . To complete the model and allow for correlation between treatment and sexual activity, we assume the errors $\epsilon_{S,i}$ and $\epsilon_{T,i}$ are jointly distributed as bivariate normal with correlation coefficient ρ . It is useful to think of this correlation ρ as unobserved health. That is, patients with poor unobserved health are more likely to get treatment and they are also less likely to be sexually active. We estimate this joint model via maximum likelihood.¹⁴

B A Count Data Model of Sex Partners

While the simple treatment effects approach is standard, its limitation is the modeling of sex partners as a continuously distributed variable, instead of an integer-valued count. Therefore, we present an alternative “count data” approach where the number of partners follows a negative binomial distribution, but HAART receipt continues to be modeled simply as a probit [see Deb and Trivedi, 2004, for a detailed exposition of the model].

Suppose there is some common unobserved component in the treatment and sex partner equation. Concretely, one can think of this as health status. This component, η , is assumed to be distributed as a standard normal random variable, so that the latent treatment equation

¹⁴Maddala [1986] derives the maximum likelihood estimator. Computationally, we use the standard “treatreg” command (with weights and robust standard errors clustered at the state level) in Stata version 8.2.

reads as:

$$T_i^* = \delta_0 + \delta_1 X_i + \delta_2 Z_i - \nu_{T,i} - \eta_i \quad (7)$$

We assume that $\nu_{T,i}$ and η_i are distributed standard normal, and that T_i continues to follow the latent index T_i^* . Therefore, conditional on η , T_i follows a probit as before. Effectively, unobserved heterogeneity in the treatment equation is decomposed into one component that is not correlated with sexual behavior ($\nu_{T,i}$) and another one that is (η_i).

The number of sex partners is distributed as a Poisson process with mean/variance parameter $\exp(\theta_0 + \theta_1 X_i + \theta_2 T_i - \lambda \eta_i - \nu_{S,i})$. Just as in the treatment equation, heterogeneity in the number of sex partners is decomposed into a component that is uncorrelated with treatment, $\nu_{S,i}$, and one that is, η_i . Notice that λ is the covariance between the correlated errors in the two equations. If $\nu_{S,i}$ is assumed to follow a $\Gamma(\alpha)$ distribution, S_i is distributed as a negative binomial with mean $\mu \equiv \exp(\theta_0 + \theta_1 X_i + \theta_2 T_i - \lambda \eta_i - \nu_{S,i})$. We estimate the parameters of this model via Maximum Likelihood; details of the likelihood function are given in Appendix IV.

C Identification

We use state Medicaid policies as the instrumental variables Z_i assumed to affect treatment status but not sexual activity directly. Medicaid is the most common form of insurance for the HIV+ population in care, covering 51% of the HIV+ and 61% of the insured HIV+. HIV+ patients can qualify for Medicaid through at least two pathways.¹⁵ First, Supplemental Security Income (SSI) beneficiaries are automatically eligible for Medicaid in 38 states. The other states have different standards for eligibility either as a 209(b) state or a waiver state. Section 209(b) of the Social Security Amendments Act of 1972 allows States to include more restrictive definitions of “disability” and lower income and assets standards for Medicaid

¹⁵The third pathway is Aid to Families with Dependent Children (AFDC): patients who meet the state’s income eligibility and family composition requirements for AFDC as they existed on July 16, 1996 qualify for Medicaid coverage. We do not use AFDC eligibility as an instrument, since this affects incentives for family formation and possibly sexual activity as well.

eligibility. Second, Medicaid eligibility is also available through a “medically needy” program for individuals who meet Medicaid’s disability criteria but have incomes that exceed the financial eligibility limit. The program allows individuals to “spend-down” to Medicaid eligibility by deducting medical-related expenses from their reported income. States have the option but are not required to establish a medically needy program. In addition, states vary in their income eligibility levels for the medically needy program.

As our instruments, we use the following two variables:

- “Medically Needy Threshold” is the state’s income eligibility threshold for the medically needy program expressed as a percentage of the federal poverty line.¹⁶
- “*SSI > 65% FPL*” is an indicator variable for whether the state’s income eligibility threshold for Medicaid eligibility through the “SSI” category was less than 10 percentage points lower than the federal guideline of 75 percent of the federal poverty line.¹⁷

Our state policy instruments could fail if they are correlated with unobserved state-level determinants of sexual activity.¹⁸ To address this issue, we develop evidence that our instruments are related to state-specific differences in sexual activity only through their effects on HIV treatment receipt, and not otherwise.

Valid instruments would affect sexual activity only through their impact on HAART. Before the advent of HAART around 1996, there should have been no relationship between

¹⁶States which did not institute a medically needy program were coded as having an income threshold of zero percent. Specifying the state’s medically needy threshold as a continuous variable was a better predictor of treatment (in the sense of a higher partial chi-squared statistic) than a dummy variable for whether states have a medically-needy program or various alternative nonlinear specifications. The main results for the effect of HAART on sexual activity are highly robust across these different specifications.

¹⁷We coded the SSI threshold as a dummy variable, because there were only four states that implemented a significantly more restrictive eligibility standard than the federal guideline of 75% of the federal poverty line.

¹⁸If respondents with more severe disease moved to states with more generous Medicaid policies our instruments would be correlated with unobserved severity of disease. However, less than 3% of the HCSUS sample migrated across states between Baseline and the Second Follow-Up interviews, despite dramatic improvement in HIV treatment during this period.

Medicaid eligibility policies and sexual activity. A nonzero pre-HAART relationship would suggest the presence of unobserved state-specific determinants of sexual activity, correlated with Medicaid policies. After HAART, generous Medicaid rules would increase the survival of HIV+ patients; according to theory, this would promote precaution among the uninfected.¹⁹ If the instruments are valid, a negative relationship between Medicaid generosity and sexual activity may emerge after HAART, but none should be present before HAART.

Motivated by these observations, we examine the relationship between 1996 state Medicaid rules and sexual activity among the *uninfected*, both before and after the advent of HAART. It is particularly useful to investigate uninfected populations with similar risk characteristics to the HIV+. We do this using high HIV risk subsamples from the General Social Surveys (GSS).

The GSS collect data from a nationally representative sample of individuals on social attitudes, beliefs, and behaviors, including sexual behavior. These include questions about lifetime sexual history: number of partners since age 18, by gender; and ever having paid for or having been paid for sex. Respondents are also asked more specific questions about their sexual experiences during the last twelve months: the number and gender of sex partners; and the type of sex partner, e.g., prostitute, acquaintance, neighbor, friend, stranger, spouse, long-term lover, and so forth. From these questions, we measure the recent and long-term riskiness of each respondent's behavior.

For the pre-HAART period, we use all GSS survey years from 1988 to 1994;²⁰ these include 1988-1991 annually, 1992, and 1994. For the post-HAART period, we use all years from 1998 to the present; these include 1998, 2000, and 2002. These choices end up providing about as much power in the pre-HAART period as post-HAART, a desirable property since we are testing for a zero effect pre-HAART and nonzero post-HAART.

¹⁹There may be an offsetting positive relationship if uninfected people in more generous states face a lower cost of disease, but this presumes uninfected people heavily discount the possibility of cross-state migration before or at the future date of infection.

²⁰The sexual behavior questions were redesigned beginning in 1988.

We construct high-risk samples from the GSS in three ways. Our simplest approach is to isolate males who report ever having had sex with someone previously unknown to them, i.e., respondents who report: (1) Sex with a prostitute in the last 12 months; or (2) Sex with a casual date or “pick-up” in the last 12 months; or (3) Ever having paid for or ever having been paid for sex.²¹ We pick males here, because the HIV+ population is three-quarters male. We call this the “Risky Males” sample.

While the simplicity of the above approach is attractive, there are differences in observable characteristics (race, age, gender, and sexual orientation) between risky males and the HIV+. To complement it, therefore, we re-weight the sample of all risky (male and female) respondents to match the age, gender, race, and sexual orientation composition of the HIV+ population.²² Specifically, we re-weight the sample so that it matches HCSUS proportions in terms of: age category (18-29, 30-39, 40-49, and 50+), race (white and non-white), gender, and sexual orientation (straight male or non-straight male).²³ Henceforth, this is the “Weighted Risky” sample.

Finally, we also re-weight the entire universe of adult GSS respondents, according to the characteristics above. We call this the “Weighted Adults” sample. This provides us with a larger, more powerful sample that still matches the HCSUS on key observables.

Summary statistics for the three GSS subsamples are shown in Table II. By construction, the age, race, and gender composition of the re-weighted samples is the same as the HIV+ population in HCSUS Follow-Up 2. Risky males are older and more likely to be white. All groups are somewhat more educated than the HIV+.²⁴

Comparing this table to the characteristics of the HIV+ population in Table I suggests that the sexual activity of the HIV+ is likely to be similar. Direct comparison is difficult,

²¹The latter question is asked from 1991 onwards.

²²Unfortunately, more detailed weighting is not possible (e.g., adding education groups), because empty cells emerge in the GSS among young, gay black males.

²³Since very few (less than 2%) HIV+ females report being gay or bi-sexual, we do not explicitly consider the proportion of non-straight females.

²⁴This has ambiguous and likely small impacts on the sexual risk-taking of these groups, as education has ambiguous and nonmonotonic effects on the number of sexual partners reported in the GSS.

Table II
Descriptive statistics for high-risk and low-risk uninfected GSS subsamples.

Variable	Weighted Risky ^a	Risky Males ^b (N=1,056)	Weighted Adults ^c (N=11,817)
Demographics			
Age (years)	39	43	39
Non-white (%)	60	19	60
Female (%)	28	0	28
Years of Schooling			
Less than 12th Grade (%)	19	16	18
12th Grade (%)	17	29	33
1-3 years of College (%)	38	28	31
4+ years of College (%)	26	27	25
Number of Partners in Last 12 months			
0-1 partner (%)	35	50	74
2 partners (%)	26	14	11
3 or more partners (%)	39	36	15

Notes: Sample is GSS years 1988-91 annually, 1992, 1994, 1998, 2000, and 2002.

^a"Weighted Risky" sample consists of risky males and females (with "risky" defined as in note b) reweighted to match HCSUS Follow-Up 2 composition in terms of age (18-29, 30-39, 40-49, and 50+), race (white or non-white), gender, and sexual orientation (straight male or non-straight male).

^bHigh-risk males are GSS (years given above) respondents who reported: (1) Sex with a prostitute in the last 12 months, (2) Sex with a casual date or "pick-up" in the last 12 months, or (3) Ever having paid or ever having been paid for sex.

^cHIV-weighted sample is the entire GSS sample reweighted to match HCSUS Follow-up 2 composition in terms of age category (18-29, 30-39, 40-49, and 50+), race (white or non-white), gender, and male sexual orientation (straight male or non-straight male).

because the HCSUS asks respondents to report partners over the last 6 months, while the GSS asks about the last year. However, mean partners among the HIV+ seem comparable, particularly to the weighted risky sample. When we consistently topcode the number of partners at 5 in both data sets, mean partners are 1.36 in 6 months for HCSUS, and 2.46 in 12 months for the weighted risky GSS sample.²⁵ We can overstate the true 12-month partners mean for the HIV+ by doubling the 6-month mean, to obtain 2.72. This is an overstatement, because it topcodes the variable at 10 rather than 5, and because some people will remain monogamous or not acquire additional partners in the extra 6 months. In spite of these factors, it is within 10% of the mean for the weighted risky sample. As discussed in Appendix I, the *distribution* of partners for the HIV+ likely lies somewhere between our three subsamples.

Table III presents the effect of Medicaid eligibility rules on the sexual activity of high-risk uninfected groups, pre- and post-HAART. For each sample and subperiod, we run an OLS regression and a negative binomial regression²⁶ of sexual activity on: 1996 Medically Needy Thresholds; whether the 1996 SSI threshold was above or below 65% of the poverty line; year dummies; gender; black; annual income decile; age category (18-24, 25-29, 30-34, 35-39, 40-44, 45-54, 55-64, 65+); and years of schooling group (less than 12th grade, 12th grade, 1-3 years of college, or 4+ years of college). The number of partners is topcoded at 5 (less than 2% of the sample is topcoded). Due to the three subsamples, two subperiods, and two models, the table reports the results of 12 different models. To conserve space, we report only the coefficients on the state policy variables. All standard errors are clustered by state.

Without exception, there is no statistically significant relationship between 1996 state policies and sexual activity pre-HAART. The majority of the point estimates lie at or below their standard errors. Post-HAART, on the other hand, more generous medically needy thresholds are always associated with less sexual activity in these high-risk groups. The

²⁵The means are 2.14 for unweighted risky males, and 1.42 for the full re-weighted GSS sample.

²⁶The negative binomial coefficients can be interpreted as percent changes in the number of partners.

Table III
State policy and the sexual activity of high-risk uninfected groups, before and after the
advent of HAART.

	Weighted Risky ^a		Risky Males ^b		Weighted Adults ^c	
	Pre- ^d	Post- ^e	Pre- ^d	Post- ^e	Pre- ^d	Post- ^e
<u>OLS on Number of Partners^f</u>						
Med. Needy Thresh.	-0.0033 (0.0046)	-0.0135*** (0.0035)	-0.0022 (0.0023)	-0.0042** (0.0018)	-0.0012 (0.0016)	-0.0065*** (0.0023)
SSI Thresh. >65% FPL	-0.3564 (0.3002)	-0.2521 (0.2428)	-0.0204 (0.0908)	-0.0473 (0.1247)	0.0688 (0.2020)	-0.1632 (0.1954)
R-Squared	0.37	0.29	0.40	0.29	0.12	0.14
<u>Negative Binomial on Number of Partners^f</u>						
Med. Needy Thresh.	-0.0010 (0.0019)	-0.0058*** (0.0013)	-0.0011 (0.0010)	-0.0020** (0.0008)	-0.0009 (0.0011)	-0.0043*** (0.0014)
SSI Thresh. >65% FPL	-0.1550 (0.1219)	-0.1124 (0.0933)	-0.0104 (0.0379)	-0.0279 (0.0570)	0.0490 (0.1519)	-0.1056 (0.1149)
Sample Sizes	506	640	497	559	6338	5479

Robust standard errors, clustered by state, appear in parentheses, underneath coefficients.

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

Notes: All models include dummies for survey year, gender, black, (annual) income decile, age (18-24, 25-29, 30-34, 35-39, 40-44, 45-54, 55-64, 65+), and years of schooling group (less than 12, exactly 12, 13-16, and 16+).

^a"Weighted Risky" sample consists of risky males and females (with "risky" defined as in note b) reweighted to match HCSUS Follow-Up 2 composition in terms of age (18-29, 30-39, 40-49, and 50+), race (white or non-white), gender, and sexual orientation (straight male or non-straight male).

^b"Risky Males" are defined as GSS males who reported: (1) Sex with a prostitute in the last 12 months, (2) Sex with a casual date or "pick-up" in the last 12 months, or (3) Ever having paid for or having been paid for sex (available from 1991 onwards).

^c"HIV-weighted adults" sample reweights GSS so that it matches the composition of the HIV+ population in terms of: age categories (18-29, 30-39, 40-49, 50+), race (white and non-white), and gender risk group (gay male, straight male, or female). HIV+ population proportions are estimated from the HCSUS Follow-Up 2 sample.

^dDefined as GSS survey years 1988, 1989, 1990, and 1991.

^eDefined as GSS survey years 1998, 2000, and 2002.

^fNumber of partners in the past 12 months, which we topcode at 5. Above this level (1.92% of the sample), the GSS reports partners categorically not continuously.

insignificant coefficients on the SSI variable are consistent with our later finding that the medically needy threshold is by far the more individually powerful predictor of HAART receipt.²⁷

Moreover, with the one exception of the risky males subsample, the significant post-HAART effects survive (at the 1% level) the full battery of region dummies and state-specific controls we use in our later HCSUS analysis.²⁸ The pre- and post-HAART effects of the medically needy threshold are statistically different (at 5%) across the board for the weighted adults sample, and statistically different across the board (at 10%) for the weighted risky sample. We cannot statistically reject equality across periods for the smaller risky male subsample.

While one can in principle always find a statistically insignificant relationship in a small enough sample, it is important to note that post-HAART samples of comparable size and power²⁹ consistently yield significant relationships between the medically needy threshold and the number of sex partners among high-risk subpopulations. This is true regardless of the model or sample used, and these relationships are fairly robust to a range of state- and region-level controls. If a correlation between policies and unobserved state-level sexual preferences existed, it would likely appear in both the pre- and post-HAART periods, but a causal effect appears only post-HAART. Therefore, our findings are consistent with the contention that state Medicaid policies have causal effects rather than correlations with unobserved determinants of sexual activity. Moreover, even a less generous interpretation of our results provides some support. If the post-HAART relationship is somehow an artifact

²⁷However, adding the SSI threshold as an instrument significantly increases the joint explanatory power of the two instruments together.

²⁸These include region dummies (Northeast, Midwest, South, and West), state per capita income, state-level frequency of church attendance, state level attitudes about whether homosexuality is wrong, percent living in urban areas, and abortion rates. Details regarding these variables can be found in Section V and Appendix II.

²⁹In all 6 specifications, the pre-HAART data would have had enough power to identify the estimated post-HAART effect: Of the 6 post-HAART coefficients on the Medically Needy Threshold, 5 would be significantly different from zero at 95% using the standard error in the corresponding pre-HAART model, and the last would pass a 90% test. Conversely, in all 6 specifications, the post-HAART data would have failed to reject equality with zero (at 10%) for a coefficient of the size estimated in the corresponding pre-HAART data.

of unobserved state-level heterogeneity, it biases us against finding an effect of HAART on sexual activity, since more generous states exhibit less sexual activity.

As a final check on the plausibility of our results, we compared our estimates for risky groups to estimates for the population at-large, where one would expect much smaller post-HAART effects. We do in fact find little to no post-HAART effect in the sample of all adult males (as opposed to high-risk males). OLS regressions for all adult males produces a pre-period coefficient on the medically needy threshold of -0.0004 with a standard error of 0.0005, and a post-period coefficient of -0.0013 with a standard error of 0.0010. The coefficients on the SSI threshold were likewise insignificant in the pre- and post-periods (sample sizes are 3096 and 2554, respectively).

V Results

A HAART and the Number of Partners

Table IV reports the results from the treatment effects model where HAART receipt is modeled as a probit and number of partners is modeled linearly. In this table and throughout the balance of this section, we investigate the impact of HAART treatment on the sexual activity of the HIV+. To illustrate the consequences of selection bias, we first estimate a naïve model (left-hand column) where we do not include our state policy instruments and treat HAART as an exogenous variable. We then estimate instrumental variables models with progressively larger numbers of controls for state-specific heterogeneity.

Comparing the naïve model to our core instrumental variables specification (IV Model 1) again suggests that unobserved health generates a spurious negative correlation between HAART treatment and sexual activity. The naïve model associates HAART with a statistically significant reduction in sexual activity. Instrumenting for HAART, however, reveals a statistically significant increase in sexual activity for the treated population. A treated

Table IV
Effect of HAART on the number of sex partners for the HIV+, estimated from joint treatment effects model.

	Naive Model		IV Model 1		IV Model 2		IV Model 3	
	Partners	HAART	Partners	HAART	Partners	HAART	Partners	HAART
Age	-0.080*** [0.016]	-0.003 [0.006]	-0.076*** [0.009]	-0.002 [0.005]	-0.078*** [0.011]	-0.001 [0.006]	-0.076*** [0.011]	-0.002 [0.006]
Non-white	-0.543** [0.227]	-0.229** [0.107]	0.117 [0.278]	-0.228* [0.124]	-0.056 [0.212]	-0.175 [0.111]	-0.075 [0.212]	-0.167 [0.110]
Female	-1.743*** [0.429]	-0.164** [0.072]	-1.064*** [0.180]	-0.140** [0.065]	-0.980*** [0.204]	-0.151** [0.070]	-0.946*** [0.208]	-0.168** [0.069]
Less than HS degree	-2.548*** [0.527]	-0.310** [0.132]	-1.509*** [0.432]	-0.319** [0.145]	-1.364*** [0.421]	-0.353** [0.145]	-1.200*** [0.413]	-0.392*** [0.145]
High school degree	-2.825*** [0.579]	-0.262** [0.107]	-1.866*** [0.500]	-0.262** [0.117]	-1.803*** [0.504]	-0.268** [0.108]	-1.719*** [0.511]	-0.291*** [0.109]
Some college or AA degree	-1.965*** [0.341]	-0.228** [0.098]	-1.010*** [0.291]	-0.271** [0.107]	-0.989*** [0.294]	-0.283*** [0.104]	-0.924*** [0.288]	-0.303*** [0.104]
HAART	-5.290*** [0.624]		4.384*** [0.962]		4.237*** [0.998]		4.184*** [1.038]	
Instruments ^a								
Medically needy threshold				0.003** [0.001]		0.004 [0.003]		0.002 [0.003]
SSI threshold >65% of FPL				0.082 [0.076]		0.274 [0.176]		0.197 [0.179]
Chi-2 Test for Joint Significance of Instruments				11.91		5.79		1.75
p-value of Chi-SQ test				0.003		0.0553		0.417
Correlation (Rho) ^b	0.14***		-0.13***		-0.13***		-0.13***	
State Controls ^c	NO		NO		YES		YES	
Region Fixed Effects	NO		NO		NO		YES	
Observations	1396	1396	1396	1396	1396	1396	1396	1396

Robust standard errors clustered by state appear in brackets underneath coefficients.

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

All models were estimated using "treatreg" (with weights and state clusters) in STATA version 8.2.

^a State instruments were obtained from Westmoreland (1999)

^b Correlation coefficient between error terms: a negative sign means someone with a positive shock to treatment probability has a negative shock to his number of partners.

^c State controls include state per capita income in 1998, percent living in urban areas, abortion rate, proportion of GSS respondents (1988-2002) in state who think sexual relations between persons of the same sex are always wrong or almost always wrong, proportion of GSS respondents (1988-2002) in state who pray several times a week. Data on state per capita income in 1998 are based on data from Bureau of Economic Analysis available online at <http://www.bea.gov/bea/regional/spi/>. Urbanization data are based on the 2000 Census as reported in the *Statistical Abstract of the United States: 2004-05*. Data on abortion rates (abortions per 1000 females aged 15-44) in 1999 are based on the survey by The Alan Guttmacher Institute (Finer and Henshaw, 2003), and population estimates from the US Census Bureau. We thank Ted Joyce for the abortion rates data.

individual has 4.4 additional partners, which corresponds to a more than 100% increase. The instrumented model provides further evidence of the role played by unobserved health, since it estimates a significantly negative correlation ($\rho = -0.13$) between unexplained treatment and unexplained sex partners. There is some unobserved factor (e.g., poor health) that raises the probability of treatment but reduces the number of sex partners.

Jointly, the instruments produce a Chi-2 statistic near 12 ($p = 0.003$). The medically needy threshold is the more powerful instrument,³⁰ but the addition of the SSI threshold improves joint significance. Not surprisingly, the first-stage power of the instruments declines as we add more state-level controls: IV Model 2 adds 5 additional state-level variables, and IV Model 3 adds 3 more. At the end of this subsection, we present a more detailed discussion of first-stage power and show that we have enough power in all specifications to make meaningful statistical inferences.

IV Models 2 and 3 check the robustness of our results to the inclusion, in both the first- and second-stage equations, of state controls and region dummies. In IV model 2, we add: 1998 state per capita income, from Bureau of Economic Analysis Data; the 1999 state-specific abortion rate, from a survey by the Alan Guttmacher Institute;³¹ the percent of the state population living in urban areas, based on the 2000 Census and as reported in *The Statistical Abstract of the United States: 2004-05*; and measures of state-specific church attendance and tolerance of homosexual activity, both of which are estimated from 1988-2002 GSS respondents. Further details regarding variable construction are found in Appendix III. The results from IV Model 2, with state controls, are extremely similar to the results from IV Model 1. The coefficient on HAART falls to 4.24 from 4.38; this decline represents less than one-sixth of a standard deviation.

In IV Model 3, we add region dummies to all the state-level controls mentioned above. If our instruments are merely picking up regional differences in sexual activity (e.g., the South

³⁰This is likely to be because only four states had SSI thresholds below the federal maximum of 75% of the poverty line.

³¹We thank Ted Joyce for providing us with the abortion data.

versus the West), the estimated effect of HAART ought to be significantly affected by this inclusion. It is encouraging that the estimated HAART effects are robust to this inclusion, and are quantitatively similar to the results of the earlier models. The effect of HAART remains highly significant with all these additional controls.³²

As discussed earlier, an alternative approach that accounts for the integer-valued nature of the sex partner distribution is negative binomial regression. Table V displays the coefficients and standard errors estimated from the negative binomial sexual partner equation, estimated jointly with a probit for HAART receipt. The first column shows the naïve model estimated without instruments. As before, HAART is naïvely associated with a significant reduction in the number of partners. Instrumenting for HAART once again flips the sign of this effect. Since negative binomial coefficients have the natural interpretation of percentage changes, IV Model 1 implies that HAART increases the number of partners by 134%. This is similar (though slightly larger) in magnitude to the treatment effects estimates, since the mean number of partners is approximately 3.8. The estimated effect is robust to the inclusion of region dummies and state controls in IV Model 2, where the estimated effect is 133%.

We now turn to the instruments' strength. In our core treatment effects model (IV Model 1), the first-stage Chi-2 statistic exceeds the Staiger and Stock "rule of thumb" for strong instruments, namely a test statistic of 10.0 or higher. However, the core negative binomial model shows a Chi-2 statistic of 7.62, somewhat below the recommendation of 10. In spite of this, a review of the literature on weak instruments, along with two Monte Carlo experiments of our own, suggests that our instruments are strong enough to deliver valid statistical inferences, not just in the core models, but even in the most expansive specifications with all 5 state-level controls and 3 region dummies.

Staiger and Stock [1997] observe that weak instruments create two problems: they bias

³²Separately, we also studied the impact of HAART on reported safe sex practices. We found that HAART provision (instrumented by Medicaid eligibility rules) had no significant effect on the frequency with which individuals practiced safe sex. This implies that treatment increases the number of partners but does not induce more safe sex practices that might offset some of the resulting increase in the risk of transmission.

Table V
Effect of HAART on the number of sex partners for the HIV+, estimated from joint negative binomial count data model.

	Naïve Model		IV Model 1		IV Model 2	
	Partners	HAART	Partners	HAART	Partners	HAART
Age	-0.048*** [0.007]	-0.002 [0.007]	-0.048*** [0.005]	-0.002 [0.007]	-0.048*** [0.005]	-0.002 [0.007]
Non-white	-0.160** [0.068]	-0.290* [0.157]	0.055 [0.089]	-0.292* [0.163]	-0.031 [0.074]	-0.199 [0.147]
Female	-0.764*** [0.167]	-0.239** [0.099]	-0.587*** [0.136]	-0.214** [0.084]	-0.498*** [0.130]	-0.262*** [0.095]
Less than HS degree	-0.998*** [0.202]	-0.400** [0.196]	-0.710*** [0.130]	-0.430** [0.211]	-0.596*** [0.141]	-0.522** [0.206]
High school degree	-0.960*** [0.180]	-0.360** [0.151]	-0.652*** [0.189]	-0.383** [0.159]	-0.618*** [0.170]	-0.403*** [0.146]
Some college or AA degree	-0.693*** [0.112]	-0.322** [0.148]	-0.384*** [0.094]	-0.404** [0.147]	-0.393*** [0.101]	-0.423*** [0.146]
HAART	-1.639*** [0.168]		1.339*** [0.071]		1.327*** [0.076]	
Instruments ^a						
Medically needy threshold				0.004** [0.002]		0.003 [0.005]
SSI threshold >65% of FPL				0.066 [0.144]		0.356 [0.290]
Chi-2 Test for Joint Significance of Instruments				7.62		1.67
p-value of test				0.02		0.43
Lambda ^b		1.319*** [0.161]		-1.278*** [0.060]		-1.269*** [0.055]
State Controls ^c		NO		NO		YES
Region Fixed-Effects		NO		NO		YES
Observations	1396	1396	1396	1396	1396	1396

Robust standard errors clustered by state appear in brackets underneath coefficients.

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

Notes: Except where noted, all data are from HCSUS follow-up 2 sample. Number of partners is top-coded at 24 (1.9% of the sample). All estimates were obtained using ml maximize (with weights and state clusters) in Stata 9.0.

^a State instruments were obtained from Westmoreland (1999)

^b Loading factor on common unobserved heterogeneity. The negative sign of this coefficient implies errors of Partner and HAART equations are negatively correlated

^c State controls include state per capita income in 1998, percent living in urban areas, abortion rate, proportion of GSS respondents (1988-2002) in state who think sexual relations between persons of the same sex are always wrong or almost always wrong, proportion of GSS respondents (1988-2002) in state who pray several times a week. Data on state per capita income in 1998 are based on data from Bureau of economic Analysis available online at <http://www.bea.gov/bea/regional/spi/>. Urbanization data are based on the 2000 Census as reported in the *Statistical Abstract of the United States: 2004-05*. Data on abortion rates (abortions per 1000 females aged 15-44) in 1999 are based on the survey by The Alan Guttmacher Institute (Finer and Henshaw, 2003), and population estimates from the US Census Bureau. We thank Ted Joyce for the abortion rates data.

IV estimates towards the naïve estimates, and cause estimated confidence intervals to be too narrow. The first issue seems minor in our context, since instrumenting for HAART dramatically flips the sign of the effect,³³ but the second issue—known as “test size distortion”—warrants closer consideration. The “size” of a test is the proportion of times a true null hypothesis is rejected; e.g., an ideally sized 95% test has size 5%.

In their Monte Carlo analysis, Staiger and Stock compute test size distortions for IV estimators as a function of the number of instruments, the correlation between error terms across the first- and second-stage equations, and the explanatory power of the instruments. For a correlation between errors of $|0.2|$ (ours is $|0.13|$), they find maximum likelihood estimates with one or two instruments produce very little test size distortion: 95% Wald tests are rejected roughly 10% of the time or less.

To investigate the applicability of their results to our estimation context, and to quantify the impact on our estimated standard errors, we performed our own Monte Carlo experiments patterned after theirs. The adjustment to our standard errors is based on a key idea in Stock and Yogo [2002]. They propose defining instruments as “weak” if the test size distortion exceeds a certain cut-off. They suggest the following alternatives: instruments are “weak” if they reject a true 95% Wald test more than 10%, more than 15%, more than 20%, or more than 25% of the time. We adopt their most stringent definition and ask: by how much would we need to inflate our standard errors to ensure that a true 95% Wald test be rejected less than 10% of the time?

To answer this question, we ran two Monte Carlo experiments, one for the treatment effects data generating process, and another for the joint negative binomial data generating process. Details appear in Appendix VI. In both experiments, we introduced one continuous instrument of varying power (similar to the medically needy threshold), and one binary instrument with less power (like our SSI threshold). We chose error covariance structures

³³This finding is also consistent with theory: Stock and Yogo [2002] suggest that LIML estimates exhibit comparatively little bias from weak instruments.

to match the ones we estimate. Our results were qualitatively similar to those of Staiger and Stock, in the sense that the MLE estimates produced size distortions well below those typically seen in 2SLS. For the treatment effects model (which displays more vulnerability to weak instruments), inflating the standard errors by 2.08 yields hypothesis tests of acceptable size (i.e., a 95% test is rejected no more than 10% of the time) even in the most severe case of truly uncorrelated instruments. Even after applying this worst-case correction to our standard errors in Table IV, we are left with second-stage estimates that are significant at 5% across the board, with the one exception of IV Model 3, which is significant at 10% ($p=0.06$). The negative binomial model is even more robust to weak instruments, requiring little if any inflation of standard errors to achieve acceptably sized tests.

B External Validity Tests

The magnitude of our estimated effects seems consistent with external clinical evidence on the effects of HAART. We obtained estimates from clinical trials of HAART's effect on CD4 cell counts — a proven marker of health status and immune function for the HIV+. The clinical evidence suggests that HAART increases CD4 cell counts by 220 for the median patient [Hammer et al., 1997; Gulick et al., 1997].³⁴ We then estimated the likely impact of a 220 point CD4 increase in sex partners. Specifically, we used our entire HCSUS Follow-Up 2 sample, and regressed number of partners on categorical measures of current (self-reported) CD4 cell count (the categories were: 0-50, 51-200, 201-500, and 500+). Based on this regression, we then calculated the increase in partners that would occur if every member of the HCSUS population had his CD4 count increased by 220. This translates into an increase of 4.1, or a roughly 110% increase in partners. Since these are quite close to the estimates of the treatment effects and negative binomial models, respectively, improved health alone is a plausible mechanism through which HAART could achieve the effects we estimate.

³⁴This is the median change in CD4 after 52 weeks of treatment with HAART. The 25th percentile change was about 140, and the 75th percentile was just over 230 [Gulick et al., 1997].

In addition, helping individuals qualifying for Medicaid also seems a plausible mechanism through which Medicaid rules influence HAART receipt in our models. Most persons with HIV/AIDS gain Medicaid coverage as a result of illness-related disability [Westmoreland, 1999]. Therefore, Medicaid policies should have a larger impact on receipt of HAART for the disabled population. To test this, we re-estimated our first-stage probit model for HAART receipt, but this time allowed the effect of the medically needy threshold to vary with AIDS diagnosis.³⁵ The results suggest that increasing the medically needy threshold from 0% to 100% of the Federal Poverty Line increases the probability of receiving HAART by 26 percentage points for those with AIDS, but by only 6 percentage points for those whose disease has not yet progressed that far. This difference is statistically significant at the 1% level.

VI Treatment Breakthroughs and Infection Risk

HAART increases the sexual activity and the longevity of HIV+ individuals. Both these effects will increase the average risk of encountering an HIV+ sex partner. However, HAART also decreases the concentration of the HIV virus in an infected individual. This lowers infectivity, or the risk of transmission conditional on sexual contact. In this section, we present some simple calculations that weigh these various effects.

A Calculating the Impact of Treatment on Risk

The first-order impact of HAART on infection risk is the change in risk, holding precautionary behavior constant. To estimate this, we begin with the fact that infection risk is the probability that one's partner is HIV+, multiplied by the probability of HIV transmission conditional on contact with an HIV+ partner. Accounting for the clinical evidence that the

³⁵This model included exactly the same demographic and state controls as IV model 2 in Table IV, which is our "middling" specification.

risk of transmission differs by treatment status, we can represent infection risk as:

$$\phi \equiv \Pr(\text{Infection}) = \Pr(\text{Untreated HIV+ Partner}) \Pr(\text{Transmission}|\text{Untreated HIV+ Partner}) + \Pr(\text{Treated HIV+ Partner}) \Pr(\text{Transmission}|\text{Treated HIV+ Partner}) \quad (8)$$

Recall from the theory that π is prevalence, r^+ is the risky behavior of the HIV+ and r the risky behavior of the uninfected. For the purposes of this discussion, we define risky behavior as the number of sex partners. Define τ as the proportion of HIV+ on HAART, r_T^+ the number of sex partners for a treated HIV+ person, r_U^+ the number of partners for an untreated HIV+ person, and r the number of partners for an uninfected person. Given these quantities, the probability of encountering a treated HIV+ partner is: $\Pr(\text{Treated HIV+ Partner}) = \frac{\pi\tau r_T^+}{\pi\tau r_T^+ + \pi(1-\tau)r_U^+ + (1-\pi)r}$. The denominator represents the average number of sex partners population-wide, while the numerator is the portion of the average accounted for by the treated HIV+. $\Pr(\text{Untreated HIV+ Partner})$ is computed similarly.

Since prevalence is quite close to zero (2002 prevalence was 3 per 1,000 US adults) and r is held constant in this first-order welfare calculation, HIV treatment will have little impact on the average number of sex partners in the population. Therefore, the only parameters to consider are: the proportion treated (τ), prevalence (π), sex partners among the HIV+ (r_U^+ and r_T^+), and infectivity ($\Pr(\text{Transmission})$) due to treatment. The percentage impact of treatment on the risk of infection in Equation 8 can thus be approximated using these quantities as:³⁶

$$\frac{\Pr(\text{Infection}|\text{HAART})}{\Pr(\text{Infection}|\text{Post-HAART})} \approx (1 + g_\pi) [(1 - \tau) + \tau ((1 + g_{r^+})(1 - g_I))] \quad (9)$$

g_I is the percentage decline in infectivity due to treatment, g_{r^+} is the percentage increase

³⁶To be precise, $\frac{\Pr(\text{Infection}|\text{HAART})}{\Pr(\text{Infection}|\text{Post-HAART})} = (1 + g_\pi) [(1 - \tau) + \tau ((1 + g_r)(1 - g_I))] \frac{\pi r^+ + (1-\pi)r}{\pi\tau r_T^+ + \pi(1-\tau)r_U^+ + (1-\pi)r}$. If π is close to zero, the last term approaches unity.

in the number of sex partners among the HIV+ due to treatment, and g_π is the percentage increase in HIV prevalence associated with the new treatment (i.e., the increase in prevalence that occurs simply because treatments reduce mortality among the HIV+). Intuitively, the effect of HAART is the growth in prevalence plus the growth in new infections per HIV+ person. The latter depends on the growth in HIV+ sexual activity, minus the reduction in infectivity.

B Estimated Parameters

To estimate the impact of HAART on the infection risk of the uninfected, we need to know: (1) the proportion of HIV+ individuals treated (τ); (2) HAART's percentage impact on sex partners (g_r); (3) HAART's percentage impact on prevalence (g_π); and (4) HAART's percentage impact on infectivity (g_I).

According to HCSUS Follow-Up 2, about 60% of HIV+ individuals were receiving HAART. The percentage change in sex partners can be read directly from the negative binomial coefficients on HAART, which indicates a 133% increase.

The term g_π reflects how much prevalence would rise if HAART were the only thing that changed. We calculate the introduction of HAART by itself would have increased the long-run prevalence of HIV by at least 50%. Long-run prevalence is approximately equal to the incidence of the disease multiplied by the longevity of an HIV+ individual.³⁷ Holding sexual activity constant, incidence will certainly not fall. Moreover, clinical trials show that HAART increases the life-span of treated individuals by 50% to as much as 80% [Freedberg et al., 2001]. Therefore, the decline in mortality on its own increases prevalence by at least 50%.

Our estimate of g_I acknowledges the medical evidence that HAART reduces the infectivity

³⁷Annual steady-state HIV prevalence is equal to $\phi * L$, where ϕ is incidence and L is the expected longevity in years of the HIV+. Defining the survival rate of HIV+ persons as s , and the overall population as Pop , the evolution of the HIV+ population is governed by $HIV_t = sHIV_{t-1} + \phi Pop$. The steady-state expression comes out of this equation, coupled with a constant mortality hazard.

of a treated HIV+ individual [Barroso et al., 2000; Gray et al., 2001; Fiore et al., 2003; Porco et al., 2004]. The clinical reduction in infectivity is tempered by a selection effect, which can cause HAART to increase mean viral load in the sexually active population: reducing viral loads among the healthiest sexually *inactive* patients can move them into the sexually active population, where they may raise mean viral loads. The question, therefore, is what the *empirical* viral load distribution looks like with and without HAART, not just what HAART does to viral load in a randomized trial.

We estimate g_I in two steps. We begin with the empirical viral load distributions of the sexually active HIV+ people in HCSUS, by HAART status. We combine these with estimates of infectivity by viral load from Gray et al. [2001] and Quinn et al. [2000] to arrive at an estimate of average infectivity by HAART status (that is, infectivity weighted by the empirical viral load distributions by HAART status). The Gray et al. [2001] and Quinn et al. [2000] papers are based on a study of 415 heterosexual, monogamous, and discordant (one HIV+ and one HIV- partner) couples in Uganda. The investigators measured viral load for each HIV+ respondent, tested each initially HIV- respondent regularly for HIV seroconversion, and calculated the probability of transmission based on observed coital frequency. The result was a table of infectivity risk by viral load; these are the only such estimates in the literature. Combining this with the empirical effect of HAART on viral loads in the HCSUS, we calculate that the sexually active treated population would exhibit infectivity about 13% lower than the sexually active untreated population.

This leads to an estimated 142 % increase in infection risk:

$$\frac{\Pr(\text{Infection}|\text{HAART})}{\Pr(\text{Infection}|\text{Post-HAART})} \approx (1 + 0.5) [(1 - 0.6) + 0.6 ((1 + 1.33)(1 - 0.13))] = (1.50)(1.62) = 2.42 \quad (10)$$

This is higher than the actual change in incidence (about 42%), which also includes the effect of rising precaution in the uninfected population. All else equal, the pure mortality

reduction led to a 50% increase in risk, sexual activity contributed a 79 % increase, and reduced transmissivity lowered risk by 8%.

The calculation above uses the lowest estimate of HAART’s impact on mortality and the Ugandan study that is widely considered to be the best available analysis of HIV infectivity. However, our qualitative conclusions hold up even when we use the least generous available estimate of infectivity. Porco et al. [2004] calculate a 60% decline in infectivity by comparing empirical rates of transmission pre- and post-HAART. Using this number leads to an estimated 44% increase in the risk of infection. Later, we show that this lower number still suggests a decrease in welfare for uninfected Americans.

VII Normative Implications of Treatment Breakthroughs

While new and improved treatments vastly raise the welfare of the infected, they can sometimes lower welfare for the uninfected by exposing them to a larger and more sexually active pool of HIV+ people. Therefore, new treatments can call for increases in government subsidies for HIV prevention.

A Welfare Effects of Treatment Breakthroughs

Permanent breakthroughs in treatment obviously improve the lifetime welfare of the infected. Less clear, however, is the effect on the private welfare of the uninfected. In the long-run (steady-state), the lifetime welfare of an uninfected person satisfies,

$$v(\pi^{ss}) = \frac{u(r^{ss})}{1 - \beta(1 - \phi(r^{ss}, \pi^{ss}))} + \frac{\beta\phi}{1 - \beta(1 - \phi)} \frac{w(r^+)}{1 - s\beta} \tag{11}$$

This consists of the expected utility an individual derives before falling ill ($\frac{u(r^{ss})}{1 - \beta(1 - \phi(r^{ss}, \pi^{ss}))}$), and the utility expected from the infected state ($\frac{\beta\phi}{1 - \beta(1 - \phi)} \frac{w}{1 - s\beta}$).

Calibrating this expression suggests that treatment breakthroughs are most likely to have first-order lowered the welfare for the uninfected. We can rewrite the lifetime welfare of the uninfected as

$$v(\pi^{ss}) = \frac{u(r)}{1 - \beta(1 - \phi)} \left[1 + \phi \frac{\beta}{1 - s\beta} \frac{w(r^+)}{u(r^{ss})} \right] \quad (12)$$

Consider the first-order welfare change that occurs when infection risk ϕ rises. Take β to be 0.97,³⁸ and pre-HAART survival to be 0.8, based on a constant mortality hazard and an expected lifetime of 5 years from initial infection [Freedberg et al., 2001]. We calculate pre-HAART ϕ to be approximately 0.08%.³⁹ Since a life-year spent with disability is commonly estimated to be about half as valuable as a healthy life-year, we assume that pre-HAART $\frac{w(r^+)}{u(r^{ss})} = 0.5$ [Erickson et al., 1995]. For the post-HAART values, we use the maximum clinical estimate of an 80% increase in longevity from HAART, our most conservative estimate that first-order infection risk ϕ rose by 44%, and conservatively assume that HIV+ individuals become perfectly healthy, so that $w = u(r^{ss})$. Even under these conservative assumptions, calibrating Equation 12 reveals a first-order welfare decline for the uninfected.⁴⁰

B Optimal Public Prevention

Welfare decline for the uninfected can never in itself justify the end of treatment provision, whether from a Paretian, public health, or pragmatic point of view. The relevant question is how the government can manage and ameliorate this effect. Prevention subsidies can be an efficient way to do so.

Prevention subsidies have value because a healthy uninfected person does not consider the impact of his actions on tomorrow's rate of infection. In a long-run steady-state, the

³⁸This is based on a real interest rate of 3%, assuming that β is the reciprocal of unity plus the real rate.

³⁹In 1995, the incidence of HIV was about 4.75 per 100,000 people. In that year, about half the US population was over age 18, and (according to the 2000-02 GSS) approximately 13% of US adults were unmarried with multiple sex partners in the last year. Assuming conservatively that only such people are at risk for HIV, the incidence is 5 per 6500 people at risk, or 0.08%.

⁴⁰This result is less likely to hold in developing countries where prevalence and infection risk are extremely high, or in extremely high-risk subpopulations. However, for the average sexually active, non-monogamous American, expected utility is likely to have fallen.

external costs of risk-taking by the uninfected can be written as

$$\underbrace{\phi_r(1-\pi)}_{\text{Marginal Incident Cases}} * \underbrace{\left\{ \frac{\beta\phi_\pi}{1-\beta(s+\phi_\pi(1-\pi))} \right\}}_{\text{Incidence Multiplier}} * \underbrace{\left\{ \beta \left(\frac{w}{1-s\beta} - v \right) \right\}}_{\text{Cost of becoming HIV+}} \quad (13)$$

The long-run external cost of risk-taking consists of: (1) The number of incident cases that result from a unit change in risky behavior; (2) The long-run increase in the number of cases that results from a single incident case, which we call the “incidence multiplier;” and (3) The cost to an individual of becoming infected. Since the external cost can be interpreted as the per unit tax on risky behavior, understanding changes in external costs provides insight into the optimal subsidy for prevention. HIV breakthroughs always raise the first component of the external cost,⁴¹ have ambiguous effects on the second, and always lower the third. The need for and degree of government prevention subsidies thus turn on the relative importance of these three effects.

A variety of subtle comparisons can be made based on the analysis above, but the broad outlines are straightforward. Increased prevention is called for when new treatments significantly raise the prevalence of the disease, relative to the cost of being infected. In some cases, sufficient prevention may be undertaken by the private sector alone. When this fails to be enough, prevention subsidies may a tool to mitigate the erosion of welfare for the HIV-negative.

The case for public prevention subsidies is particularly strong in the case of foreign aid to countries ravaged by HIV. Treatment aid may need to be systematically coupled with resources for prevention, because a one-time foreign aid gift does not reduce the long-run cost of infection, but may raise long-run prevalence.

⁴¹Formally, ϕ_r rises with π . This could be offset if r rises, but this is not the empirical experience of the US. In addition, if π is extremely large, it is possible that increases in prevalence shrink the pool of uninfected people to such an extent that there are fewer new cases of HIV for a given increase in risk-taking.

VIII Conclusions

Treating an incurable infectious disease like HIV involves unique issues for individual behavior and social policy. While improvements in treatment clearly raise the welfare of infected individuals, societies must cope with the reality that treatment can fuel the further spread of the disease. We have presented empirical evidence consistent with this argument and have developed a few of its implications for private behavior and public policy.

HIV treatment appears to increase the number of sex partners of HIV+ individuals by 133%. For HIV- people, this increases infection risk by at least 44%, and possibly more than doubles it. If this increased infection risk outweighs the reduction in the cost of the disease, HIV breakthroughs might actually lower the lifetime welfare of the uninfected. A simple calibration of the theoretical model suggests that it does so.

The government's best response to treatment breakthroughs may be to increase public spending on prevention. When treatment involves large negative external costs, it is efficient for the government to respond by boosting prevention incentives, which benefit the uninfected. An incurable infectious disease like HIV can generate a unique complementarity between treatment breakthroughs and prevention. Looking to the longer-run, our results suggest that a cure for HIV would have extraordinary welfare benefits, because it would break the link between treatment and risky behavior by infected individuals.

Our research also suggests the importance of investigating the impact of treatment breakthroughs on the behavior of the uninfected. Elsewhere, we have found that the introduction of HAART coincided temporally with an increase in condom usage by unmarried and non-monogamous people, but no change in condom usage by married or monogamous people [Lakdawalla and Sood, 2004]. Further work is needed to determine whether HIV breakthroughs were the dominant force in generating this behavior. These results will help further quantify the long-run welfare impacts of breakthroughs on the uninfected population.

APPENDIX

I GSS Data

This section describes the construction of the GSS variables used in the paper.

Education is reported by grade attainment and postsecondary years completed. Respondents are asked how many primary and secondary grades they have completed, and how many years of college they attended. Like in the HCSUS analysis, therefore, we have information on the attainment of a high school degree. However, we have years of postsecondary schooling, rather than the postsecondary degree attainment data in HCSUS.

Income is reported categorically, and the categories change across survey years. In our analysis, we consistently use annual income deciles, according to the relevant GSS income categorization for that survey year. For 1988 to 1990, the categories are: under \$1K, \$1K-2.999K, \$3K-3.999K, \$4K-4.999K, \$5K-5.999K, \$6K-6.999K, \$7K-7.999K, \$8K-9.999K, \$10K-12.499K, \$12.5K-14.999K, \$15K-17.499K, \$17.5K-19.999K, \$20K-22.499K, \$22.5K-24.999K, \$25K-29.999K, \$30K-34.999K, \$35K-39.999K, \$40K-49.999K, \$50K-59.999K, and \$60K+. The 1991-6 surveys added the categories: \$60K-74.999K, and \$75K+. In addition to these, the 1998-2002 surveys added: \$75K-89.999K, \$90K-109.999K, and \$110K+.

Sexual Partners and Risk Categories are based on the following series of questions. “How many sex partners have you had in the last 12 months?” “Was one of the partners your husband or wife or regular sexual partner?” “If you had other partners, please indicate all categories that apply to them: Close personal friend; Neighbor, Co-worker, or long-term acquaintance; *Casual date or pick-up*; *Person you paid or paid you for sex*; Other (Please Specify).” We use the two italicized partner types as two of our three criteria for risky sex. The last criterion is obtained from the question, “Thinking about the time since your 18th birthday, have you ever had sex with a person you paid or who paid you for Sex?” The latter question is available from 1991 onwards, and is not in the 1988-90 survey years.

Sexual Orientation (for males) is used in our re-weighting of the GSS. We define as gay any man who reports having had a gay sexual encounter since the age of 18 (more stringent definitions produced qualitatively similar results). Specifically, the GSS asks, “Now thinking about the time since your 18th birthday (including the past 12 months) how many male partners have you had sex with?”

Distribution of Partners in HCSUS and GSS subsamples. The distribution of HIV+ partners likely lies somewhere between the risky males sample and the weighted adults. Consider two scenarios: (1) HCSUS respondents who reported zero or one partners in the last 6 months remained abstinent or monogamous for the last 12 months, but that other respondents ended up with twice as many partners over 12 months as they would have in 6 months; or (2) HCSUS respondents who reported abstinence over the last 6 months remained abstinent or monogamous over 12 months, but all others had twice as many partners over the 12-month period. These alternative approaches produce partner distributions [0-1 partners, 2 partners, and 3+ partners] of [74%,0%,26%] and [34%,40%,26%], respectively. Both these would be less risky than the distribution for the weighted risky sample, but both are also more risky than the weighted adults sample.

II HCSUS Data

In this section, we describe the variable definitions and sample selection for the HCSUS data. Detailed technical descriptions of the HCSUS sample and weighting procedure appear in Duan et al. [1999]. Questionnaires and information about the survey are available on the web at <http://www.rand.org/health/projects/hcsus/index.html>.

We construct analytic weights to calibrate the sample to the reference population. A respondent’s analytic weight, which may be interpreted as the number of people in the population represented by that respondent, is the product of three patient-specific quantities: the sampling weight, the multiplicity weight, and the non-response weight. The sampling

weight adjusts for oversampling (of women, for example); the multiplicity weight adjusts for patients who could potentially enter the sample via multiple providers; and the non-response weight adjusts for differential cooperation [Duan et al., 1999]. All analyses presented in this paper use these weights.

Number of partners and **number of new partners** were self-reported. Respondents were asked “how many different people have you had either oral, anal, or vaginal sex in the last 6 months? Your best estimate is fine.” Respondents who reported at least one sex partner were asked “Of these people, how many of them were new partners, that is, people you had oral, anal, or vaginal sex with for the first or only time in the last 6 months?”

The indicator for **HAART** use was constructed by Andersen et al. [2000] and is based on recommendations published by the Centers for Disease Control and Prevention. It is based on detailed self-reports of prescription drug use. Specifically, HAART was defined as using a protease inhibitor (PI), a non-nucleoside reverse transcriptase inhibitor (NNRTI), or a nucleoside reverse transcriptase inhibitor (NRTI) in various combinations. For example, HAART includes: two or more NRTIs in combination with at least one PI or one NNRTI; or one NRTI in combination with at least one PI and at least one NNRTI. Combinations of older drugs such as zidovudine, which is an NRTI, with either a PI or NNRTI, were not considered HAART.

Education is reported by degree attainment and years completed. Respondents were asked, “How many years of regular school or college did you ever complete and get credit for?” Respondents were also asked about highest degree or diploma they have attained. Based on the responses to these questions we classified respondents into 4 mutually exclusive education groups: (1) Less than high school degree – highest degree is “none/less than high school” and years of schooling is less than or equal to 9 years, (2) High school degree – highest degree is “high school diploma or G.E.D.” and years of school is less than or equal to 12 years, (3) Some college or AA degree – highest degree is high school diploma, AA degree

or 2 year college degree and years of schooling is less than or equal to 15 years, (4) College degree – highest degree is 4-year college degree, or professional/graduate degree and years of schooling is 16 years or more.

We used 2 measures of health status from the HCSUS – **lowest-ever CD4 count** and **AIDS**. CD4 cells are immune cells and lower cell counts imply worse health. The data on lowest ever CD4 cell count are self reports based on the the question “Of all your CD4 tests, what was your lowest count?”. AIDS is based on self reports of a detailed list of clinical conditions and symptoms. Respondents were coded as having AIDS if they had any of the “AIDS-defining or indicator conditions” outlined by the CDC definition of AIDS, or if their CD4 count was sufficiently low as to indicate AIDS according to CDC guidelines.

We used a measure of **unsafe sex** from the HCSUS, based on the question: “When you have had sex since finding out you are HIV+, how often have you had unsafe sex? (1) Always, (2) Usually, (3) About half the time, (4) Sometimes, or (5) Never.” In our analysis, people who responded about half the time, sometimes, or never were coded as practicing unsafe sex. This information was combined with information on the number of partners in the last 6 months to classify respondents into 4 mutually exclusive categories: abstinent, monogamous, safe sex with multiple partners, or unsafe sex with multiple partners.

III State-Level Characteristics Data

Data on **Medicaid policies** (Medically Needy threshold and SSI threshold) is from the Kaiser Family Foundation publication on Medicaid policy related to HIV/AIDS [Westmoreland, 1999]. Both income eligibility thresholds are expressed as a percentage of the Federal Poverty Line (FPL).

Data on **state per capita income** in 1998 is from the Bureau of Economic Analysis, available online at <http://www.bea.gov/bean/regional/spi/>. Per capita income is total personal income divided by total midyear population. Data on personal income are from

national income accounts; population data are from the Bureau of the Census.

Data on **abortion rates** (abortions per 1000 females aged 15-44 years) in 1999 are based on an abortion provider survey conducted by The Alan Guttmacher Institute and population estimates from the Bureau of the Census. Details of the survey design are available in Finer and Henshaw [2003]. These data were provided by Ted Joyce.

Data on **attitudes towards homosexuality** are based on author tabulations from all the 1988-2002 GSS surveys. Respondents were asked: “What about sexual relations between two adults of the same sex—do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?” Pooling all years, for each state we calculated the proportion of adults that thought homosexual relations are always wrong or almost always wrong.

Data on **religiosity** are also based on tabulations from all the 1988-2002 GSS surveys. Respondents were asked: “About how often do you pray? Several times a day, once a day, several times a week, once a week, less than once a week, or never?” Pooling all years, for each state we tabulated the proportion of adults who pray at least several times a week.

Data on **urbanization** measures percent of the population in a state living in urban areas. These data are based on the 2000 Census and reported in the *Statistical Abstract of the United States: 2004-05*.

IV Negative Binomial Model

Under the negative binomial model we have constructed, the probability that an individual i will have s_i sex partners in a given year is:

$$\Pr(S_i = s_i | Z_i, \lambda, \delta_2, \eta_i) \equiv NB2(s_i | Z_i, \lambda, \delta_2, \eta_i) = \frac{\Gamma(s_i + \frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})\Gamma(s_i + 1)} \left(\frac{\frac{1}{\alpha}}{\mu_i + \frac{1}{\alpha}} \right)^{\frac{1}{\alpha}} \left(\frac{\mu_i}{\mu_i + \frac{1}{\alpha}} \right)^{s_i} \quad (\text{IV-1})$$

The contribution of patient i to the log-likelihood function is then given by:

$$l_i = \left[\int NB2(s_i|Z_i, \lambda, \delta_2, \eta_i) \Phi(\delta_1 X_i - \eta_i) \phi(\eta_i) d\eta_i \right]^{T_i=1} * \left[\int NB2(s_i|Z_i, \lambda, \delta_2, \eta_i) (1 - \Phi(\delta_1 X_i - \eta_i)) \phi(\eta_i) d\eta_i \right]^{T_i=0} \quad (\text{IV-2})$$

Since it is not feasible to integrate over η directly, we approximate this by drawing J values of η from a standard normal distribution. The estimates in the paper are based on analysis with 1000 draws using Halton sequences. The approximate contribution of patient i to the log-likelihood function is given by:

$$l_i \approx \left[\frac{1}{J} \sum_{j=1}^J NB2(s_i|Z_i, \lambda, \delta_2, \eta_j) \Phi(\delta_1 X_i - \eta_j) \right]^{T_i=1} * \left[\frac{1}{J} \sum_{j=1}^J NB2(s_i|Z_i, \lambda, \delta_2, \eta_j) (1 - \Phi(\delta_1 X_i - \eta_j)) \right]^{T_i=0} \quad (\text{IV-3})$$

V Steady-State Dynamics

The value function $v(\pi)$ is convex decreasing, because its associated functional mapping satisfies Blackwell's sufficient conditions for a contraction mapping, and because that mapping takes a convex decreasing function and produces another convex decreasing function: if v is convex decreasing, and ϕ is concave increasing in π , $\beta\phi * \left[\frac{w}{1-s\beta} - v(\pi') \right] + \beta v(\pi')$ is convex decreasing in π . Therefore, the fixed point of the mapping must also be convex decreasing.

To ensure the uniqueness and existence of a steady-state, we make two assumptions:

Assumption V.1. $\frac{\partial}{\partial \pi} \left[\frac{du}{dr} - \beta \frac{\partial \phi}{\partial r} \left(v^{ss} - \frac{w}{1-s\beta} \right) \right] < 0$

Assumption V.2. $1 - s - \phi_\pi > 0$

Assumption V.1 means that the private marginal utility of risk-taking falls when prevalence rises. Assumption V.2 implies that higher steady-state prevalence levels must be sup-

ported by higher levels of risk-taking by the uninfected.

Assumption V.1 is nontrivial, because increased prevalence has offsetting effects on the marginal utility of risk-taking: it raises the impact of risk-taking on exposure ($\phi_{r\pi} > 0$), but it lowers the lifetime utility of a healthy person relative to a sick one. High prevalence could encourage risk-taking, because staying healthy becomes more costly. This point is emphasized by Kremer [1996]. For observed US prevalence levels (lower than developing country levels), it appears that prevalence discourages risk-taking [Ahituv et al., 1996; Philipson, 2000].

Assumption V.2 is very likely to hold in the pre-HAART equilibrium, and thus is useful for assessing the effect of HAART. Note that, at any steady-state, it must be true that $1 - \phi_\pi > 0$. Therefore, when survival rates s are low, the assumption condition is also likely to hold. Comparative dynamics that begin with the post-HAART equilibrium might be different, if rates of healthy survival increase dramatically so as to violate this assumption. This would lead to unstable equilibria.

Given Assumption V.1, the optimal risk-taking level falls with prevalence (which lowers the marginal utility of risky behavior), and rises with breakthroughs in treatment s (which lower the cost of infection, holding prevalence fixed). Given Assumption V.2, the steady-state risk level rises with steady-state prevalence, but falls with s . When s is higher, infected people live longer, and it requires less risk-taking to support a given level of prevalence. This leads to the two curves in Figure II that define a unique steady-state equilibrium.

Finally, we make an assumption analogous to Assumption V.1.

$$\text{Assumption V.3. } \frac{\partial}{\partial r^+} \left[\frac{du}{dr} - \beta \frac{\partial \phi}{\partial r} \left(v^{ss} - \frac{w}{1-s\beta} \right) \right] < 0$$

This assumption implies that the increase in risky behavior by the HIV+ lowers risk-taking among the HIV-. Generally, there are two offsetting effects: increases in r^+ lower the relative benefit of remaining healthy, but raise the risk of infection. Just as with prevalence, we assume that the latter effect dominates the former in developed countries.

VI Monte Carlo Analyses

To investigate the impact of weak instruments on our ability to make valid statistical inferences, we conducted Monte Carlo analyses modeled on those of Staiger and Stock [1997], who performed similar experiments for continuously linear models. Following their lead, we focus purely on the weak instruments problem, and abstract from other potential concerns like misspecification or omitted variables. We performed two sets of experiments: one where the data-generating process matches the treatment effects model, and one where it matches the joint negative binomial model.

A TREATMENT EFFECTS MODEL

Suppose that the data-generating process is characterized by the following equations:

$$T_i^* = \beta_0 + \frac{\beta_1}{\sqrt{N}}Z_1 + \beta_2Z_2 + \epsilon_T \tag{VI-4}$$

$$T_i = (T_i^* > 0) \tag{VI-5}$$

$$S_i = \alpha_0 + \alpha_1T_i + \epsilon_S \tag{VI-6}$$

$$(\epsilon_S, \epsilon_T) \sim BVN(\mu_S = \mu_T = 0, \rho = -0.2, \sigma_S^2 = 1, \sigma_T^2 = 1) \tag{VI-7}$$

Just as in Staiger and Stock, all parameters are normalized to standard values except for the between-equation correlation term. The correlation between error terms is picked to match our empirical estimate of $\rho = -0.2$.

T_i^* is the latent treatment index, and T_i is the binary treatment variable. Z_1 is the continuous instrument, modeled as a standard normal. We normalize its coefficient by \sqrt{N} (N is the number of observations) to make our test statistics invariant to sample size; this is the same approach adopted by Staiger and Stock [1997]. Z_2 is a binary variable with mean

0.9 (to match our SSI threshold variable) and generated as:

$$Z_2 = (Z_2^* > 0.9) \tag{VI-8}$$

$$Z_2^* \sim \text{Uniform}[0, 1] \tag{VI-9}$$

For ease of interpretation and testing, we set all the true parameter values to be zero, except for β_1 , the power of the continuous instrument, which we allowed to vary. The uncorrelatedness of the binary instrument Z_2 is a conservative approach to our finding that the SSI threshold was not individually significant in any of our models.

For each replication of the model above, we generated 2000 observations using the above equations, and (based on these generated observations) calculated the treatment effects estimate of $\hat{\alpha}_1$, its associated standard error, and the joint chi-squared statistic associated with testing the hypothesis that $(\beta_1 \ \beta_2) = \vec{0}$. To analyze the consequences of weak instruments, we allowed the coefficient β_1 to take on the values 0, 1, 3, 5, where 0 is the worst-case of uncorrelated instruments.

We replicated the model 5000 times.⁴² In other words, for each value of c , we had 5000 estimates of $\hat{\alpha}_1$, its standard error, and the first-stage chi-squared. With this set of estimates, we calculated the proportion of times the estimator rejected the hypothesis $\alpha_1 = 0$ with 95% confidence. We call this the “size” of the hypothesis test.

Finally, for each value of β_1 , we calculate the extent to which we would have to inflate the estimated standard error in order to produce a size of 10% or less, which would meet Stock and Yogo [2002]’s most stringent definition of strong instruments.⁴³ We call this the “standard error inflation factor.” These factors are calculated via grid search.

⁴²We ran two sets of 5000 replications each and found virtually identical estimates from each set.

⁴³They propose the following alternative tests of “weak” instruments: an instrument is weak if it rejects a true 5% Wald test more than 10%, more than 15%, more than 20%, or more than 25% of the time.

B NEGATIVE BINOMIAL MODEL

We repeated the Monte Carlo exercise above for the joint negative binomial-probit data-generating process. Here, we assumed that the data were generated according to:

$$T_i = \delta_0 + \delta_1 Z_1 + \delta_2 Z_2 - \nu_{T,i} - \eta_i \quad (\text{VI-10})$$

$$\mu = \exp(\delta_3 - \theta T_i - \lambda \eta_i - \nu_{S,i}) \quad (\text{VI-11})$$

$$S_i \sim \text{Poisson}(\mu) \quad (\text{VI-12})$$

$$\nu_{T,i} \sim N(0, 1) \quad (\text{VI-13})$$

$$\eta_i \sim N(0, 1) \quad (\text{VI-14})$$

$$\nu_{S,i} \sim \Gamma(\alpha) \quad (\text{VI-15})$$

$$\alpha = 1, \lambda = -1.3 \quad (\text{VI-16})$$

Z_1 and Z_2 are defined as before, and as before, we set $\delta_2 = 0$ to model the weakness of the binary instrument. The value for λ is chosen to match our empirical estimates of the correlation structure. We set $\delta_3 = 1$, since this generates a mean of 3.8, close to the empirical mean for partners. For convenience, we set $\delta_0 = \theta = 0$. We allow the power of the continuous instrument to vary across $\delta_1 = 0, 1, 3, 5$.

For each value of δ_1 , we ran 1000 replications, at which point our results were quite stable. For each replication, we estimated the joint negative binomial-probit model via the method of simulated maximum likelihood.⁴⁴ This yielded an estimate $\hat{\theta}$, a corresponding standard error, and a first-stage chi-squared statistic. Based on these estimates, we computed test sizes for 95% confidence intervals around $\hat{\theta}$, and calculated inflation factors for the standard error of $\hat{\theta}$.

⁴⁴The SMLE estimator was computed using 100 draws.

C RESULTS

The results of the two Monte Carlo experiments are given in Table VI-1. The “Size” column reports the actual size of a 95% Wald test that compares the relevant estimate to its mean. “Size” is defined as the proportion of times the null hypothesis is rejected; therefore, an ideal estimator has size 5%. The inflation factor is the extent to which we would have to inflate the standard error on the second-stage treatment effect estimate, in order to obtain size less than or equal to 10%.

The Monte Carlo experiments demonstrate that the negative binomial model is not very sensitive to the weak instruments problem; test sizes are similar to those found by Staiger and Stock [1997] for LIML estimates of linear models. Sizes are all under 10%; therefore, even for uncorrelated instruments, the estimator does not meet the Stock and Yogo [2002] definition of a weak instrument, which should have size greater than 10%.

The treatment effects model is much more susceptible to the weak instruments problem. However, even in the worst-case scenario of uncorrelated instruments, inflating the second-stage standard errors by 108% would produce acceptably sized hypothesis tests. Applying this correction to all our treatment effects estimates would still yield significant estimates across the board. The estimates in IV Models 1 and 2 would be significant at 5%, and the estimate in IV Model 3 has a p-value of 0.06.

References

- Ahituv, A., V. J. Hotz, and T. J. Philipson (1996). “The Responsiveness of the Demand for Condoms to the Local Prevalence of AIDS.” *Journal of Human Resources* 31(4):869–897.
- Andersen, R., et al. (2000). “Access of Vulnerable Groups to Antiretroviral Therapy Among Persons in Care for HIV in the U.S.” *Health Services Research* 35(2):389–416.

Table VI-1

The consequences for standard errors of weak instruments in the Treatment Effects and Joint Negative Binomial Models.

Strength of Instruments	Size^a	Standard Error Inflation Factor^b
<i>Treatment Effects Model^c</i>		
Uncorrelated with endogenous variable	41.3%	108%
Mean Chi-square = 2.7	38.7%	106%
Mean Chi-square = 7.8	28.3%	82%
Mean Chi-square = 12.3	21.9%	57%
<i>Joint NB2-Probit Model^d</i>		
Uncorrelated with endogenous variable	7.7%	0%
Mean Chi-square = 2.4	5.3%	0%
Mean Chi-square = 4.8	5.2%	0%
Mean Chi-square = 9.8	4.8%	0%

Notes:

^a Percent of replications that reject the null hypothesis that the estimated treatment effect is equal to the true mean of the estimate, at a 95% level of significance. Size greater than 5% implies that computed standard errors are lower than true standard errors.

^b Standard error inflation factor is the percentage increase in the treatment effect standard error test for non-weak instruments.

^c Based on results from 5000 replications of a monte carlo experiment that estimated a treatment effects model with a binary endogenous (treatment) variable and a continuous outcome variable. The latent treatment variable and the outcome variable were jointly distributed standard bivariate normal with correlation coefficient -0.2. The treatment effect was assumed to be zero and was identified using 2 instruments -- one distributed normally and the other as a binary variable.

^d Based on results from 1000 replications of a monte carlo experiment that estimated a joint NB2-Probit model with a binary endogenous (treatment) variable and an integer-valued outcome. Unobserved treatment propensity and the outcome were assumed to be distributed standard normal with covariance -1.3. Conditional on this common unobserved heterogeneity, the latent treatment variable was distributed standard normal and the outcome variable was distributed NB2. The treatment effect was assumed to be zero and was identified using 2 instruments -- one distributed normally and the other as a binary variable.

- Auld, M. C. (2003). “Choices, Beliefs, and Infectious Disease Dynamics.” *Journal of Health Economics* 22(3):361–77.
- Barroso, P., et al. (2000). “Effect of Antiretroviral Therapy on HIV Shedding in Semen.” *Annals of Internal Medicine* 133:280–284.
- Bhattacharya, J., D. Goldman, and N. Sood (2003). “The Link between Public and Private Insurance and HIV-Related Mortality.” *Journal of Health Economics* 22(6):1105–22.
- Carpenter, C., et al. (2000). “Antiretroviral Therapy in Adults: Updated Recommendations of the International AIDS Society-USA Panel.” *Journal of the American Medical Association* 283(3):381–90.
- Centers for Disease Control (1998). “Diagnosis and Reporting of HIV and AIDS in States with Integrated HIV and AIDS Surveillance — United States, January 1994 - June 1997.” *Morbidity and Mortality Weekly Report* 47(15):309–314.
- Crepaz, N., T. A. Hart, and G. Marks (2004). “Highly Active Antiretroviral Therapy and Sexual Risk Behavior.” *Journal of the American Medical Association* 292(2):224–236.
- Deb, P., and P. Trivedi (2004). “Provider Networks and Primary Care Signups: Do They Restrict the Use of Medical Services?” *Technical report*. Hunter College, CUNY.
- Detels, R., et al. (1998). “Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators.” *Jama* 280(17):1497–503.
- Duan, N., et al. (1999). “HCSUS Baseline Methods Technical Report: Weighting, Imputation and Variance Estimation.” *RAND Report MR-1060-AHCPR*. Santa Monica, CA: RAND.
- Duggan, M. G., and W. N. Evans (2005). “Estimating the Impact of Medical Innovation: The Case of HIV Antiretroviral Treatments.” *Working Paper 11109*. Cambridge, MA: National Bureau of Economic Research.
- Dukers, N., et al. (2002). “HIV Incidence on the Increase Among Homosexual Men Attending an Amsterdam Sexually Transmitted Disease Clinic: Using a Novel Approach for Detecting Recent Infections.” *AIDS* 16:F19–F24.
- Erickson, P., R. Wilson, and I. Shannon (1995). “Years of Healthy Life.” *Healthy People 2000 Statistical Notes* (7):1–15.
- Fauci, A., et al., eds. (1998). *Harrison’s Principles of Internal Medicine, 14th Edition*. New York: McGraw Hill.
- Finer, L. B., and S. K. Henshaw (2003). “Abortion Incidence and Services in the United States in 2000.” *Perspectives on Sexual and Reproductive Health* 35(1).
- Fiore, J., et al. (2003). “Correlates of HIV-1 shedding in cervicovaginal secretions and effects of antiretroviral therapies.” *AIDS* 17(15):2169–76.

- Freedberg, K. A., et al. (2001). "The cost effectiveness of combination antiretroviral therapy for HIV disease." *N Engl J Med* 344(11):824–31.
- Goldman, D., et al. (2001). "The Effect of Insurance on Mortality in an HIV+ Population in Care." *Journal of the American Statistical Association* 96(405):883–894.
- Gray, R. H., et al. (2001). "Probability of HIV-1 Transmission Per Coital Act in Monogamous, Heterosexual, HIV-1-Discordant Couples in Rakai, Uganda." *Lancet* 357:1149–53.
- Gremy, I., and N. Beltzer (2004). "HIV Risk and Condom Use in the Adult Heterosexual Population in France Between 1992 and 2001: Return to the Starting Point?" *AIDS* 18:805–809.
- Gulick, R., et al. (1997). "Treatment with Indinavir, Zidovudine, and Lamivudine in Adults with Human Immunodeficiency Virus Infection and Prior Antiretroviral Therapy." *New England Journal of Medicine* 337:734–739.
- Hammer, S. M., et al. (1997). "A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team." *N Engl J Med* 337(11):725–33.
- Jordan, R., et al. (2002). "Systematic Review and Metaanalysis of Evidence for Increasing Numbers of Drugs in Antiretroviral Combination Therapy." *British Medical Journal* 324:1–10.
- Katz, M., et al. (2002). "Impact of Highly Active Antiretroviral Treatment on HIV Seroincidence Among Men Who Have Sex with Men: San Francisco." *American Journal of Public Health* 92:388–394.
- Kremer, M. (1996). "Integrating Behavioral Choice Into Epidemiological Models of AIDS." *Quarterly Journal of Economics* 111(2):549–572.
- Lakdawalla, D., and N. Sood (2004). "HIV Breakthroughs and the Growth in Safe Sex." *Technical report*. RAND Corporation. Mimeo.
- Lucas, G. M., R. E. Chaisson, and R. D. Moore (1999). "Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions." *Ann Intern Med* 131(2):81–7.
- Lucas, G. M., R. E. Chaisson, and R. D. Moore (2003). "Survival in an urban HIV-1 clinic in the era of highly active antiretroviral therapy: a 5-year cohort study." *J Acquir Immune Defic Syndr* 33(3):321–8.
- Maddala, G. (1986). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Moatti, J., et al. (2003). "Access to Antiretroviral Treatment and Sexual Behaviours of HIV-Infected Patients Aware of their Serostatus in Cote d'Ivoire." *AIDS* 17:S69–S77.

- Over, M., et al. (2004). *HIV/AIDS Treatment and Prevention in India: Modeling the Costs and Consequences*. Human Development Network Health, Nutrition and Population Series, Washington, DC: World Bank.
- Palella, F., et al. (1998). “Declining Morbidity and Mortality Among Patients with Advanced Human Immunodeficiency Virus Infection.” *New England Journal of Medicine* 338(13):853–60.
- Philipson, T. (2000). “Economic Epidemiology and Infectious Diseases.” *Handbook of health economics. Volume 1B*. A. J. Culyer and J. P. e. Newhouse, eds. New York: Elsevier Science. 1761–1797.
- Philipson, T., and R. Posner (1993). *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective*. Cambridge, MA: Harvard University Press.
- Porco, T., et al. (2004). “Decline in HIV infectivity following the introduction of highly active antiretroviral therapy.” *AIDS* 18(1):81–88.
- Quinn, T., et al. (2000). “Viral Load and Heterosexual Transmission of HIV Type I.” *New England Journal of Medicine* 342(13):921–929.
- Staiger, D., and J. H. Stock (1997). “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65(3):557–586.
- Staszewski, S., et al. (1999). “Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. Study 006 Team.” *N Engl J Med* 341(25):1865–73.
- Stock, J. H., and M. Yogo (2002). “Testing for Weak Instruments in Linear IV Regression.” *Working Paper 284*. National Bureau of Economic Research.
- Vittinghoff, E., et al. (1999). “Combination antiretroviral therapy and recent declines in AIDS incidence and mortality.” *J Infect Dis* 179(3):717–20.
- Westmoreland, T. (1999). *Medicaid and HIV/AIDS Policy: A Basic Primer*. Washington, DC: Henry J. Kaiser Family Foundation. Available online at <http://www.nastad.org/PublicPolicyResources/1891-KFF.pdf>.