Forum for Health Economics & Policy

Volume 10, *Issue* 2 2007 *Article* 5

(ECONOMICS OF THE HIV EPIDEMIC)

Risky Sexual Behavior, Testing, and HIV Treatments

Stéphane Mechoulan*

*University of Toronto, s.mechoulan@utoronto.ca

Risky Sexual Behavior, Testing, and HIV Treatments*

Stéphane Mechoulan

Abstract

This paper studies the impact of antiretroviral therapies (ARVs) on HIV testing and risky sexual behavior. I use data collected in San Francisco among a high-risk population from 1994 to 2002. The evidence supports the hypothesis of a causal link between the introduction of ARVs in late 1996 and the sharp increase in risky sexual behavior that ensued. Further, following ARVs, testers take more risks while non-testers take fewer risks. The proportion of testers remains stable, which was ambiguous *a priori*. To the extent that ARVs may induce changes in the composition of the testing and non-testing groups, such effects do not seem to affect the results.

KEYWORDS: HIV, AIDS, testing, HAART, ARV, risk, UAI

---

*I am indebted to Gordon Anderson, David Bishai, Don Dewees, William Evans, Andrew Francis, Dino Gerardi, Anupam Jena, Emily Oster, Andreas Park, Neeraj Sood, two anonymous referees and seminar participants at the University of Toronto, participants in the 4th Canadian Health Economics Study Group in Toronto, and in the 5th meetings of the International Health Economics Association in Barcelona for their helpful comments and suggestions. I thank all the undergraduate students enrolled under the University of Toronto Research Opportunity Program who worked on this project for their input. All errors are my own. Contact: s.mechoulan@utoronto.ca

I. Introduction

The HIV virus has caused worldwide devastation. At least 60 million people have been infected, 90% of them living in developing countries. It is believed that over 20 million people have died of AIDS (WHO, 2005). The emergence of highly active Anti-Retro-Viral Therapies (ARVs) in the course of 1996 has been one of the most radical steps in the treatment of AIDS.

Expanding access to treatments forms a key strategy of the international community to fight AIDS: the stated objective of the WHO is to have universal access to comprehensive prevention programs, treatment, care and support by 2010. It is urgent to consider the behavioral impacts of HIV drugs in order to understand their consequences on the spread of the disease. In particular, by inducing "treatment optimism," they may change the fear of contracting HIV and the incentives for testing for HIV.

In this paper, I present a theoretically motivated empirical analysis of the effects of ARVs on testing, unprotected sex, and number of sexual partners. Beyond characterizing the separate impacts of ARVs on testing and risky sexual behavior, I examine the differential effects of ARVs on risky sexual behavior between testers and non-testers. This also entails an analysis of how the new treatments may alter the composition of the tested and non-tested groups.

Using intuition alone, the effects of better treatments are complex to interpret. Assuming a constant population, a couple of preliminary thought experiments should help fix the ideas.

(a) First, holding testing likelihood constant, ARVs lower the cost of getting HIV: this is the direct and most straightforward effect. (b) However, as treatment quality improves, HIV prevalence is expected to rise: this stems from the increase in the proportion of healthier HIV+ individuals present in the market for risky exposures: this is the indirect effect. (c) Finally, ARVs reduce the average concentration of virus, thus lowering the risk of transmission per risky act. Margins (a) and (c) suggest more risk-taking while margin (b) suggests less: which will dominate is thus theoretically ambiguous.

Now let us hold risky behavior constant. (a) ARVs increase the benefit of testing because one can then initiate more efficient treatments conditional on a positive test. (b) For the reasons described above, ARVs increase the prevalence of HIV in the general population: an increase in the prevalence of HIV enhances the value of testing − irrespective of advances in treatment. (c) On other hand, it may be that prospective partners become less demanding of HIV tests once ARVs become available. Margins (a) and (b) suggest more testing while margin (c) suggests less: again, which will dominate is uncertain *a priori*.

Clearly, the decision regarding risk-taking *and* testing is the product of these offsetting effects. A partial equilibrium model with heterogeneity in risk aversion

illustrates formally the multiple channels that affect this joint decision. Leaving aside for simplicity the two margins (c) above, the key idea that emerges is that risky sex and testing are complement goods: if new treatments lower the cost of being infected, they increase the incentive to test. This is because testing is necessary to access treatments. To summarize, the model predicts that the more risk-averse agents will choose not to test either before or after the release of ARVs, and will decrease their risk level after the release of ARVs: for those people, the indirect effect dominates the direct effect. In contrast, the more risk-inclined agents will choose to test before and after the release of ARVs, and will increase their risk level after the release of ARVs: for those people, the direct effect dominates. Depending on the net product of the direct and indirect effects, agents at the margin may shift from lower risk and no testing to test and riskier behavior, or from higher risk and testing to lower risk and no testing: this constitutes a selection effect.

Intuitively, if we were to make a single empirical prediction, it would be that the direct effect of ARVs introduction among susceptibles is presumably first-order: in other words, ARVs should translate into an increase in risky sexual behavior. This is my first testable hypothesis.

Second, because risky sex and testing are complement goods, regardless of the direction of the selection effect, testers should take *relatively* more risks than non-testers following the introduction of ARVs. This represents a second testable hypothesis.

Furthermore, with heterogeneity in preferences for risky sex, and assuming selection effects do not matter too much, better treatments may result in a polarization of behaviors: with people who test engaging in more risky sex while those who do not test actually engaging in less risky sex. This conjecture is my third testable hypothesis.

To validate these insights, I exploit a large data set of men who have sex with men (MSM) collected in San Francisco from 1994 to 2002. The data reveal an increase in overall risky sexual behavior starting in the first semester of 1997. I use a difference method to purge the causal effect of ARVs from other potential confounding factors: as expected, Whites and gays (as opposed to bisexuals) respond to ARVs more strongly than others. In addition, testers do take relatively more risks than non-testers (as measured by their likelihood to practice unsafe sex). In fact, non-testers take fewer risks altogether (as measured by their average number of partners). I implement the first bivariate application of the Anderson (2004) polarization test to confirm this result non-parametrically. Finally, I find that the proportion of testers remains stable, which was ambiguous *a priori,* and that there is no evidence of significant migration from one group to the other. Although there is some indirect suggestion that ARVs attract some newcomers in

the MSM community, the changes that they may bring to the composition of the testing and non-testing groups do not seem to be driving the results.

II. Literature Review and Motivation

A. Background

The parameters of the HIV infection changed in 1996 when the 11th International Conference on AIDS announced major therapeutic improvements in HIV treatment. Subsequent advances made undergoing ARVs less excruciating by reducing the number of pills to be taken and side effects experienced. ARVs have reduced AIDS-related death rates by more than 80% for HIV+ people taking these drugs (CASCADE, 2003; Duggan and Evans, 2005). Furthermore, by reducing the viral load to infinitesimal (yet, still, strictly positive) proportions, ARVs decrease the rate of transmission of the disease (per coital act), conditional on unprotected sexual intercourse (Quinn *et al.*, 2000; Gray *et al.*, 2001).

Shortly after 1996, surveys documented an increase in unsafe sex in countries where ARVs had been introduced. Reported cases of STDs other than HIV confirm the pattern. In the U.S., the incidence of gonorrhea increased by 9% between 1997 and 1999. Prior to that upturn, it was declining at a rate of roughly 10% per year between 1986 and 1996 (Fox *et al.*, 2001). Although 1998-2001 marked a spike in the incidence of gonorrhea, it is now decreasing much more slowly than in previous decades. Similar trend breaks have been found in Canada, (Health Canada, 2000), in Europe (Eurosurveillance, 2002) and in Australia (Grulich, 2000).

The public health and epidemiology literature has suspected for a few years that the belief in the efficacy of ARVs increases risky sexual behavior (Crepaz, Hart and Marks, 2004), but it has failed to link the change in incentives with the issue of testing. In addition, some of these studies rely on questioning individuals directly about their views on ARVs; hence they waver between a causal path from optimism to risk and optimism about treatments as a form of *post hoc* rationalization following risky encounters (Huebner, Rebchook and Kegeles, 2004). In any event, observing an increase in risky sexual behavior after 1996 is not enough to isolate a causal effect from ARVs since the observation could, theoretically, be driven by other reasons, notably time effects, *e.g.*, the so-called "prevention fatigue", a generational effect, the decreasing impact of prevention campaigns, or any combination of such reasons.

At the same time, more effective treatments could be expected to increase the demand for testing. In the pre-ARVs era, one could excuse non-testers as potentially rational, albeit selfish individuals. Limited therapeutic options meant the lead time gained by knowing one's status did virtually nothing to lengthen

life, and simply extended the time spent worrying about dying. Yet, the impact of new treatments on the incidence of testing seems, if anything, modest (see for example the Report on HIV/AIDS in Ontario, 2003). This apparent puzzle has been neglected.

This paper examines the impact of HIV treatments from the perspective of economic epidemiology (Philipson, 2000; Gersovitz and Hammer, 2003). It extends the literature in economics estimating the impact of HIV testing or public health interventions – such as subsidies for safe sex or information campaigns – on sexual behavior (Philipson and Posner, 1994 and 1995; Geoffard and Philipson, 1996; Boozer and Philipson, 2000). More specifically, it seeks to expand upon an analysis developed in Geoffard and Mechoulan (2004): according to that paper, under the existence of ARVs, susceptible individuals who undergo testing are expected to increase risky exposures. On the other hand, the paper argued that those who do not test do not face a change in their incentives. Consistent with these arguments, Geoffard and Mechoulan (2004) found a significant increase in risky sex among testers and no significant increase among non-testers.

This previous investigation left several issues unresolved. A problem with that model is that it does not explain why some people undergo testing while others do not. It simply implies that those who do test will take more risks after the introduction of ARVs. Geoffard and Mechoulan (2004) used the relative stability of the proportion of testers as an implicit argument for treating non-testing individuals as a control group. The identification of ARVs thus hinged on a questionable foundation. Moreover, the stability in the proportion of testers is a rather counter intuitive pattern and deserves some elaboration. In particular, it stands at odds with the sharp increase in unprotected sex observed after 1996 (among those who test). In this paper, I therefore improve on that previous work by acknowledging the endogeneity of the testing decision and suggest new arguments to account for the empirical stability of testing. Also, I propose a second-best method to support the causal effect of ARVs availability beyond its identification against a flexible time trend. Finally, Geoffard and Mechoulan (2004) ignored the variability of the number of partners as a key alternative measure of sexual risk. I now analyze two margins of risk, *i.e.*, both protection and number of partners.

Another closely related paper is that by Lakdawalla, Sood and Goldman (2006) which analyzes risky sexual behavior among HIV+ individuals using access to health insurance as an instrument for treatment status. They find that treatment results in more sexual risk-taking by HIV+ adults. Insofar as my population is mostly composed of HIV- individuals, this work can be thought as a complement to theirs. Further, since they focus on HIV+ people, the dimension of the testing decision is missing in their analysis.

I qualify their conclusions on two related points. First, Lakdawalla, Sood and Goldman (2006) note that (observed) incidence of HIV increased following the introduction of ARVs. At the same time, they report that condom use among unmarried Americans increased. However, HIV incidence is still overwhelmingly fuelled by individuals belonging to high risk populations. Therefore, linking the two observations and abstracting from the segmentation of the market (in other words, assuming random mixing) tends to exaggerate the increase in unsafe behavior on the part of HIV+ individuals as the primary causation mechanism. Further, measurement of HIV incidence itself is endogenous (recall improved therapies increase the benefit of testing), making the connection even more difficult to establish. Second, regarding HIV prevalence, the suggestion that ARVs increase the spread of the disease in the general population seems premature. Lakdawalla, Sood and Goldman's (2006) calculation is based on the increase in risky activity from HIV+ individuals, after correctly accounting for the diminished risk of HIV transmission when HIV patients are under ARVs (see Blower, Gershengorn and Grant, 2000). Yet, the explicit assumption of holding the behavior of the general uninfected population constant (along both risk margins) is unreasonable: recall the authors themselves acknowledge that it is not supported in the data. Given that this element plays a first order role in the computation, their results are therefore too preliminary not to invite skepticism. Alternatively, the results of the present paper suggest that the rise in HIV incidence is more likely attributable to the deliberate choice of the high risk, susceptible population to increase its risk level.

B. Hypotheses

Theoretical hypotheses

The Appendix provides the derivation of a partial equilibrium model with heterogeneity in risk aversion to formally illustrate the multiple channels that affect the joint risk-taking/testing decision. The arguments underlying the model build on the literature that addresses the interaction between primary and secondary prevention. See Kenkel (2000) and Eeckhoudt *et al*. (2001) for details.

In the model, susceptible individuals jointly choose a level of risky exposure to HIV and decide whether they should get tested or not. Testing is costly but gives access to a better outcome should one test positive. The optimal level of risk is determined by the tradeoff between the perceived benefits of risky sex (relative to safer sex) and the expected cost to be infected, which depends on one's number of risky sexual exposures, the prevalence and disease transmission rates, and the availability/quality of treatment.

The model reveals several insights. It first shows that there exists a critical value of treatment quality such that *ceteris paribus*, testing is optimal if and only if the quality of treatment is larger than that cutoff. Second, it reproduces the

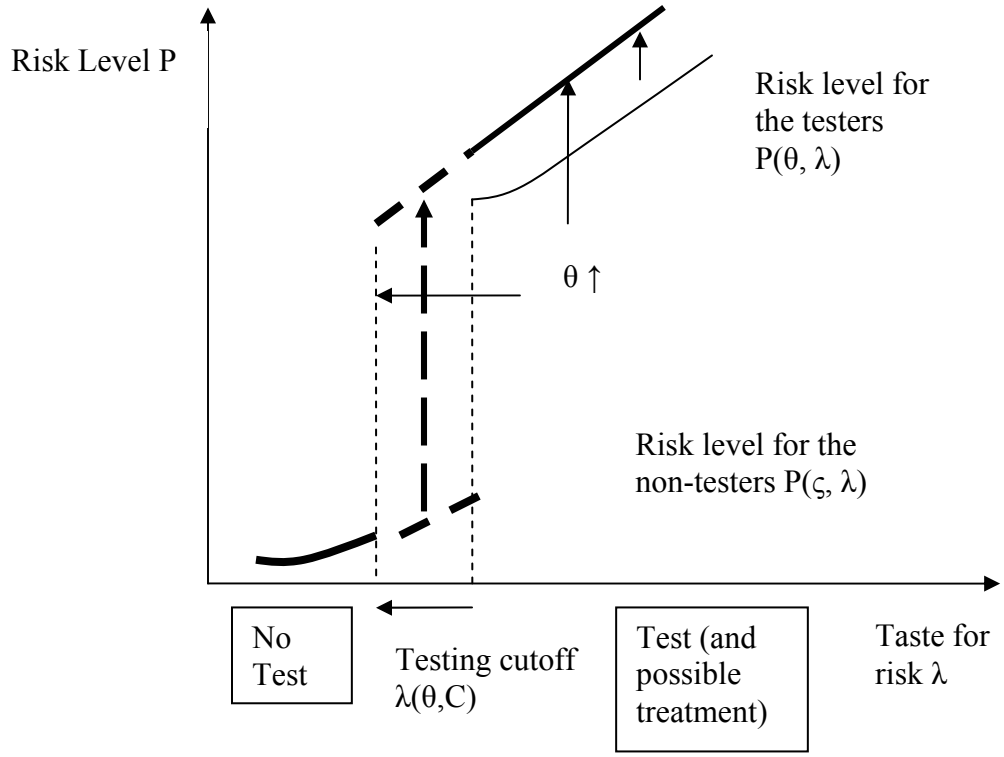
fundamental result of economic epidemiology that the optimal number of risky exposures decreases with the prevalence of the disease. However, for a given number of exposures, an increase in prevalence also directly increases the risk level. The net effect on the risk level is negative only if the demand for risky behavior is sufficiently prevalence-elastic. One implication is that the optimal behavior at the margin may be a shift from test and high risk to no test and low risk. In short, if the demand for risk is elastic enough, an increase in prevalence unambiguously decreases the risk level, and also decreases the demand for tests. Symmetrically, in the case of an inelastic demand, an increase in prevalence increases the risk level for those who test, and may even induce some individuals to shift from safer sex (and no testing) to testing and risky behavior. This is because for some individuals who adopt safer sex practices, the behavior loss from reducing their risk level even further would be too large, and outweigh the benefit of risk avoidance.

The model then considers aggregate behavior, and assumes some form of heterogeneity in risk, such that the critical value of treatment quality above which one decides to test differs across individuals. This defines a threshold of risk-tolerance above which all individuals decide to test. This threshold depends on treatment quality and disease prevalence. An increase in treatment quality lowers the threshold. However, an increase in prevalence has ambiguous effects. It increases the incentive to test for risk inclined-agents more than it does for the risk-averse ones; indeed, it may even decrease the incentive to test for those with a low enough tolerance for risk.

Then, how does one disentangle the effects of better treatments? The model helps sort out the different elements at stake, principally: direct (or substitution) effect, indirect (or equilibrium) effect and selection (or composition) effects. As mentioned earlier, the key intuition is that risky sex and testing are complement goods.

Assuming a constant prevalence, an increase in treatment quality has two effects. The first, standard one is the direct effect of new treatments on the price of risky behavior: conditional on a positive test, the consequences of infection are less severe. Hence there should be an increase in the risk level among individuals for whom testing is already optimal. Second, there is a selection effect from those individuals at the margin who switch from no testing (and low risk) to testing (and higher risk). On the other hand, no behavioral change occurs among individuals who choose not to test, since they are indifferent to treatment opportunities. This is summarized in Figure 1.

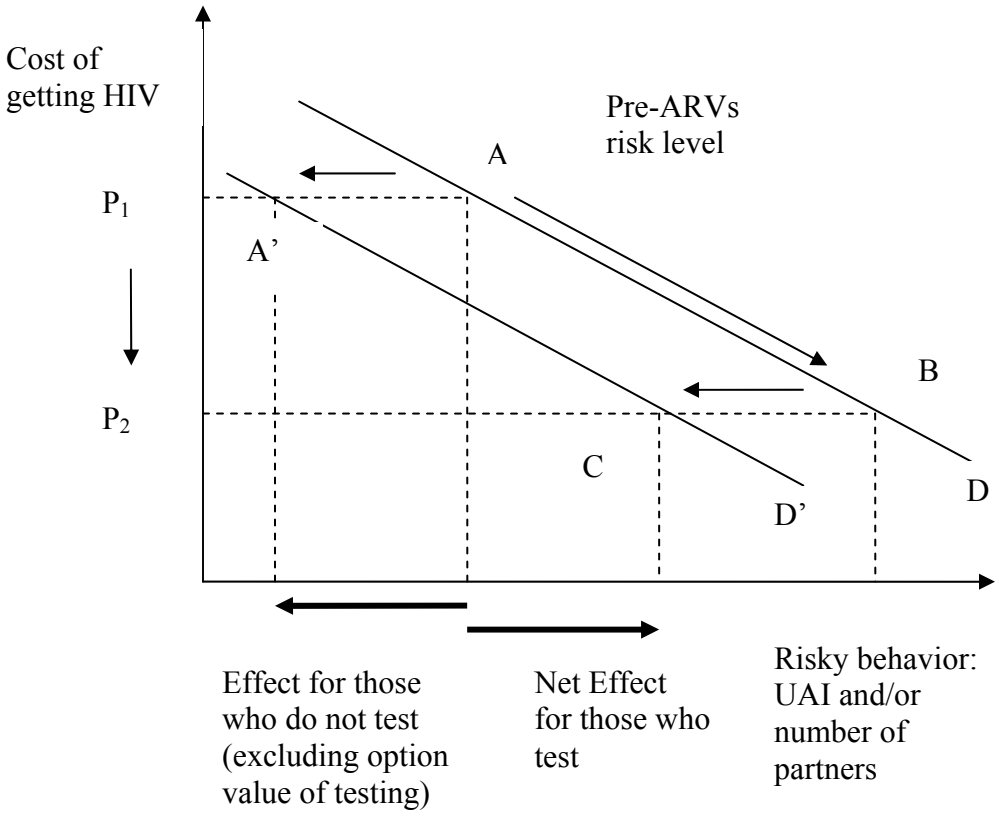
Figure 1: Hypotheses regarding the effects of ARVs on (susceptible) testers and non-testers (*excluding equilibrium effects*)



ARVs decrease the cost of getting HIV ($\theta \uparrow$) while testing cost (C) stays constant, so a larger fraction of individuals tests and increases their risk level. The new testers have a lower taste for risk than the other testers but their change in risk level is more dramatic. There are fewer non-testers, who do not change their risk level because the cost of the disease for them stays unchanged at ς (excluding option value of testing); recall $\varsigma < \theta < 0$.

The model finally studies what happens if new treatments increase prevalence: this would be the case if the dominant effect comes from healthier seropositive selfish individuals on the market for risky trades, offsetting the decrease in the transmission parameter. The susceptibles should in turn become relatively more cautious, and this would act as a counteracting force against the substitution effect. In that case, Figure 1 needs to be supplemented with Figure 2.

Figure 2: Hypotheses regarding the effects of ARVs on (susceptible) testers and non-testers (*excluding selection effects*)



HAART decreases the cost of getting HIV from P_1 to P_2. Those who decide to test take more risks (A→B). However, an increase of prevalence as a result of HAART shifts the demand for risky sex D to D' (not necessarily parallel). Accordingly those who decide not to test take fewer risks (A→A'). Those who decide to test take fewer risks than had the demand for risky sex remained constant (B→C).

Furthermore, this indirect effect also affects the selection effect. Depending on the magnitude of the equilibrium effect, the direction of the selection effect may even be reversed, *i.e.*, individuals at the margin could switch to no-testing and decrease their risk level. For those different reasons, the effect of treatments on the risk level becomes ambiguous.

However, the above discussion abstracts from the option value of treatment. This is why positing the *relative* negative response of the non-testers compared to the testers is a weaker hypothesis than positing the *absolute* negative response of the non-testers.

Empirical hypotheses

Empirically, the direct and presumably first-order effect should result in an increase in risky sexual behavior and intentions to practice unsafe sex following the introduction of ARVs.[1] This leads to:

Hypothesis (1): ARVs availability leads to a global increase in risk-taking

Second, because of the complementarities between sexual risk and test, regardless of the direction of the selection effect,[2] we can formulate the following:

Hypothesis (2): ARVs availability increases risk-taking *relatively more* among testers than among non-testers

In the limit, with enough heterogeneity in preferences for risky sex, and second-order selection effects, better treatments may result in a *polarization of behaviors* whereby individuals who test would engage in a higher amount of risky sex, while those who do not test would actually engage in less risky sex. Notice this possibility goes against the (naïve) view that no behavioral change should occur among individuals who choose not to test, the argument being that they should be indifferent to treatment opportunities. Whether that is the case or not, these people are not indifferent to the expected prevalence change induced by ARVs. This leads to:

Hypothesis (3): ARVs availability increases risk-taking among testers, decreases risk-taking among non-testers,

with the provision that risk may be measured in different ways (both protection and number of partners) to establish those hypotheses.

Recall a central assumption underlying Hypotheses (2) and (3) is that the option value of treatments when non-testing (through the possibility of a test in the future) is significantly lower than the value of immediate treatment access under a positive test. The static model cannot capture this element by construction, so it is worth exploring its implications in detail.

---

[1] Using the model notations we could say $\frac{dP(\theta, \lambda, p(\theta))}{d\theta}$ is theoretically ambiguous and the hypothesis is that the direct effect offsets the equilibrium effect, so that $\frac{dP(\theta, \lambda, p(\theta))}{d\theta} > 0$.

[2] Empirically it will be argued that $\frac{d\lambda}{d\theta} \simeq 0$.

Obviously, only testing opens access to treatment conditional on being HIV+. Yet, why would the cost of getting the disease become lower for testers given that, if contamination happened tomorrow, there would typically be no physiological benefit from knowing this for several years? A high-risk selfish individual may thus be tempted to wait for AIDS symptoms, get tested, and then fully benefit from treatments. Things are not that simple: after infection, there is no gain from HIV testing *per se* but there is a gain from regular (every 3-6 months) CD4 count testing so that the start of ARVs can be optimally timed. Therefore, waiting for diarrhea and wasting or even later for an opportunistic infection probably entails a loss of expected life years.[3] It is difficult to fathom that many non-testers would adopt such a risky strategy, let alone that they would (or could) get a regular blood test for CD4 monitoring only but not for HIV. More importantly, the preceding discussion reflects the state of our medical knowledge in 2007: in the mid to late 1990s, when physicians did not know (or neglected) HIV drug resistance, treatments began much earlier than today (the "hit early, hit hard" strategy, officially abandoned in 2000). Finally, the reasons that some individuals have not yet performed a test (stigma, optimism, delusion, etc.) indicate that they are also less likely to get tested in the future absent symptoms, and therefore less likely to fully benefit from ARVs. These factors therefore support Hypotheses (2) and (3).

These three hypotheses may be regarded as variations on the classical seatbelt argument, whereupon drivers wearing seat belts feel more secure, and drive less carefully, leading to more traffic accidents (see Peltzman, 1975). Here, susceptible individuals solve the joint problem of optimizing testing decision and risk levels, and adjust that risk level along two margins (number of partners and protected sex) instead of just one as in the driving case. Note that I limit my analysis to behavioral changes among susceptibles. The overall impact of treatments on HIV prevalence (the analog of traffic fatalities) is even more complex to assess and would require richer data.

Further, HIV prevalence among those who do not test is, by definition, unobserved. Instead, what is observed is the prevalence among those who get tested (*i.e.*, the proportion of positive tests). Recall that new treatments may lead some previously low risk individuals to test. The risk level of these individuals, who are presumably at the margin, would be lower than the average risk level among the whole population of testers. Therefore, changes in the observed proportion of positive tests may provide a downwardly biased estimate of changes

[3] ARVs seem to offer the best clinical response when it is started at CD4 counts at just under 350. Those who wait until CD4 falls below 200, never seem to regain the full function of their immune system. Within the next few years epidemiologists will have collected enough cohort data to estimate the percentage reduction in survival entailed by the decision by an asymptomatic patient to delay HIV testing until symptom onset.

in actual risky behavior. Combine this possibility with the reduction in infectiousness, and we can see that even if the proportion of positive tests decreases empirically, actual risky behavior may have increased. This is a phenomenon I actually observe in the present data set over 1997-2000.

III. Data

A. Sources

My data come from the Stop Aids Project (SAP). This San Francisco-based agency has gathered information on populations at risk for HIV since 1994. Precisely, the SAP data set is made of non-repeated cross sections and is principally directed towards MSM.

The sampling frame is as follows. The SAP volunteers randomly intercept men on the streets of gay districts and participants of a variety of gay-oriented venues (*i.e.*, bars, clubs, events, parades, street fairs) in the course of outreach education.[4] The interviewees then respond to a peer-administered, detailed questionnaire. Some people may have been sampled more than once and if so, it is indicated (variable "Ever answered"). They are excluded if they have participated previously during the same year. Data are collected throughout the year and each interview has a record of the exact time when it was made.[5]

My collection runs from April 1994 to June 2002. Since it goes back as far as 1994 before the introduction of ARVs, it is essential in identifying treatment effects on behavior.[6] For the complete years, the number of individual records ranges from 9,942 in 1997 to 2,657 in 2001. Overall, the data base contains answers from 48,888 interviews.

The data are rich with questions pertaining to sexual practices in the six months prior to the interview: number of partners, type of intercourse, whether a condom was always used, etc.; although regrettably, the formulation of the

[4] Note that individuals who like going to bars, for example, will certainly be over-represented, while individuals who are in the late-stages of HIV (or any other debilitating disease for that matter) will be under-represented.

[5] Note that since these are not longitudinal data, I cannot track the same individuals over time; accordingly I can only observe changes in the composition of the testing and non-testing group through their average characteristics.

[6] Such information, with a large sample size on a high risk population, and collected consistently over such a period, is exceptional. The Multicenter AIDS Cohort Study, in comparison, has fewer than 1,500 observations in the critical mid 1990s period. The SAP data have been used in the public health literature over the past several years and their validity is well documented.

questionnaire does not allow to know the frequency of risk taking among those who declare not always using protection, let alone the nature of the risk taken for each of the sexual encounters. On the other hand, the questionnaire is relatively poor with socio-demographic variables. Only age[7], ethnicity, and zip code are available throughout. That is why I merged the SAP data with the STBF3 file of the U.S. Census to obtain information on median education and income in the SAP respondents' zip codes.[8] Additionally, respondents were asked if they ever attended a Stop AIDS meeting (variable "Ever attended").

A key problem is that the question regarding HIV status was not introduced until the middle of 1997. Thus, I cannot compare the behaviors of HIV+ persons before and after the advent of the new treatments although it is plausible that behavioral adjustments to therapeutic improvements differ by serological status. Epidemiological studies contend that this is not the case (Crepaz, Hart and Marks, 2004).[9] What I can observe in the SAP data set after June 1997 supports that assertion. HIV+ respondents, although more likely to practice unsafe sex in absolute terms (44% vs. 32% on average), do not do so relatively more than HIV- individuals over time (+1.2% vs. +1.6% per year on average). The data on number of partners reveal a similar pattern: HIV+ respondents have a higher number of partners (14.7 vs. 9 on average) yet increase their average number of partners at a slower rate: +0.6% vs. +2.4% per year on average. "Only" 14% of individuals interviewed after June 1997 who got tested declare they are HIV+, and the proportion is 13% out of all interviewees. This proportion is stable over 1997-2002. The population is therefore fairly homogenous with respect to HIV status.[10]

---

[7] I kept the two observations with age=99 because I have seven others with age ranging from early to late nineties, thus it may not code for non-response. With less than 1% of the sample over 60 years old deleting any fraction of the oldest respondents would not affect the results in any significant way.

[8] Because the Census does not cover all zip codes, some individuals do not have observations on median household income or education in their zip code.

[9] Theoretically, this may be justified by the fact that "superinfection" (reinfection by a second strain of virus), or secondary infections (such as syphilis), are possible and may lead to disease progression. Lakdawalla, Sood and Goldman (2006) found that HAART provision (instrumented by Medicaid eligibility rules) had no significant effect on the frequency with which seropositive individuals practiced safe sex.

[10] However, this may not necessary reflect the true HIV prevalence in the MSM community of the San Francisco area which was been estimated to be closer to 20% in the late 1990s (Catania *et al*., 2001).

The benchmark empirical estimations ignore this corruption problem. Still, to check whether it is serious, I perform a two-step sensitivity analysis. First, I simply remove those individuals who are known to be HIV+ from mid-1997 on. In this case, among the testers, I have a mix of HIV+/HIV- before July 1997 and HIV- individuals exclusively afterwards. Assuming the same proportion of HIV+ before July 1997, this leaves now 6% of unidentified, HIV+ respondents aware of their status in the sample. Second, I impute the missing values on HIV status for 1994-mid 1997 based on the exogenous characteristics of HIV+ that are almost always available throughout (Age, Age^2, whether White, whether resident of the Bay area, whether resident of California, median household income in zip code, median education in zip code).[11] I thus construct a counterfactual sample of non-testers and susceptible testers by removing the 14% known HIV+-testers from July 1997 on, and the 14% "most likely HIV+"-testers before July 1997, based on imputed HIV status.

Another important caveat is that I restrict attention to sexual behavior.[12] I cannot analyze change in intravenous drug use risk, and therefore miss an important risk factor of HIV transmission. Further, the data set only captures attitudes and behaviors of a subset of adults affected by HIV risk and infection, namely MSM.

An ideal data set would be longitudinal in nature while allowing for new entrants into the survey, therefore enabling the researcher to distinguish who changes their behavior over time. It would contain a larger set of risk factors (including a distinction between insertive vs. receptive anal sex, number of unprotected sexual encounters, condom use with persons whose HIV serostatus was unknown or known to be positive), information on HIV status throughout, and more socio-demographics, including health insurance coverage. The sampling would capture a more representative fraction of the population at risk for HIV, targeting beyond those clearly visible or active in the gay community,[13] perhaps through telephone interviews. Nevertheless, I believe that the new insights that can be collected from the present data outweigh the limitations of the design.

---

[11] I have also checked that these predictors of seropositivity do not change over time.

[12] Such a sensitive topic might result in a higher proportion of intentional reporting error than in a typical survey. Be that as it may, the proportion of implausible or inconsistent responses (people who declare being seropositive without ever testing for example) is very low.

[13] The location of the interview is recorded for almost all observations. At least two thirds of the interviews took place in or around the Castro district. Lack of familiarity with the area prevents me from identifying the rest of the venues with certainty.

B. Summary Statistics

Table 1a provides the descriptive statistics of the sample with definitions of all the variables. The main variables are defined as follows. "After" is the dummy variable corresponding to the introduction of ARVs: After =1 starting from 1997, 0 otherwise. Although the ARVs are phased in during 1996 (especially in the second semester), recall questions concerning sexual practices are retrospective of the last six months.[14] "Test" is a dummy indicating whether the individual declares "knowing his serological status" (before February 1997) or "performed an HIV test" (after January 1997).[15] The consistency of the answers before and after the change of formulation, which can be checked on a daily basis, has convinced me that I am in fact dealing with the same question, which defines who belongs to the "test group." Regrettably, there is about 18% non-response for this question, but most of it (in fact more than 80%) comes from a narrow period: in November and December 1996 and from February to June 1997 the question relative to testing disappears from the survey.

A large majority of the respondents are testers.[16] It is critical to understand that almost all non-testers have never performed an HIV test: in other words, they do not declare non-testing because they already know that they are HIV+ and are therefore in absorbing state. From the question on sero-status result introduced in 1997, I checked that 98% of non-testers declare not knowing their status, whereas 98% of testers declare knowing it, 14% of them declaring being HIV+.[17] Therefore, the corruption of the sample for the testers (where the HIV+ are undistinguishable from the HIV- in the first three years of the survey) does not apply for the non-testers.

---

[14] All results on unsafe sex are robust to coding After=1 starting October 1996; on the other hand, false experiments yield smaller and less significant point estimates as the coding of the structural break moves away from the late1996/ early 1997 period - regressions available upon request.

[15] A question regarding how much time elapsed since the most recent test was introduced in 1997: 73% of respondents who answer yes to "Have you had an HIV test?" did have a test less than a year before the interview, 82% within two years or less, etc. Therefore, having respondents who will always appear in the testing group, even though their HIV status information is too old to be relevant does not seem to be not a major problem.

[16] It is possible that some non-testers are people who recently became sexually active in the gay community: the non-testers are slightly younger (31 vs. 34 on average for the testers) and the proportion declaring zero or one partner is slightly higher (38% vs. 34% for the testers). It appears that this argument is actually more plausible after 1996, and its possible implications are analyzed later. I thank an anonymous referee for this observation.

[17] The variable Test Result presented in Table 1a is a binary recode from an original question with three possible answers: "HIV+", "HIV-", and "Do not know".

Those variables measuring sexual activity in the six months prior to the interview are "numbers" (number of partners)[18], "Anal intercourse" (at least one case of anal intercourse, henceforth AI), "Anal intercourse no condom" (at least one case where no condom was used), and "Vaginal Sex" (at least one case of vaginal sex). I use Vaginal Sex as a proxy for sexual orientation (*i.e.*, bisexual vs. gay). Finally, the question "Do you intend to practice safe sex?" was dropped after January 1997 but resumed in the second half of 2000.

Because as mentioned earlier, we neither know the frequency nor the nature of risk taking among those who declare not always using protection, my empirical analysis deals with a more conservative (if not more robust) indicator. I recombined variables "Anal intercourse" and "Anal intercourse no condom" into a variable indicating whether the individual declares exposing himself to some AI risk in the last sixth months. Therefore my main variable of interest is risk on the intensive margin: UAI. This dummy takes value one if the individual declares any unprotected AI in the last six months. Considering the extensive margin of risk, the number of partners does not mean only the number of male partners with AI, *i.e.*, the riskiest kind of sex. It is therefore an imperfect measure of risk. Yet it turns out that, not surprisingly, the higher the number of partners, the higher the chance that the respondent declares some UAI.

It is also useful to consider a synthetic indicator of risk that combines intensive and extensive margins. I construct $\text{Syn} = 1_{\{\text{Numbers} < 2\}} \times (1 - \text{UAI})$, where $1_{\{\text{Numbers} < 2\}}$ is an indicator for whether the individual declares fewer than two partners in the last six months. In other words, Syn = 1 if the individual takes no risks on either margin, 0 otherwise.

IV. Empirical Analysis

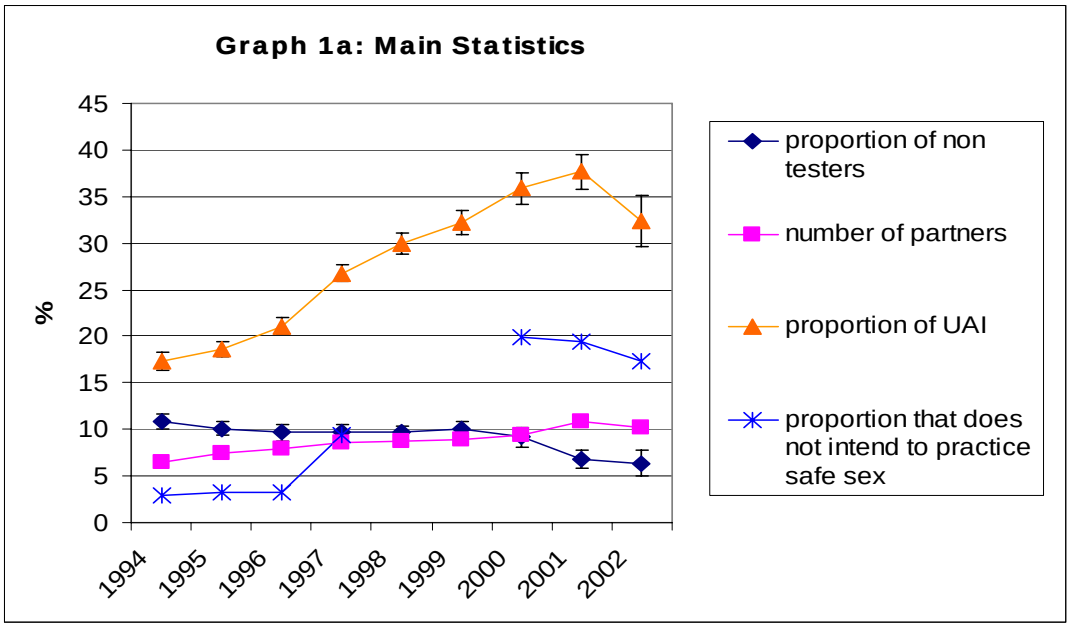
A. Data Patterns

Table 1b provides some average characteristics for the whole sample and for the main five digit zip codes for each year, enabling us to see composition changes over time. The most visible one is perhaps the declining proportion of Caucasians over time. I also note a small increase in the proportion of Bay area residents in 2001-2002. This may reflect labor market conditions following the recession.

The following diagrams illustrate some of the most interesting features conveyed by the raw data. Whenever informative, I report 95% confidence intervals around the means. Graph 1a summarizes the evolution of the main

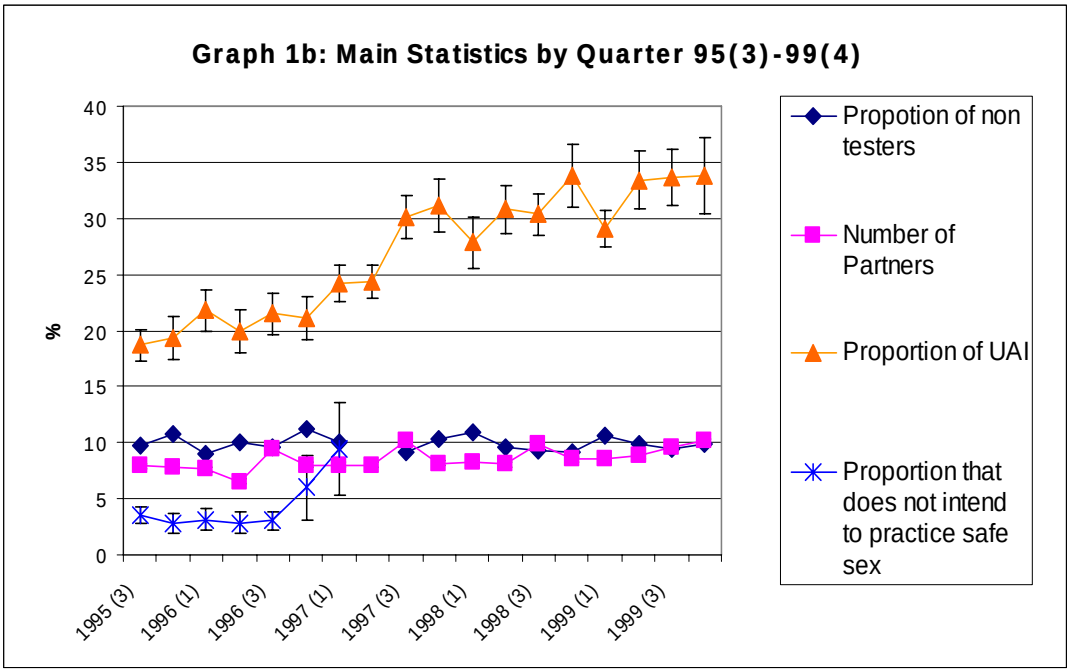
[18] The variable numbers (number of partners) is censored at 999. Since nobody else declares a number in the 800-900 range I ignored such responses. However, highly promiscuous respondents make only a tiny fraction of the resulting sample: 0.62% declares a number of partners greater than 100 over six months.

variables. The most striking patterns are the increase in UAI over the period (see Page-Shafer *et al*., 1999; Chen *et al*., 2002), together with the increase in the proportion of individuals who do not intend to practice safe sex. These two features of the graph provide strong support for Hypothesis (1).[19] At the same time, testing odds remain fairly constant at a high level (88 to 93%). Recall that I only have the first semester of 2002. On top of a smaller sample size than in previous years there are seasonality issues (on average, I find that individuals take more risks in the second semester), so any break observed that year — for this graph and the following ones — should be interpreted with caution.



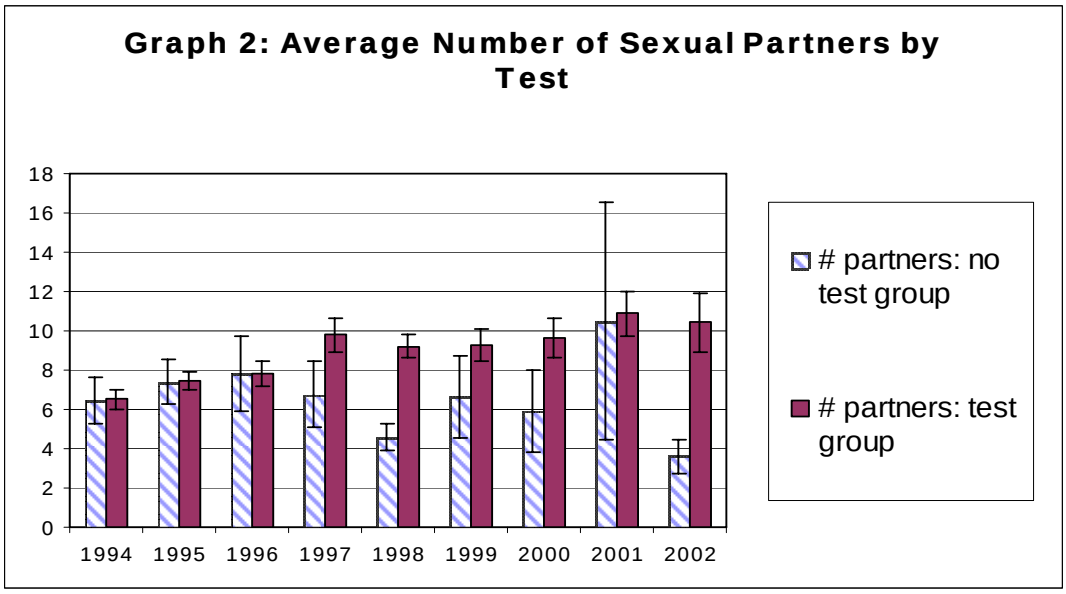
Since I know the day when each data point is collected, I also looked at those patterns by quarter to check more precisely when the break in UAI occurs: it clearly takes place mid 1997. Note the consistency between the declaration of intentions in 1996 (4) / 1997 (1) and the six month-retrospective data on UAI in the course of 1997.

[19] I also uncovered that the perception of risk from peers increases sharply between mid 1994-1996 and January 1997, after which data collection on this topic stops. This further supports Hypothesis (1).

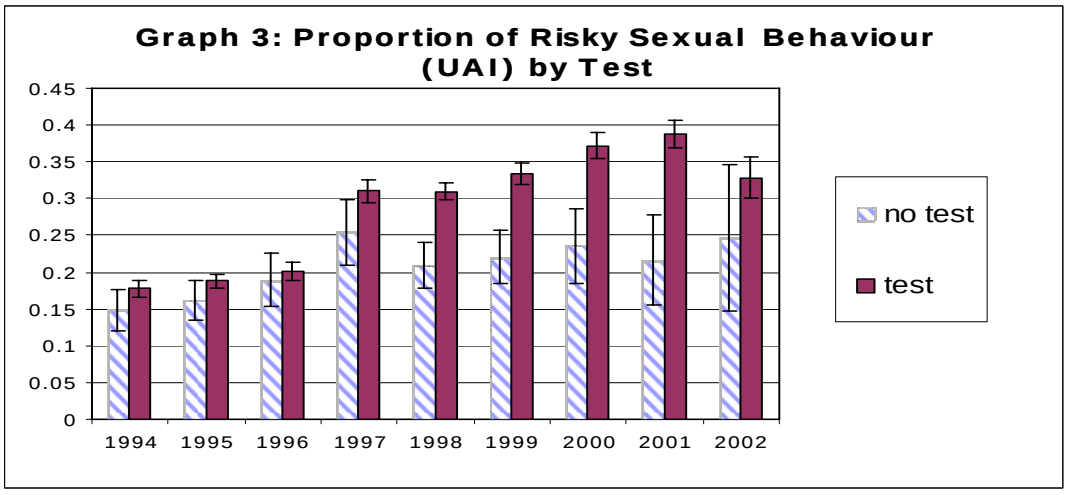


I start with the analysis of risk taking along the two dimensions of number of partners and UAI. I first compare the evolution of the number of partners for those who test and those who do not. Graph 2 illustrates a key element: the difference in number of partners between those who test and those who do not test increases sharply after 1996.[20] The feature supports at least Hypothesis (2) and to a lesser extent Hypothesis (3).

[20] Performing the same analysis by quarter, I verified that the break occurs no earlier than 1997.



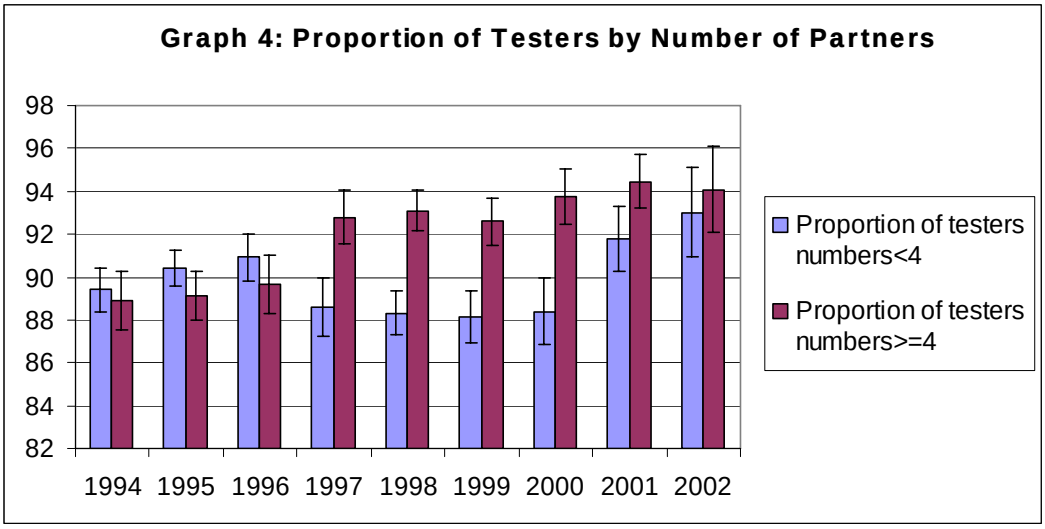
When looking at the proportion of UAI by testing category (Graph 3), I note that from 1997 on, the difference in the probability of UAI increases discretely between testers and non-testers.[21] This significant and dramatic increase in UAI, specifically after 1996 among those who test, further supports Hypothesis (2). On the other hand, the proportion of UAI among those who do not test increases as well, but almost linearly, and to a smaller extent.



[21] Again, I checked that there is no graphical evidence for a break in the data before 1997.

As for the proportion of testers, most noticeably I have just shown that it increases overall only in the last two years of the survey.[22] Since P(Test) is relatively flat over time, it comes as no surprise (given Bayes' rule) that P(Risk | Test) and P(Test | Risk) show the same time-series patterns: indeed, the difference in the proportions of testers between the two groups UAI=0 and UAI=1 which was decreasing between 1994 and 1996 widens significantly afterwards. However, one should bear in mind that the changes in the proportion of testers are small.

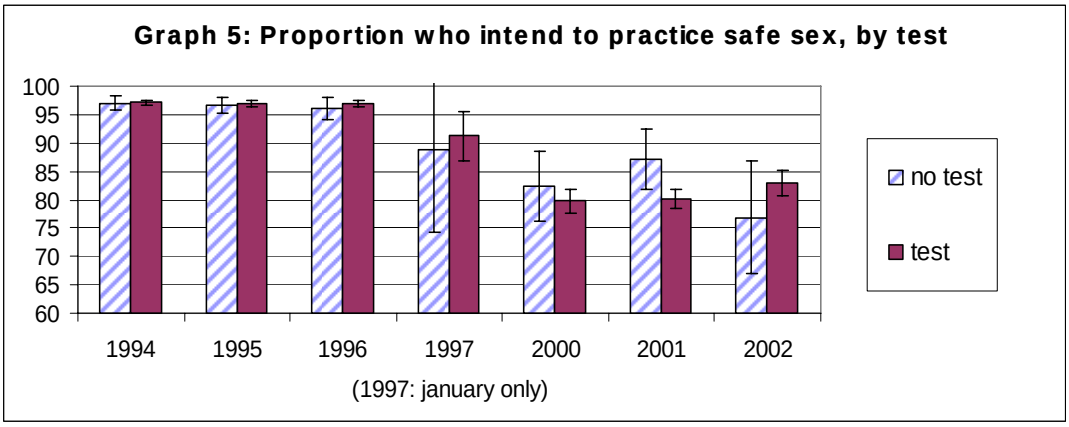
The proportion of testers by number of partners (Graph 4) reveals a more spectacular pattern. Among those with more than three partners (heuristically the most revealing cutoff when I characterize the behavior of high-risk individuals, but other presentations would confirm the pattern) the proportion of testers increases strikingly after 1996 when an important reversal occurs: from then on, testing is significantly more frequent in the multi-partner group than among its complement. This feature indirectly supports Hypothesis (3).



Regarding intentions (Graphs 5), one can already see a change in 1997 (from the observation of January alone). Looking at the mid-2000 to mid-2002 data, the decreased intentions to practice safe sex are consistent with the actual decrease in safe sex over the period. Regrettably, it is hard to tell whether the decrease in

[22] This discontinuity is puzzling. However, I note that a similar increase in the proportion of testers is observed in Ontario in those two years (Report on HIV/AIDS in Ontario 2003) and is unexplained by the principal investigator responsible for the data collection. A sharp increase has been signaled in France as well after 2000.

intention to practice safe sex appears more pronounced among those who declare testing, which could have been another argument in support of Hypothesis (2).



Those patterns are the motivation for the econometric analysis that follows.

B. Econometric Analysis

Before testing my three hypotheses, I first verify whether ARVs significantly influence the testing decision. My empirical check is whether I detect a break in the overall proportion of testers controlling for a linear-quadratic time trend and socio-demographic characteristics, in the form of linear probability models.[23]

$$Test = \alpha + \beta After + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\ Characteristics_i + \varepsilon \qquad (1)$$

It has been shown earlier that testing is relatively stable over the 1994-2000 period: it seems to increase over 1994-96, to decrease over 1997-2000 and to increase again after 2000. The different specifications presented in Table 2 confirm this initial view. Table 2 tells us that the coefficient of After, if anything, is actually negative.

A closer look at the stable proportion of testers following the introduction of ARVs reveals that it is driven down by those with no UAI, which was to be expected. This pattern is consistent with the proposition of complementarities between risk and testing. What is more puzzling is the non-significance of After when regressing Test over the sub sample where UAI=1 exclusively: if we set aside the fact that so many of them are testers in the first place, one could expect risk takers to be still more likely to test if testing becomes more valuable. According to the preceding discussion, one explanation for the stability of testing

[23] See Ai and Norton (2003) who discuss problems with interaction terms (which appear in Tables 4 and 5) in nonlinear models.

after 1996 could be that the prevalence effect offsets the direct incentive to get tested.

However, I do not observe a significant increase in prevalence in the sample from the time this information becomes available on (even though, as mentioned earlier, observed prevalence may be a downwardly biased predictor of actual prevalence). Further, the same way the fraction of positive tests does not increase, what I may call practical prevalence — the probability of meeting an HIV+ partner up for risky sex at random — likely does not increase either: recall that, if anything, the risk level, measured by UAI and number of partners, increases faster among those who test negative from mid 1997 on. Hence the need for alternative hypotheses explaining why we do not see a significant increase in testing following the advent of ARVs. For example, as first suggested in the introduction, partners may have become less strict or demanding. The less deadly the disease is, the less the need to show a negative test as proof of lower risk. Finally, mistaken perceptions of prevalence may have overestimated the true risk level.

The finding that the proportion of testers is approximately stable led Geoffard and Mechoulan (2004) to conveniently consider non-testers a control group. Table 3 presents an analysis of UAI within the testing and non-testing groups where I estimate the following model:

$$UAI = \alpha + \beta After + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\ Characteristics_i + \varepsilon \qquad (2)$$

The first two columns refine and confirm the key result in Geoffard and Mechoulan (2004) while offering a counterpart to Graph 3. Column (3) runs the same regression on the sub sample of testers, with observable HIV+ removed (from July 1997 on). Column (4) does the same on the counterfactual sub sample of susceptible testers based on the removal of observable HIV+ and of the unobservable, most likely HIV+ given the imputation method described earlier. It is comforting to see that estimates from columns (2), (3) and (4) are not statistically different from each other and significantly different from zero. This offers a reasonable indication that the contamination problem is not too serious. Finally, column (5) runs the regression on the entire sample and finds a significant positive effect, which was expected given the small weight of the non-testers in the population. To summarize, the increase in UAI occurs within the testing group (as expected) but not in the non-testing group (where the effect should be at least smaller). This finding constitutes a validation of Hypothesis (2). Yet, if Table 3 is to present conclusive evidence of the causal impact of ARVs on behavior in the entire population – Hypothesis (1) – the exercise is not entirely convincing because testing is in principle endogenous.

The main identification challenge is the possibility that, despite a flexible time trend, the before/after treatments dummy variable may pick not only the new

treatments but other elements that change over precisely that same period.[24] A flexible time trend buys some identification power because the usual factors of sexual risk among MSM are known to show a high level of persistence at the individual level. The public health literature teaches us that high-risk behavior among MSM is most commonly associated with the expression of sexual identity and sexual addiction, "sensation-seeking" personality traits, depression (hence desire to escape from the reality of HIV), recreational drug consumption (such as methamphetamine or nitrite inhalants, *i.e.*, "poppers"), sero-sorting, or safer sex / prevention fatigue. It follows that, on average, the determinants of sexual activity for a given MSM vary slowly over time. However, an important confounder of any trend here is the migration of individuals in and out of San Francisco (and in and out of the gay bar scene).[25] In fact the data show that, relative to a linear-quadratic time trend, Bay area residents are more represented after 1996. Consequently, a flexible time trend may not be enough to capture most of the secular changes in risk-taking. Nevertheless, if the significance of a break in the data "survives" the inclusion of such controls, it should at least point toward a major change, such as the announcement that the cause of AIDS is sexually transmittable.

This being said, I cannot exploit the same variations in treatment access as in Lakdawalla, Sood and Goldman (2006). However, it is intuitive that certain groups of individuals should respond to treatment availability more strongly than others. First, an abundant public health literature documents that Whites have better access to treatments, and are more likely to seek treatments and to receive better care than Blacks and Hispanics in the U.S.[26] The medical literature further reveals that, for genetic reasons, Blacks are less responsive to ARVs.[27] Consequently, among Black and Hispanic men, the death rate from AIDS is significantly greater than for Whites. The behavioral response of Whites should

[24] To my knowledge, the only other element that meets those criteria and may have affected sexual behavior is a significant expansion of the Californian AIDS Drug Assistance Program (ADAP) — itself a consequence of ARVs. Therefore, my main indicator captures not only the discovery of new treatments but possibly also their availability through extended insurance coverage, which does not affect the main argument of the analysis.

[25] I thank an anonymous referee for making this point.

[26] See for example Gebo *et al*. (2005), Moore *et al*. (1994), Kass *et al*. (1999), Kahn *et al*. (2002), Morin *et al*. (2002), Villarosa (2004), Lopez-Quintero *et al*. (2005), Campo *et al*. (2005). I have also uncovered some evidence that Asian gays are sexually less risk-taking than Whites (Van de Ven, Mao and Prestage, 2004).

[27] See for example http://www.thebody.com/tpan/novdec_05/afam_health.html (Malebranche, 2005)

therefore be stronger than that of non-Whites. Similarly, "pure" homosexuals or gays, by definition, do not substitute sexual activity toward women, whereas other MSM – roughly speaking opportunistic homosexuals or bisexuals – may do so (see Posner, 1992).[28] I thus form the hypothesis that gay men will be taking relatively more UAI risk following ARVs compared to bisexuals.[29] Recall I do know throughout if a respondent has had sex with a woman in the last six months. In the data, I therefore expect those who do not have sex with women to respond more strongly to ARVs than those who do. The proportion of respondents who have sex with women is stable at 9% before and after 1996 which supports the idea that sex with women is exogenous with respect to ARVs.[30]

I first checked that simple uncontrolled difference in differences pre-1997/post 1996 between Whites and other groups, and between those who recently had sex with a woman and those who had not, support my hypotheses. I then defined two difference estimators: White × After, and Vaginal sex × After. The idea behind the interaction term White × After, for example, is to capture the differential impact of being White after 1996 – controlling for being White, the before/after 1996 effect, other individual characteristics (Bay area resident, age etc.), and a flexible time trend. The same idea applies for the interaction term: Vaginal sex × After.

Yet, this is not a standard difference-in-differences exercise: Whites are not a perfect control group for non-Whites; those who recently had sex with a woman are not a perfect control group for those who had not, even though being White or having sex with women does not significantly influence the probability of UAI in the pre-ARVs era. Rather, I am evaluating treatment intensities in groups for which differences that may affect UAI need be controlled for. To best cope with this problem, I obviously control for the before/after 1996 effect, individual characteristics and a flexible time trend, but also add all these controls interacted with the race and sexual orientation variables. To take again the first source of variation, the interpretation of the interaction term coefficient White × After becomes then the pure differential effect of ARVs between Whites and non-

[28] Francis (2005) explores this substitution effect empirically in the context of the appearance of HIV in the 1980s.

[29] An analogy would be individuals who have access to the stock market and the bond market compared to others who are constrained to only invest in stocks (assumed to be the riskier asset). A decrease in the relative risk of stocks should lead to more of an increase in stock buying in the constrained group than in the unconstrained group.

[30] The result is confirmed when regressing Vaginal sex on the other exogenous variables: the coefficients of After and of the linear-quadratic time trend are non-significant individually and jointly.

Whites. The coefficient of the variable After becomes mechanically what it is when the sample is restricted to the non-White group only. Note the particular importance of adding the variables White × Time and White × Time^2: these interaction terms capture the secular differential impact of being White over the whole time frame and enable me to purge the estimate I am focusing on from any smooth trend in differential effects.

I provide those estimates in Table 4a and 4b. For example, the preferred specification for the first model is of the form:

$$UAI = \alpha + \beta After + + \phi After \times White + \gamma Time + \delta Time^2 + \eta Time \times White + \kappa Time^2 \times White + \sum_i \lambda_i Individual\ Characteristics_i + \sum_i \mu_i \{Individual\ Characteristics_i \times White\} + \varepsilon \quad (3)$$

Estimates from Table 4a and 4b support Hypothesis (1). In both cases, column (1) presents the benchmark estimation with no interaction term, and column (2) adds the main interaction term only. Recall that the identification of the preferred model relies on the assumption that differences in unobservables between the different groups can be ignored.[31] At least, when adding different interaction terms using the observables in columns (3) and (4), the coefficients are statistically significant and not different from the coefficient in column (2), which suggests that unobserved heterogeneity may not be harmful in this case. As in the previous Table, models (5) and (6) address the problematic presence of HIV+ testers. Estimates from columns (4), (5) and (6) are not statistically different from each other and significant, which constitutes yet another indication that the main results are robust to the contamination problem. Thus the weight of the cumulative evidence, from two distinct sources of variations, supports Hypothesis (1), namely a causal impact of ARVs on risk, as measured here by UAI.

I then move the investigation to the least intuitive conjecture, Hypothesis (3), *e.g.*, the possibility of a polarization of risky behavior between testers and non-testers. While I have established that testers take more risks, I cannot assert that non-testers take fewer risks when looking at the UAI dimension alone. Still, I follow up on this idea with the study of number of partners by test group which I previewed in Graph 2.

Table 5 presents an analysis of number of partners and the previously defined synthetic indicator of risk within the testing and non-testing groups where I estimate the following models:

[31] See Meyer (1995) for a comprehensive discussion. However, note that even if White × After picked a confounding factor such as the differential effect of being wealthy after 1996 (despite the control for median zip code income), it would merely change the interpretation of the result, not the result itself. Indeed, I expect wealthier people to receive better medical care. So at the margin, if they can expect more benefits from ARVs, they should take more risks.

$$Numbers = \alpha + \beta After + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\ Characteristics_i + \varepsilon \quad (7)$$

$$Syn = \alpha + \beta After + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\ Characteristics_i + \varepsilon \quad (8)$$

The estimates in Table 5 confirm a decrease in the number of partners among non-testers. No significant effect is found among testers. Models (1)-(4) analyze changes in number of partners while models (5)-(8) analyze the composite indicator of risk Syn. Models (1) and (5) support the idea that non-testers adopt safer attitudes following ARVs.[32] From models (2)-(4), I can rule out the possibility that those who test have increased their number of partners because of the large proportion of testers before and after 1996, leaving little room for composition changes within that group: since there are 88-90% of testers initially and the proportion of testers remains stable (except for a slight change in 2001-2), at the very most I only have 11-13% of new members in the test group after 1996 and I would need an equal number of former testers becoming non-testers, which is unlikely, as will be argued shortly. In contrast, the interpretation of behavior changes among non-testers is more difficult than among testers. The composition of the non testing group could be, in principle, vastly different before and after 1996.

In fact, given the ambiguity of the selection effect in the theoretical model, one could interpret the results in Table 5 in three different and non-exclusive ways, which would be equally consistent with the observation of a larger proportion of testers among those with multiple partners as well as a larger proportion of non-testers among those with few or zero partners (Graph 5). The interpretations are the following: (1) those who did not test and had many partners have moved to the test group; or (2) those who tested and had few or zero partners have moved to the no-test group; or (3) those who do not test decrease their number of partners.

It appears least plausible (and impossible in the theoretical model) that the first and second effects forming a double movement, in opposite directions, are leaving the proportion of testers almost intact while replacing a large portion of the population of non-testers by former testers. In the following, I present some arguments against the first two interpretations, based on the groups average characteristics, with uncontrolled difference-in-difference estimates and corresponding p-values for the before/ after 1996 comparison in parentheses.

[32] I obtained qualitatively similar results when analyzing the dependent variables "whether four or more partners", "whether fewer than two partners," "number of partners conditional on that number being greater than one," etc.

When the composition changes between the two groups are in fact statistically significant, they appear to be small in magnitude.[33]

For example, the proportion of men who have sex with women is 13% in the no-test group before and after 1996, it is 8% before and after 1996 in the test group (DD=0.01; p-value = 0.63). The proportion of California residents is 71% before and 72% after 1996 in the no-test group; it is 78% in the test group before 1996, and 80% afterwards (DD=0.01; p-value = 0.66). The proportion of Bay area residents is 62% before 1996, in the no-test group, and 60% afterwards; it is 69% in the test group before 1996 and 71% afterwards (DD=0.04; p-value = 0.004 – most of the DD comes from the period 2000-2002 – and in both cases the testers have lived in San Francisco for a longer period of time – DD= -0.83; p-value = 0.09). Respondents in the no-test group are younger in both periods – by two years before 1996, and by four years afterwards (DD=1.83; p-value $\approx$ 0). They are also less likely to be White – by 2% before 1996, and by 8% afterwards (DD=0.06; p-value $\approx$ 0).

Investigating further these composition changes, Table 6 shows how the average characteristics of the two groups evolve over time. Most noticeably, we observe a drop in the proportion of Whites among non-testers in 1997, together with an increase in the proportion of young adults.[34] Since the proportion of (non) testers is approximately stable, especially up to 2000, one possibility to account for this puzzling pattern is to invoke sampling variability. Indeed, there are only a few hundreds observations per year in the non-test group (and only a few dozens in 2002). Another, more speculative explanation, is that more efficient drugs bring new individuals to the MSM market altogether.[35] Within that hypothesis, we could have again three scenarios. Assuming a stock of potential MSM participants crosses the line, such individuals could enter the test and non-test groups according to the preexisting proportions. Alternatively, these newcomers may select themselves predominantly into one group, while some members of that same group, in roughly equal amounts, would join the other group. While it is impossible to firmly decide between those scenarios, the fact

[33] Note that the differences with the test group become attenuated when removing the known (and imputed) HIV+ from the analysis. But since most of the variation comes from the non-test group, the argument offers little to account for the changes in difference we observe.

[34] Given that younger cohorts are less likely to be White, the two features are consistent: the DD is indeed driven by those under age 30. Additionally, these features support the idea that following a decrease in the price of risk, those who start taking risks are those who could not afford them previously.

[35] This is akin to the sexual orientation margin Francis (2005) examines in the context of the HIV epidemic.

that both groups become less White and younger in 1997 pleads against a simple transfer from one group to the other.

To the extent that this selection into the MSM group may be indeed factual, the key question is whether this phenomenon could be driving the results, especially in the non-test group. Although we cannot track respondents individually over time, we can make conjectures based on who is more or less likely to have recently joined the MSM community. First, I checked that neither the younger nor the non-Whites respondents are responsible for the outcomes within the non-test group. Second, the results are still robust to excluding those individuals with zero (or even those with less than two) partners. Most importantly, if we select the sample to be those who declare having answered the questionnaire before or have attended a Stop AIDS meeting in the past, the results are still statistically significant in each case (and even more so when combined).

If we thus interpret the cumulative evidence by viewing potential variations in the composition of the populations of testers and non-testers as second order or irrelevant, the most satisfying interpretation is still that ARVs availability decreases risk-taking among non-testers. In other words, Table 5 (combined with Table 3) supports Hypothesis (3). Again, even if (observed) prevalence does not increase, the perceptions of a higher risk level may have convinced the non-testers to decrease their exposures.

Finally, I propose a bi-dimensional application of the non parametric polarization test created by Anderson (Anderson, 2004). In this context, the Alienation Index is 1 – {proportion of overlap of the tester-non-tester frequency distributions}. The result indicates that the joint distribution of number of partners and UAI in the test and no-test group has significantly less overlap in 1997-2002 than in 1994-1996 (at the 1% level), confirming the parametric analysis.[36]

In summary, the no-test group becomes more conservative and the test group more risk inclined. This polarization result does not appear to stem from selection effects between the two groups; it supports the theory of complementarities between test and risk.

V. Conclusion

Highly Active Anti-Retroviral Therapies have vastly improved the quality and length of life of people infected with HIV. However, by lowering the cost of the disease and inducing "treatment optimism," they may have changed incentives for

[36] Program code and results available from the author.

risky sexual behavior and testing. In this paper, I develop and empirically test a rational model of individual behavior where susceptible agents decide whether to get tested and whether to take risks.

Using a unique data set that collects information on sexual behavior and testing in a high risk population between 1994 and 2002 I first substantiate the prediction of a global increase in risk following improved treatments in late 1996 through two distinct difference estimators. I show that those who get tested take relatively more risks than those who do not. Further, because of heterogeneity in preferences for risky sex, a polarization of behaviors emerges: people who have a high taste for risky sex being more likely to test and to take more risks, while those with a low taste for risky sex being less likely to test and taking fewer risks. However, I could not predict that testers and non testers use different margins to control their risk level: empirically, in response to the new treatments (and relative to a flexible time trend), testers increase their risk through UAI while non-testers decrease their number of partners. I performed an Anderson (2004) polarization test to confirm this result non-parametrically. If any, changes in the composition of the two groups over time plausibly emanate from newcomers in the MSM community, which does not seem to affect the main results.

These findings may be linked with Kremer and Morcom's (1998) conclusion that HIV may spread faster if those who take few risks take even fewer risks because they are least likely to contaminate others. In particular, Kremer (1996) shows, without reference to therapeutic improvements, that a high prevalence, "fatalistic" steady-state equilibrium with polarization of behaviors between low-activity and high-activity individuals is a noticeable possibility. In the light of my empirical results, it is plausible that ARVs fostered a new separating equilibrium in the San Francisco MSM community in the late 1990s between testers and non testers.

Yet, in view of the limited evidence on the evolution of HIV prevalence in the data, this high-risk equilibrium does not seem to have translated into a high prevalence equilibrium. To be sure, given the data at hand, I do not know with certainty if HIV+ individuals have decreased their risk level between the pre- and post- ARVs periods. However, given the similarity in the positive trend in sexual risk between positives and negatives after 1996, this hypothesis is unlikely. The most plausible explanation is that the ARVs-induced reduction in infectiousness has offset the global increase in the risk level. This could explain the paradox of non-testers reducing their risk level and the proportion of testers remaining roughly the same even if (practical) prevalence does not increase: at the time, few knew about the reduction in infectiousness, while the dramatic health benefits of ARVs were for everyone to see. In this context, it is therefore not surprising to detect a precautionary behavior by non-testers.

The previous remarks stress the need to understand better the response to therapeutic improvements. From a strictly epidemiological standpoint, if the unintended consequence of better treatments is to increase the spread of the epidemic, as suggested by Katz *et al*. (2002), the answer may be as a shift of R&D resources towards effective vaccine rather than treatments, or increased prevention programs. Alternatively, if ARVs have slowed or better yet, reversed the spread of HIV, it would be desirable to take steps to improve access to those treatments in the developing world (Over *et al*., 2004). In economic terms, according to Lakdawalla, Sood and Goldman (2006), for the average sexually active, non-monogamous American population, the increase in infection risk outweighs the expected welfare benefit from the reduction in the cost of the disease. This would be partially offset through (costly) extra precautionary behavior. However, as mentioned earlier, the assumption of random mixing might be particularly misleading in this context. Further, Philipson, Mechoulan and Jena (2006) show that the direct and external (altruistic) welfare benefits obtained through ARVs dominate the cost of the R&D investment that led to it. This suggests that in order to get an accurate picture, those net positive welfare effects, coupled with the presumably positive effects for the high risk susceptible population, should be balanced against the change in welfare of the at-risk (but low-risk) one that may be caused by the change in infection risk.

Table 1a: Data Description and Summary Statistics

Variables	# Obs	Mean	Std. Dev	Min	Max
Dependent Variables					
Test	39,884	0.904	-	0	1
(=1 if ever tested in last 6 months)					
UAI	48,888	0.257	-	0	1
(=1 if ever practiced unprotected anal intercourse in last 6 months)					
Numbers	48,055	8.414	23.01	0	750
(Number of partners in last 6 months)					
Intentions	22,161	0.93	-	0	1
(=1 if intention to practice safe sex)					
Vaginal sex	48,189	0.092	-	0	1
(=1 if sex with a woman in the last 6 months)					
Test Result	19,512	0.14	-	0	1
(=1 if HIV+, July 1997 onward; sample of testers)					
Independent Variables					
Age	48,596	33.476	9.325	15	99
Zip Code	44,210	-	-	00600	99941
San Francisco	26,439	8	9.6	0	81
(How long have you lived in San Francisco?)					
Ever Answered	48,112	0.264	-	0	1
(=1 if answered a STOP AIDS questionnaire before)					
Time	48,888	0.396	-	0	1
(Time trend)					
White	48,888	0.666	-	0	1
(= 1 if White)					
Ever Attended	39,677	0.179	-	0	1
(=1 if ever attended a STOP AIDS meeting)					
After	48,888	0.57	-	0	1
(= 1 after 1996, 0 before 1997)					
Bay Area	48,880	0.693	-	0	1
(= 1 if resident of San Francisco Bay Area)					
California	48,880	0.791	-	0	1
(= 1 if resident of California)					
Median Education	42,368	15,837	1.589	9	20
(Median education in Zip Code)					
Median HH Income	42,352	34,450	10,898	4,999	129,654
(Median Household Income in Zip Code)					

Table 1b: Average characteristics by Year

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Whole Sample									
% White	69	71	69	64	65	63	66	64	63
% Age<30	39	34	33	41	38	34	38	37	33
% 30≤Age≤45	53	55	55	49	50	53	51	49	51
% Age>45	9	9	12	10	12	13	11	14	16
San Francisco	7.8	7.8	7.8	7.5	8.2	8.5	7.9	9.1	9.8
% Bay Area	70	67	68	71	67	69	72	75	76
% California	80	77	78	80	78	80	83	79	79
Med Educ	15.96	16	15.95	15.75	15.72	15.71	15.62	15.87	15.96
Medi HH Inc	34,810	34,955	34,709	34,057	34,294	34,272	33,999	34,096	34,444
% Vag sex	9	9	9	9	9	9	10	9	9
Same for the four most represented five digit zip codes (32% of the sample)*									
% White	71	73	73	68	71	69	69	68	69
% Age<30	38	33	32	38	32	29	32	34	28
% 30≤Age≤45	54	55	56	52	55	57	56	51	53
% Age>45	8	11	12	10	13	14	12	16	15
San Francisco	7.6	7.7	7.6	7.5	8.2	8.4	7.9	9.1	10.3
% Vag sex	9	8	9	9	9	7	10	9	7

*By construction there can be no meaningful variations in Median Education and Median HH Income based on zip codes in the 2000 Census for those respondents.

TABLE 2: LINEAR PROBABILITY REGRESSIONS WITH DEPENDENT VARIABLE: Test

	(1)	(2)	(3)	(4)	(5)	(6)
After	-0.009	-0.003	0.001	-0.007	-0.014	-0.02
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)*	(0.008)**
Adjusted R^2	0.001	0.033	0.034	0.033	0.03	0.032
# observations	39,884	33,755	28,634	33,755	31,310	28,892

Model (1) controls for Time, $Time^2$

Model (2) controls for Time, $Time^2$ Age, Age^2, Bay Area, California, White, Ever Answered, Median Education, Median Household Income

Model (3) controls for Time, $Time^2$, Age, Age^2, Bay Area, California, White, Ever Answered, Median Education, Median Household Income, Ever Attended

Model (4): same as Model (2) with After=1 from October 1996 on

Model (5): same as Model (2) without known HIV+ testers (identifiable from July 1997 on)

Model (6): same as Model (5) without most likely HIV+ testers based on imputed HIV status (before July 1997). See text for details.

Robust standard errors in parentheses.

TABLE 3: LINEAR PROBABILITY REGRESSIONS WITH DEPENDENT VARIABLE: UAI

	non-testers	testers			Entire Sample
	(1)	(2)	(3)	(4)	(5)
After	0.016	0.08	0.059	0.056	0.034
	(0.035)	(0.01)***	(0.012)***	(0.013)***	(0.008)***
Adjusted R^2	0.01	0.038	0.033	0.03	0.03
# observations	2,951	30,804	28,359	25,941	41,635

All samples control for: Time, Time^2, Age, Age^2, White, Bay Area, California, Ever Answered, Median Household Income, Median Education

Samples (3): known HIV+ testers (14% of the testers) are removed from the sample of testers (from July 1997 on)

Samples (4): counterfactual sample of "susceptible testers only", same as Models (3) with 14% of most likely HIV+ testers removed (before July 1997) based imputed HIV status. See text for details.

Robust standard errors in parentheses. ***: 1% significance level; **: 5% significance level; *: 10% significance level.

Table 4a: Linear Probability Models for Difference Estimates
Dependent Variable: UAI

	(1)	(2)	(3)	(4)	(5)	(6)
White × After		0.037	0.038	0.036	0.042	0.03
		(0.009)***	(0.017)**	(0.017)**	(0.017)**	(0.018)*
After	0.034	0.01	0.008	0.01	-0.002	0.002
	(0.008)***	(0.01)	(0.01)	(0.014)	(0.014)	(0.015)
Adjusted R^2	0.03	0.03	0.03	0.031	0.026	0.024
# observations	41,635	41,635	41,635	41,635	39,174	36,756

All regressions contain controls for: Time, Time^2, Age, Age^2, White, Bay Area, California, Ever Answered, Median Household Income, Median Education

Model (3) contains controls for: White × Time, White × Time^2
Model (4) contains controls for: White × Ever Answered, White × Age, White × Age^2, White × Bay Area, White × California, White × Median education, White × Time, White × Time^2 †
Model (5): same as Model (4) without known HIV+ testers (identifiable from July 1997 on)†
Model (6): same as Model (5) without most likely HIV+ testers based on imputed HIV status (before July 1997).† See text for details.

Robust standard errors in parentheses ***: 1% significance level; **: 5% significance level; *: 10% significance level.

† Adding Vaginal Sex, Vaginal Sex × White does not affect results

Table 4b: Linear Probability Models for Difference Estimates
Dependent Variable: UAI

	(1)	(2)	(3)	(4)	(5)	(6)
Vaginal sex × After		-0.054	-0.097	-0.095	-0.088	-0.086
		(0.015)***	(0.029)***	(0.029)***	(0.029)***	(0.03)***
After	0.037	0.042	0.046	0.046	0.036	0.032
	(0.008)***	(0.008)***	(0.008)***	(0.008)***	(0.009)***	(0.009)***
Adjusted R^2	0.031	0.031	0.031	0.032	0.028	0.025
# observations	41,077	41,077	41,077	41,077	38,653	36,263

All regressions contain controls for: Time, Time^2, Age, Age^2, White, Bay Area, California, Ever Answered, Vaginal Sex, Median Household Income, Median Household Education

Model (3) contains controls for: Vaginal Sex × Time, Vaginal sex × Time^2.
Model (4) contains controls for: Vaginal Sex × Ever Answered, Vaginal Sex × Age, Vaginal Sex × Age^2, Vaginal Sex × Bay Area, Vaginal Sex × California, Vaginal Sex × White, Vaginal Sex × Time, Vaginal Sex × Time^2,
Model (5): same as Model (4) without known HIV+ testers (identifiable from July 1997 on)
Model (6): same as Model (5) without most likely HIV+ testers based on imputed HIV status (before July 1997). See text for details.

Robust standard errors in parentheses ***: 1% significance level; **: 5% significance level; *: 10% significance level.

TABLE 5: OLS REGRESSIONS WITH DEPENDENT VARIABLE:

	Numbers				**Syn = $1_{\{Numbers < 2\}}$ × (1 - UAI)**			
	Non-testers	testers			Non-testers	testers		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
After	-4.82	0.42	-0.517	-0.56	0.09	-0.01	-0.001	0.008
	(1.57)***	(0.635)	(0.634)	(0.66)	(0.04)**	(0.012)	(0.12)	(0.013)
Adjusted R^2	0.018	0.008	0.007	0.008	0.016	0.02	0.019	0.017
# observations	2,905	30,428	28,024	25,640	2,905	30,428	28,024	25,640

All regressions contain controls for: Time, Time^2, Age, Age^2, White, Bay Area, California, Ever Answered, Median Household Income, Median Education

Samples for models (1) and (5): non-testers
Samples for models (2)-(4) and (6)-(8): testers
Models (3) and (7): same as Models (2) and (6) without known HIV+ testers (identifiable from July 1997 on)

Models (4) and (8): same as Models (3) and (7) without most likely HIV+ testers based on imputed HIV status (before July 1997).

See text for details.

Robust standard errors in parentheses ***: 1% significance level; **: 5% significance level; *: 10% significance level.

Table 6: Average characteristics by Year and Testing Status

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Testers									
% White	69	71	70	67	66	64	67	64	63
% Age<30	37	32	32	34	36	32	35	35	32
% 30≤Age≤45	54	54	56	55	52	55	53	51	52
% Age>45	9	11	12	12	12	13	11	14	16
San Francisco	8	8	7.7	8	8.2	8.4	8	9.1	9.7
% Bay Area	70	68	68	70	68	69	73	76	77
% California	81	78	77	79	79	81	84	80	80
Med Educ	15.96	16.01	15.98	15.79	15.75	15.73	15.62	15.89	15.99
Med HH Inc	34,823	35,031	34,612	33,944	34,144	34,148	33,818	34,080	34,590
% Vag sex	9	8	9	9	8	8	9	9	8
Non Testers									
% White	66	69	67	55	58	58	56	57	60
% Age<30	51	47	46	55	58	52	62	63	37
% 30≤Age≤45	41	43	47	34	31	35	32	28	38
% Age>45	8	10	8	11	11	13	7	9	25
San Francisco	6.6	6.7	6.2	7.8	7.6	8.7	5.5	7.7	12.6
% Bay Area	65	60	63	61	56	62	62	63	53
% California	74	66	73	70	70	76	76	67	58
Med Educ	16	15.89	15.95	15.63	15.44	15.48	15.47	15.48	15.39
Med HH Inc	35,041	34,083	35,693	35,717	35,668	35,290	35,927	34,925	31,707
% Vag sex	14	13	13	12	13	12	16	12	15

Appendix

I. Individual behavior[37]

I.1 Solution to the joint maximization problem

A susceptible agent simultaneously chooses a number n of risky exposures (sexual trades), and decides whether to test ($t=1$) or not ($t=0$). Testing costs C. If tested positive, the agent gets treated. Utility of health status is: $u=0$ (normalization) if not infected; $u=\varsigma<0$ if infected and not treated; $u=\theta\in[\varsigma,0]$ if tested and treated. The parameter θ represents treatment quality, with $\theta=\varsigma$ if treatment has no positive effect, and $\theta=0$ in case of a perfect cure.

The probability that a risky trade will result in a contamination is the product of the probability (denoted by β) that a trade with an infected individual results in a contamination, times the probability to meet an infected individual. Under homogenous random matching, this latter probability is equal to p, the disease prevalence; each single contact therefore gives rise to an infection with probability βp.

Let us denote by $\pi(n,p,\beta)$ the probability to get infected, given a prevalence level of p and a number of risky trades of n. Thus $\pi(n,p,\beta)=1-(1-\beta p)^n$, which can be approximated for small values of p and n by $\pi(n,p,\beta)\simeq n\beta p$.[38] Since the individual takes this risk level as given, the notation $\pi(n,p,\beta)$ is simplified into $\pi(n)$ throughout this section.

The pleasure obtained from n trades is given by a utility function $u(n)$, such that $u'>0$, $u''<0$. Overall utility is additively separable in sexual pleasure, testing costs, and expected utility of health: $\tilde{U}(t,n)$ is defined such that:

$$\tilde{U}(0,n)=u(n)+\pi(n)\varsigma \text{ and } \tilde{U}(1,n)=u(n)+\pi(n)\theta-C \qquad \text{(A1)}$$

Put differently, utility may be written as a function of t and π. Denote by $n(\pi)$ the inverse of $\pi(n)$, *i.e.* the maximum number of risky trades acceptable in order to keep infection risk under a given probability π.
Also denote by $v(\pi)$ the utility level from a risky activity of $n(\pi)$, *i.e.* $v(\pi)\equiv u(n(\pi))$.

$$U(0,\pi)=v(\pi)+\pi\varsigma \text{ and } U(1,\pi)=v(\pi)+\pi\theta-C \qquad \text{(A2)}$$

---

[37] This section of the model reproduces the framework and results in Geoffard (2004).

[38] Notice that, under alternative matching assumptions, the probability to become infected may depend on the distribution of risk across the whole population (see Geoffard and Philipson, 1996).

The overall problem may be written in a very compact form:

$$\max_{t\in\{0,1\},\pi\geq 0} U(t,\pi) \qquad \text{(A3)}$$

The solution to this problem can be easily characterized. The optimal level of risk π^* is determined by the trade off between sexual pleasure $v(\pi)$, and the expected cost to be infected, which depends on the available treatment.

The surge of a new disease like AIDS induces two kinds of costs. First, a direct health cost for those who get infected. At the onset of the epidemic, no treatment was available, and the health cost is modeled as $\varsigma < 0$. Second, the HIV epidemic induced important behavior changes: overall, when it became clear that unprotected sex was a major risk factor, many individuals modified their behavior by reducing their number of partners and/or adopting safer sex practices. The model captures this feature in the following manner.

Let us denote by $P(x)$ the optimal level of risk under a potential cost of x (in other words, $P(x)$ is the utility-maximizing probability of incurring cost x), then $P(x)$ is the solution to $\max_P[v(P)+xP]$. It is characterized by the first order condition $v'(P)+x=0$. Not surprisingly, the optimal level of risk P is increasing with x: the larger the cost (the more negative x), the lower the acceptable level of risk. Abstracting from other health costs of unprotected sex, the optimal level of risky sex was $P(0)$ before the HIV epidemic, and $P(\varsigma)$ afterwards.

The cost of changing behavior, in terms of foregone sexual pleasure, can be therefore written as: $v(P(0))-v(P(\varsigma))$. The total individual cost of HIV (before the introduction of treatments) is the sum of this behavior cost and of the expected health cost:

$$K \equiv v(P(0))-v(P(\varsigma))-\varsigma P(\varsigma) \qquad \text{(A4)}$$

(By construction, this total cost is lower than $-\varsigma P(0)$, which would be the cost under no behavioral change)

The possibility to test for the presence of the HIV virus along with the introduction of treatments reduced the cost of illness. For individuals who get tested and treated, if the health condition is $\theta > \varsigma$ when treated, the optimal level of risk is $P(\theta) > P(\varsigma)$. One can now characterize the optimal testing behavior.

Proposition 1: Assume that the cost of testing C is smaller than the individual's expected cost of HIV, K. Then there exists a critical value $\theta^c > \varsigma$ such that testing is optimal if and only if the quality of treatment θ is larger than θ^c.

Proof: Define $V(x) \equiv \max_{\pi}[v(\pi) + x\pi]$. The individual decision problem (1) can be written as: $\max\{V(\varsigma); V(\theta) - C\}$. In the absence of any treatment, *i.e.*, $\theta = \varsigma$, it is never optimal to test. In the other extreme case, for $\theta = 0$ (perfect cure), $V(0) - V(\varsigma) = v(P(0)) - v(P(\varsigma)) - \varsigma P(\varsigma) = K$. Therefore, since $K > C$, it would always be optimal to test and get treated if a perfect cure existed. Moreover, by definition of $P(.)$ we have that $V(x) = xP(x) + v(P(x))$, and by the envelope theorem, $V'(x) = P(x) > 0$. Since V is increasing, there exists a threshold value $\theta^c \in [\varsigma, 0]$, defined as the solution to $V(\varsigma) = V(\theta^c) - C$. Testing (and getting treated) is then optimal whenever treatment quality is high enough (*i.e.*, when $\theta > \theta^c$).■

This proposition simply states that risky sex and testing are complements. Put differently, safe sex and testing are substitute. If testing costs increase, then access to treatments is more difficult, and the best way to reduce risk is to decrease the probability of infection. Symmetrically, if new treatments lower the costs of being infected, they increase the incentive to test (which is necessary to obtain access to treatments).

I.2 Comparative Statics

Increases in the overall prevalence rate p have non-trivial effects. Let us make the simplifying assumption that $\pi(n) = \beta np$. Again, this implicitly relies on random homogenous matching, and is a valid approximation for small values of p and βn. In that case, only the number of sexual trades and the overall prevalence rate matters to determine the risk level of a given individual (see *e.g.*, Geoffard and Philipson, 1995; Kremer, 1996). Under this assumption, we can state the following result:

Proposition 2: The optimal number of sexual trades n^* decreases with p. If we define the prevalence elasticity of the risk level by $(p/P)((\partial P/\partial p))$ then $(p/P)((\partial P/\partial p)) < 1$. If $\sigma \equiv -n[(u''(n)/(u'(n))] < 1$, then P decreases with p and θ^c increases with p.

Proof: The utility-maximizing number of trades is the solution to $V(x, p) \equiv \max_n[u(n) + x\pi(n, p)]$, where x is θ or ς depending on whether the individual is treated or not. Therefore, it is characterized by:

$u'(n^*) = -x\pi_n(n,p) = -x\beta p$. We immediately have that $\partial n^*/\partial p = -(\beta x/u''(n)) < 0$. Moreover, since $P(x,p) = \beta n^*(x,p)p$, we have that:

$$\frac{p}{P}\frac{\partial P}{\partial p} = \frac{1}{\beta n}\frac{\partial P}{\partial p} = \frac{p}{n}\frac{\partial n}{\partial p} + 1 = -\frac{\beta p x}{nu''} + 1 = \frac{u'}{nu''} + 1 = 1 - 1/\sigma \qquad \text{(A5)}$$

This proves that the prevalence elasticity of the risk level is always smaller than 1, and smaller than 0 if $-nu''(n)/u'(n) < 1$.

Moreover, the critical treatment quality θ^c that makes it worth to test is defined as the solution to $V(\theta^c, p) - C = V(\varsigma, p)$. Therefore, $V_x\theta_p^c = V_p(\varsigma,p) - V_p(\theta^c,p)$. Straightforward computation (recall $V_x = P(x,p)$) gives: $V_{px} = P_p = n(1-1/\sigma) < 0$ if $\sigma < 1$. Since $\theta^c > \varsigma$, we have that $V_p(\varsigma,p) > V_p(\theta^c,p)$. Since $V_x > 0$, this implies that θ^c increases with p.■

Proposition 2 reproduces the fundamental result of economic epidemiology that the optimal number of risky exposures decreases with the prevalence of the disease. However, for a given number of exposures, an increase in prevalence p also directly increases the risk level P. The net effect on P is negative only if the demand for risky trades is sufficiently prevalence-elastic, which is the case if $\sigma < 1$; and therefore, we have that $(p/P)((\partial P/\partial p)) = 1 - 1/\sigma < 0$. Moreover, θ^c increases with p, which implies that the optimal behavior at the margin may be a shift from test and high risk to no test and low risk. In short, if $\sigma < 1$, an increase in prevalence rate unambiguously decreases the risk level, and also decreases the demand for tests.

Symmetrically, if $\sigma > 1$ (inelastic demand), an increase in prevalence increases the risk level for those who test, and may even induce some individuals to shift from safer sex (and no testing) to testing and risky behavior. This is because for some individuals who adopt safer sex practices, the behavior loss from reducing the number of trades even further would be too large, and outweigh the benefit of risk avoidance.

I. 3 Aggregation of individual preferences

Let us now turn to aggregate behavior, and assume some form of heterogeneity, such that the critical value of θ differs across individuals. Specifically, agents' utility functions differ by a parameter λ, distributed among the population according to a probability distribution function g. Each agent still takes the prevalence rate p as given, and I now denote by $P(x, \lambda, p)$ the (individually)

optimal risk level for an agent of type λ , under a potential cost of x and for a given prevalence level of p .

The idea is to characterize a situation where for a given value of θ, agents with low λ do not get tested, and adopt safer behavior $P(\varsigma, \lambda, p)$, whereas agents with high λ get tested and adopt risky behavior $P(\theta, \lambda, p)$. As treatments get better, the critical value $\lambda^*(\theta, p)$ would decrease, which implies that more agents adopt risky behavior. To illustrate, consider the following representation of individual preferences:

$$u(n) = -\lambda \, \mathrm{e}^{-n/\lambda} \qquad \text{(A6)}$$

Here, λ represents the "taste for risky sex": a larger value of λ leads to more partners. Under this specification, it is straightforward to show that the optimal number of trades for an individual who tests (respectively, who does not test) is given by: $n_1 = -\lambda \ln(-\beta\theta p)$ (resp., $n_0 = -\lambda \ln(-\beta\varsigma p)$. The corresponding utility levels are $u(n_1) = -\lambda\beta\theta p$ and $u(n_0) = -\lambda\beta\varsigma p$; risk levels are given by $\pi_1 = -\beta p\lambda \ln(-\beta\theta p)$ and $\pi_0 = -\beta p\lambda \ln(-\beta\varsigma p)$.

The decision to test is simple to characterize: the agent tests if $u(n_1) + \pi_1\theta - C > u(n_0) + \pi_1\varsigma$. Replacing π_0 and π_1 by their values leads to the following decision rule:

$$\text{Test if:} \qquad \frac{C}{\lambda} < [\beta\theta p(1 - \ln(-\beta\theta p)) - \beta\varsigma p(1 - \ln(-\beta\varsigma p))] \qquad \text{(A7)}$$

This defines a threshold value $\lambda^*(\theta, p)$ such that an agent tests if λ is larger than $\lambda^*(\theta, p)$. Formally, with the previous functional form, $\lambda^*(\theta, p) = \dfrac{C}{f(-\beta\theta p) - f(-\beta\varsigma p)}$, with $f(x) = x\ln(x) - x$. Notice that an increase in θ (better treatments) lowers $\lambda^*(\theta, p)$. Here, an increase in p (larger prevalence) has ambiguous effects. [Since σ is equal to $n/\lambda = -\ln(-\beta p\varsigma)$ or $-\ln(-\beta p\theta)$, it is larger for agents who test than for those who do not.] In short, an increase in prevalence increases the incentive for agents with large values of λ more than it does for small values of λ . It may even decrease the incentive to test for small enough values of λ . At the margin, the effect is unclear, but it may well be the case that an increase in prevalence reduces the incentive to test. This occurs if the increase in prevalence reduces the level of risk sufficiently, so that testing is no longer worth its cost. Formally, under the previous specification of preferences: $P(x, \lambda, p) \simeq -\beta p\lambda \ln(-\beta xp)$.

Aggregation of individual behavior

The aggregation of individual behavior determines the overall demand for risky trades, formally equal to:

$$\int_{\lambda<\lambda^*(\theta,p)} n_0(\varsigma,\lambda,p)g(\lambda)d\lambda + \int_{\lambda\geq\lambda^*(\theta,p)} n_1(\theta,\lambda,p)g(\lambda)d\lambda$$

Notice that an individual who demands a number of risky trades *n* also supplies the same quantity. To keep things simple, let us assume that each individual can meet any other individual under random matching without frictions.

Assuming a constant prevalence

The risk level among individuals who do not test is given by $\int_{\lambda<\lambda^*(\theta,p)} P(\varsigma,\lambda,p)g(\lambda)d\lambda$, and among those who test (and may get treated if found positive) by $\int_{\lambda\geq\lambda^*(\theta,p)} P(\theta,\lambda,p)g(\lambda)d\lambda$.

The overall risk level is thus given by:

$$\varphi(\theta,p) = \int_{\lambda<\lambda^*(\theta,p)} P(\varsigma,\lambda,p)g(\lambda)d\lambda + \int_{\lambda\geq\lambda^*(\theta,p)} P(\theta,\lambda,p)g(\lambda)d\lambda \qquad \text{(A8)}$$

In the example this may be written as:

$$\varphi(\theta,p) = -\beta p\ln(-\beta p\varsigma)\int_{\lambda<\lambda^*(\theta,p)} \lambda g(\lambda)d\lambda - \beta p\ln(-\beta p\theta)\int_{\lambda\geq\lambda^*(\theta,p)} \lambda g(\lambda)d\lambda \quad \text{(A9)}$$

Notice that individuals who do not get tested do not get access to treatments, and therefore take fewer risks than individuals who do get tested and treated. Put differently, individuals who get tested are those who take more risks.

Effects of Treatments Improvements on Risk

An increase in treatment quality θ has two effects. The first, standard one is the direct effect of new treatments on the price of risky behavior: conditional on a positive test, the consequences of infection are less severe. Moreover, if ARVs reduce the average concentration of virus, they lower the risk of transmission per risky act, which further compounds the incentives to take more risks. Hence there should be an increase in the risk level among individuals for whom testing is already optimal.[39] Second, there is a selection effect from those individuals at the

[39] Moreover, the release of new treatments may spur so-called "treatment optimism" whereby individuals are overconfident about the effectiveness of such new treatments and their subsequent improvements (Auld, 2003).

margin who are indifferent between not testing (and low risk) and testing (and higher risk). On the other hand, there is no behavioral change among individuals who choose not to test, since they are indifferent to treatment opportunities. This is summarized in Proposition 3.

Proposition 3: The effect of new treatments on risk is the combination of direct and substitution effects that go in the same direction, *i.e.*, more risk.

Proof: Under the assumption that prevalence is exogenous, using Leibniz' rule, the effect of an increase in θ on φ is twofold:

$$\frac{\partial \varphi(\theta, p)}{\partial \theta} = \int_{\lambda \geq \lambda^*(\theta,p)} \frac{\partial P(\theta, \lambda, p)}{\partial \theta} g(\lambda) d\lambda + g(\lambda^*(\theta, p)) \frac{\partial \lambda^*(\theta, p)}{\partial \theta} [P(\varsigma, \lambda^*(\theta, p), p) - P(\theta, \lambda^*(\theta, p), p)] \quad \text{(A10)}$$

The first effect is the direct behavior change among those who get tested. The second part represents the effects of individuals at the margin, who now test and increase their risk level from $P(\varsigma, \lambda(\theta, p), p)$ to $P(\theta, \lambda(\theta, p), p)$ (recall that $\frac{\partial \lambda^*}{\partial \theta} < 0$). These individuals who are indifferent between not testing (and a low level of risk) and testing (and a higher level of risk) decide to test and increase their risk level correspondingly, thereby substituting one risk reduction strategy (reducing the number of risky exposures) for another (reducing the consequences of infection); this substitution effect is larger, the lower testing costs. ■

II. Partial equilibrium analysis when treatments affect prevalence

I finally study what happens if new treatments affect prevalence, *i.e.*, let $p = p(\theta)$. It is reasonable to assume that new treatments increase prevalence, *i.e.*, $p'(\theta) > 0$: this would be the case if the dominant effect comes from healthier seropositive selfish individuals on the market for risky trades, offsetting the decrease in the transmission parameter β. The susceptibles should in turn become relatively more cautious, and this would act as a counteracting force against the substitution effect.[40] Furthermore, this indirect effect also affects the

[40] This indirect effect, as in most models exhibiting offsetting feedback mechanisms, is expected to be second-order. An analogy would be the standard end product of a decrease in the wage on

selection effect. Depending on the magnitude of the equilibrium effect, the direction of the selection effect may even be reversed, *i.e.*, individuals at the margin could switch to no-testing and decrease their risk level. The effect of better treatments on the risk level becomes more complex and is summarized in Proposition 4:

Proposition 4: If prevalence is endogenous, the effect of better treatments on the risk level becomes ambiguous.

Proof: Still following Leibniz' rule, we now get:

$$\begin{aligned}\frac{d\varphi(\theta,p(\theta))}{d\theta} &= (\frac{d\lambda^*}{d\theta})P(\varsigma,\lambda^*(\theta,p(\theta)),p(\theta))g(\lambda(\theta,p(\theta))+\int_{\lambda<\lambda^*(\theta,p(\theta))}\frac{\partial P(\varsigma,\lambda,p(\theta))}{\partial\theta}g(\lambda)d\lambda \\ &-(\frac{d\lambda^*}{d\theta})P(\theta,\lambda^*(\theta,p(\theta)),p(\theta))g(\lambda(\theta,p(\theta))+\int_{\lambda\geq\lambda^*(\theta,p(\theta))}\frac{dP(\theta,\lambda,p(\theta))}{d\theta}g(\lambda)d\lambda \\ &= g(\lambda^*(\theta,p(\theta))(\frac{d\lambda^*}{d\theta})[P(\varsigma,\lambda^*(\theta,p(\theta)),p(\theta))-P(\theta,\lambda^*(\theta,p(\theta)),p(\theta))] \\ &+\int_{\lambda<\lambda^*(\theta,p(\theta))}\frac{\partial P(\varsigma,\lambda,p(\theta))}{\partial\theta}g(\lambda)d\lambda+\int_{\lambda\geq\lambda^*(\theta,p(\theta))}\frac{dP(\theta,\lambda,p(\theta))}{d\theta}g(\lambda)d\lambda\end{aligned} \tag{A11}$$

The sign of this expression depends on the key term $\frac{d\lambda^*}{d\theta}=\lambda_1^*+\lambda_2^*p'$. As mentioned earlier, the change in λ^* with respect to $p(\theta)$, *i.e.*, λ_2^*, could go either way (depending on the elasticity of sexual risk with respect to prevalence). As a consequence, individuals at the margin (the first term), could thus move towards more risk and test or move towards less risk and not test. Also, $\frac{\partial P(\varsigma,\lambda,p(\theta))}{\partial\theta}$ is here negative : those who do not test after the improvement of treatments should take fewer risks. ■

---

the competitive equilibrium of workers hired: as the wage falls, the demand for labor increases but the price of output falls as well, which results in less of an increase in labor than would have been the case had output price remained constant. The main difference is that here, the behavioral change of the seropositives is unknown.

References

Ai, C. and E. C. Norton (2003) "Interaction terms in Logit and Probit Models," *Economics Letters* 80 (1): 123-129

Anderson, G. (2004) "Toward an Empirical Analysis of Polarization," *Journal of Econometrics* 122 (1): 1-26.

Auld, C. (2003) "Choices, Beliefs, and Infectious Disease Dynamics," *Journal of Health Economics* 22 (3): 361-377.

Blower S. M., H. B. Gershengorn and R. Grant (2000) "A Tale of Two Futures: HIV and Antiretroviral Therapy in San Francisco," *Science* 287: 650–654.

Boozer, M. A., and T. J. Philipson (2000) "Estimating Information Elasticities: the Case of Public Testing for HIV," *Journal of Human Resources* 35: 419-446.

Campo, R. E. *et al*. (2005) "ARVs Treatment Considerations in Latino Patients," *AIDS Patient Care and STDS* 19 (6): 366-374

Catania, J. A. *et al*. (2001) "The Continuing HIV Epidemic among Men who have Sex with Men," *American Journal of Public Health* 91 (6): 907-914.

CASCADE Collaboration (2003) "Determinants of Survival following HIV-1 Seroconversion after the Introduction of ARVs," *Lancet* 362: 1267-74.

Chen, S. Y. *et al*. (2002) "Continuing Increases in Sexual Risk Behavior and Sexually Transmitted Diseases among Men who have Sex with Men, San Francisco 1999-2001," *American Journal of Public Health* 92: 1387-1388.

Crepaz, N., T. A. Hart, and G. Marks (2004) "Highly Active ARVs Therapy and Sexual Risk Behavior: a Meta-analytic Review," *JAMA* 292: 224-236.

Duggan, M. G., and W. N. Evans (2005) "Estimating the Impact of Medical Innovation: The Case of HIV Antiretroviral Treatments," NBER Working Paper 11109

Eeckhoudt, L., C. Gollier, and G. Immordino (2001) "How Diagnostic Tests Affect Prevention: a Cost Benefit Analysis," University of Salerno CSEF Working Paper #53.

Eurosurveillance (2002) European Comuniable Disease Bulletin Vol 7 (2)

Fox, K. K. *et al*. (2001) "Gonorrhea in the HIV Era: a Reversal in Trends among MSM," *American Journal of Public Health* 91: 959-964.

Francis, A. (2005) "The Economics of Sexuality: The Effect of HIV/AIDS on Sexual Behavior, Desire and Identity in the United States," Mimeo, University of Chicago

Gebo K. A, *et al*. (2005) "Racial and Gender Disparities in Receipt of Highly Active ARVs Therapy Persist in a Multistate Sample of HIV Patients in 2001," *Acquir Immune Defic Syndr*. 38 (1): 96-103

Geoffard, P-Y, and T. J. Philipson (1996) "Rational Epidemics and their Public Control," *International Economic Review* 37, 603-624

Geoffard, P-Y (2004) "Risky Sex and Screening Decisions" Mimeo, PSE

Geoffard, P-Y, and S. Mechoulan (2004) "Risky Sexual Behavior and Incentives: the Impact of New Therapies on HIV Prevention," (in French) *La Revue Économique* 55 (5): 883-899.

Gersovitz, M., and J. S. Hammer (2003) "Infectious Diseases, Public Policy, and the Marriage of Economics and Epidemiology," *The World Bank Research Observer* 18: 129-157

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* 14357: 1149-53

Grulich, A. (2000) "Increase in Unsafe Sex and Sexually Transmissible Infections in Gay Men in Australia," Working Paper, National Centre in HIV Epidemiology and Clinical Research, UNSW, Sydney.

Health Canada (2000) Volume: 26S6 - October 2000: 1998/1999 Canadian Sexually Transmitted Diseases (STD) Surveillance Report

Huebner, D., G. Rebchook and S. Kegeles (2004) "A Longitudinal Study of the Association between Treatment Optimism and Sexual Risk Behavior in Young Adult Gay and Bisexual Men," *Journal of Acquired Immune Deficiency Syndrome* 37: 1514-1519

Kahn, J. G. *et al*. (2002) "Access to and use of HIV ARVs therapy: Variation by race/ethnicity in two Public Insurance Programs in the U.S.," *Public Health Reports* 117: 252-262

Kass N., Flynn C., Jacobson L, Chmiel J.S., and Bing E. (1999) "Effect of Race on Insurance Coverage and Health Service Use for HIV-Infected Gay Men," *Epidemiology* 20(1): 85-92.

Katz, M. H. *et al*. (2002) "Impact of highly active ARVs treatment on HIV seroincidence among Men who have Sex with Men in San Francisco," *American Journal of Public Health* 92 (3): 388-394.

Kenkel, D. S. (2000) Prevention. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* New York: Elsevier. Chap 31: 1675-1714

Kremer, M. (1996) "Integrating Behavioral Choice Into Epidemiological Models of AIDS," *Quarterly Journal of Economics* 111 (2): 549-73.

Kremer, M. and C. Morcom (1998) "The effect of Changing Sexual Activity on HIV Prevalence," *Mathematical Biosciences* 151 (1): 99-122.

Lakdawalla, D., N. Sood, and D. Goldman (2006) "HIV Treatment Breakthroughs and Risky Sexual Behavior," *Quarterly Journal of Economics* 121: 1063-1102

Lopez-Quintero, C., R. Shtarkshall, and Y. D. Neumark. (2005) "Barriers to HIV-Testing Among Hispanics in the United States: Analysis of the National Health Interview Survey, 2000," *AIDS Patient Care and STDs* 19 (10): 886-897

Malebranche, D. (2005) "HIV Treatment Series: Adverse Health Outcomes Among Black Americans -- Living With HIV: Nature or Nurture?," *Test Positive Aware Network* November / December 2005

Meyer, B. (1995) "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics* 13 (2): 151-161

Moore, Richard D. *et al*. (1994) "Racial Differences in the Use of Drug Therapy for HIV Disease in an Urban Community," *New England Journal of Medicine* 330 (11): 763-768

Morin S. F. *et al*. (2002) "Responding to Racial and Ethnic Disparities in Use of HIV Drugs: Analysis of State Policies," *Public Health Reports* 117 (3): 263-272.

Over, M, P. Heywood, J. Gold, I. Gupta, S. Hira and E. Marseille 2004 "HIV/AIDS Treatment and Prevention in India," Health, Nutrition and Population Series, World Bank

Page-Shafer, K. A. *et al*. (1999) "Increasing HIV Risk Behavior and Sexually Transmitted Disease Rates in the Interval of Improved Efficacy of ARVs Therapy Regimens, San Francisco, 1994-1997," *MMWR* 48 (3):46-48.

Peltzman, S. (1975) "The Effects of Automobile Safety Regulation," *Journal of Political Economy* 83 (4): 677-726

Philipson, T. J. (2000) Economic Epidemiology and Infectious Diseases. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* New York: Elsevier. Chap 33: 1761-1797.

Philipson, T. J. and R. A. Posner (1995) "A Theoretical and Empirical Investigation of the Effects of Public Health Subsidies for STD Testing," *Quarterly Journal of Economics* 110: 445-474

Philipson, T. J., S. Mechoulan and B. Jena (2006) "IP & External Consumption Effects: Generalizations from Health Care Markets," NBER Working Paper No. 11930

Posner, R. A. (1992) *Sex and Reason*, Harvard University Press, Cambridge, Massachussets.

Philipson, T. J. and R. A. Posner (1994) "Public Spending on AIDS Education: an Economic Analysis," *Journal of Law and Economics* 37 (1): 17-38

Quinn, T. C. *et al*. (2000) "Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1," *New England Journal of Medicine* 342: 921-29

Van de Ven P., L. Mao and G. Prestage (2004) "Gay Asian Men in Sydney Resist International Trend: no Change in Rates of Unprotected Anal Intercourse, 1999-2002," *AIDS Educ Prev* 16(1): 1-12.

Villarosa, L. (2004) "Patients with HIV Seen as Separated by Racial Divide," *NY Times* August 7th, 2004

World Health Organization (2005), Report on the global AIDS epidemic.