Risky Sexual Behavior, Testing and New HIV Treatments

Stéphane Mechoulan
University of Toronto

Abstract

This paper studies the impact of new HIV therapies (HAART) 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 HAART in late 1996 and the sharp increase in risky sexual behavior that ensued. Further, following HAART, testers take more risks while non-testers take fewer risks. The proportion of testers remains stable, which was ambiguous *a priori*, and HAART does not alter the composition of the testing and non-testing groups.

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 Therapy (HAART) in late 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 at least 3 million HIV-positive patients treated by 2006 worldwide. It is urgent to consider the behavioral impacts of those 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.

The paper presents an empirical analysis of the effects of HAART on testing, unprotected sex, and number of sexual partners. Better treatments reduce the price of risky sex, yet their effects are complex to interpret. First, holding prevalence constant, one can think of the direct, first-order effect of HAART introduction. Empirically, it should translate into an increase in risky sexual behavior. This is my first testable hypothesis.

Second, I identify a corresponding equilibrium effect. As treatment quality improves, HIV prevalence is expected to rise: this stems from the higher level of risk, together with an increase in the proportion of healthier HIV+ individuals present in the market for risky exposures. Everything else being equal, the optimal number of risky exposures decreases with the prevalence of the disease. Yet the net effect of prevalence on the risk level[1] is negative only if risk taking is sufficiently prevalence elastic. Otherwise, an increase in prevalence may induce

[1] The risk level depends on both the amount of risky exposures and the prevalence of the disease.

some individuals to shift from safer sex and no testing to test and riskier behavior. In turn, the expected increase in prevalence should increase the incentive to test for agents with a high "taste for risky sex" and should decrease the incentive to test for those with a low "taste for risky sex". Consequently I predict that testers will take relatively more risks than non-testers following the introduction of HAART, which constitutes a second testable hypothesis.

Furthermore, with heterogeneity in preferences for risky sex, 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 a sharp increase in overall risky sexual behavior starting exactly in 1997. I use a difference method to purge the causal effect of HAART from other potential confounding factors: as expected, Whites and gays (as opposed to bisexuals) respond to HAART 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 further that the average characteristics of the testing and non-testing groups are mostly unaffected by HAART.

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 HAART less excruciating by reducing the number of pills to be taken and side effects experienced. HAART has 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, HAART decreases 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 HAART 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 HAART 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 HAART; 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 HAART 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-HAART 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 HAART, 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 HAART. 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 HAART 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 HAART 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 Goldman, Lakdawalla and Sood (2004) 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 and their data allows them to consider number of partners as the only risk variable.

I qualify their conclusion on two points. First, Goldman, Lakdawalla and Sood (2004) interpret a reduction in (observed) prevalence following HAART as increased precautionary

behavior by the uninfected. Yet, such an observation would be equally plausibly consistent with a flow of new high risk testers who previously practiced safer sex with no testing and are therefore, at first, for the large majority of them, HIV-. So while I agree that it is possible that "reductions in prevalence might suggest worse welfare outcomes, not better ones," I stress that the measurement of HIV prevalence itself is endogenous. Second, the conclusion that because HIV+ individuals take more risks they increase the spread of the disease seems premature. Beyond changes in number of partners by the HIV+, the spread of HIV stems from the type of sex chosen, the behaviors of susceptibles and the diminished risk of HIV transmission when HIV patients are under HAART (Blower, Gershengorn and Grant, 2000).

B. Hypotheses

The arguments underlying my conjectures build on the literature that addresses the interaction between primary and secondary prevention. See Kenkel (2000) and Eeckhoudt *et al*. (2001) for details.

Susceptible individuals 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 overall prevalence rate (taken as given), and the availability of treatment. One of the fundamental results of economic epidemiology is 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 of a higher prevalence on the risk level is negative only if the demand for risky

exposures is sufficiently prevalence-elastic. Conversely, if that demand is not elastic enough, an increase in prevalence may convince some individuals to shift from safer sex and no testing to testing and risky behavior: for these individuals, the utility loss from further reducing their sexual activity would be too large and outweigh the benefit of risk avoidance. Safer sex and testing are thus alternate ways to lower the risk level, either by reducing the probability of infection, or by reducing its consequences.

Then, how does one disentangle the effects of better treatments? The analysis of the first, direct effect assumes a constant prevalence. There should be an increase in the risk level among individuals for whom testing is already optimal. Indeed, for those who test regularly, improved treatments unambiguously reduce the cost of getting the disease. 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). Second, there should be a selection effect, *i.e.*, individuals at the margin who switch to testing and increase their risk level.[2] These individuals would substitute one risk reduction strategy (reducing the number of risky exposures) for another (reducing the consequences of infection). Third, the effect of new treatment availability on risk taking for the non-testers should be non-negative, based on their option value of testing. Empirically, the direct and presumably first-order effect should then result in an increase in risky sexual behavior and intentions to practice unsafe sex following the introduction of HAART. Formally, defining Risk = f(HAART availability, Prevalence, Individual characteristics), this translates as:

$$\text{Hypothesis (1): } f_1 > 0 \qquad (1)$$

[2] These direct and indirect demand effects are analyzed formally in Geoffard (2004).

Simultaneously, I envision a corresponding equilibrium effect. Note that as treatment quality improves, HIV prevalence should increase. First, as explained earlier, susceptibles should take more risks. Second, the new drugs extend the lives and improve the quality of life of HIV+ individuals, which implies an increase in the proportion of those present in the market for risky exposures. The susceptibles should in turn become relatively more cautious, and this would act as a counteracting force against riskier behavior.[3] Because at the margin an increase in prevalence increases the incentive to test for agents with a high "taste for risk" (those with relatively inelastic demand for risky sex) and decreases the incentive to test for those with a low "taste for risk", I predict that testers will take more risks than non-testers following the introduction of HAART. Formally, defining [Risk | Test] = g(HAART availability, Prevalence, Individual characteristics | Test) and [Risk | No Test] = h(HAART availability, Prevalence, Individual characteristics | No Test), this translates as:

$$\text{Hypothesis (2): } g_1 > h_1 \qquad (2)$$

In the limit, with enough heterogeneity in preferences for risky sex, 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 HAART. Formally, this translates as:

[3] This offsetting indirect effect, however, is expected to be second-order. An analogy would be the standard end product of a decrease in the wage on 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.

Hypothesis (3): $g_1 > 0$ and $h_1 < 0$ (3)

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

A central assumption underlying Hypotheses (2) and (3) is that the option value of treatments when non-testing is significantly lower than the value of treatments under testing. 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 month) CD4 count testing so that the start of HAART 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.[4] 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 2006: 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

[4] HAART seems 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.

less likely to fully benefit from HAART. These factors therefore support Hypothesis (2) and Hypothesis (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, which may be simultaneously deprived of some of its lowest-risk members. Therefore, changes in the observed proportion of positive tests may provide a downwardly biased estimate of changes in actual risky behavior. In fact, 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 targets MSM.

The sampling frame is as follows. The SAP volunteers randomly intercept people 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. Those people then answer a detailed questionnaire. 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 HAART, it is essential in identifying treatment effects on behavior.[6] 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. Some people may have been sampled more than once and if so, it is indicated.

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. On

[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 data have been used in the public health literature over the past several years and their validity is well documented.

the other hand, the questionnaire is relatively poor with socio-demographic variables. Only age, 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.[7]

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).[8] 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.[9]

[7] 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.

[8] 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.

[9] 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 household education in zip code).[10] 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.

As a final caveat, it is important to note that I restrict attention to sexual behavior. 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. Nevertheless, I believe that the new insights that can be collected from those data outweigh the limitations of the design.

B. Summary Statistics

Table 1 provides the descriptive statistics of the sample with definitions of all the variables. The main variables are defined as follows. d_after is the dummy variable corresponding to the introduction of HAART: d_after =1 starting from 1997, 0 otherwise. d_test is a dummy

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

indicating whether the individual declares "knowing his serological status" (before February 1997) or "performed an HIV test" (after January 1997).[11] 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." A large majority of the respondents are testers. 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+.[12] 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.

Those variables measuring sexual activity in the six months prior to the interview are numbers (number of partners)[13], d_analsex (at least one case of anal intercourse, henceforth AI), d_analcond (at least one case where no condom was used), and d_vaginalsex (at least one case of vaginal sex). I use d_vaginalsex as a proxy for sexual orientation (*i.e.*, bisexual vs. gay).

[11] 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.

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

[13] The variable numbers (number of partners) is censored at 999 but highly promiscuous respondents make only a tiny fraction of the sample: 0.63% declares a number of partners greater than 100 over six months.

Finally, the question “Do you intend to practice safe sex?” was dropped after January 1997 but resumed in the second half of 2000.

Regrettably, the formulation of the 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; which is why my empirical analysis deals with a more conservative (if not more robust) indicator. I recombined variables d_analsex and d_analcond 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: d_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.

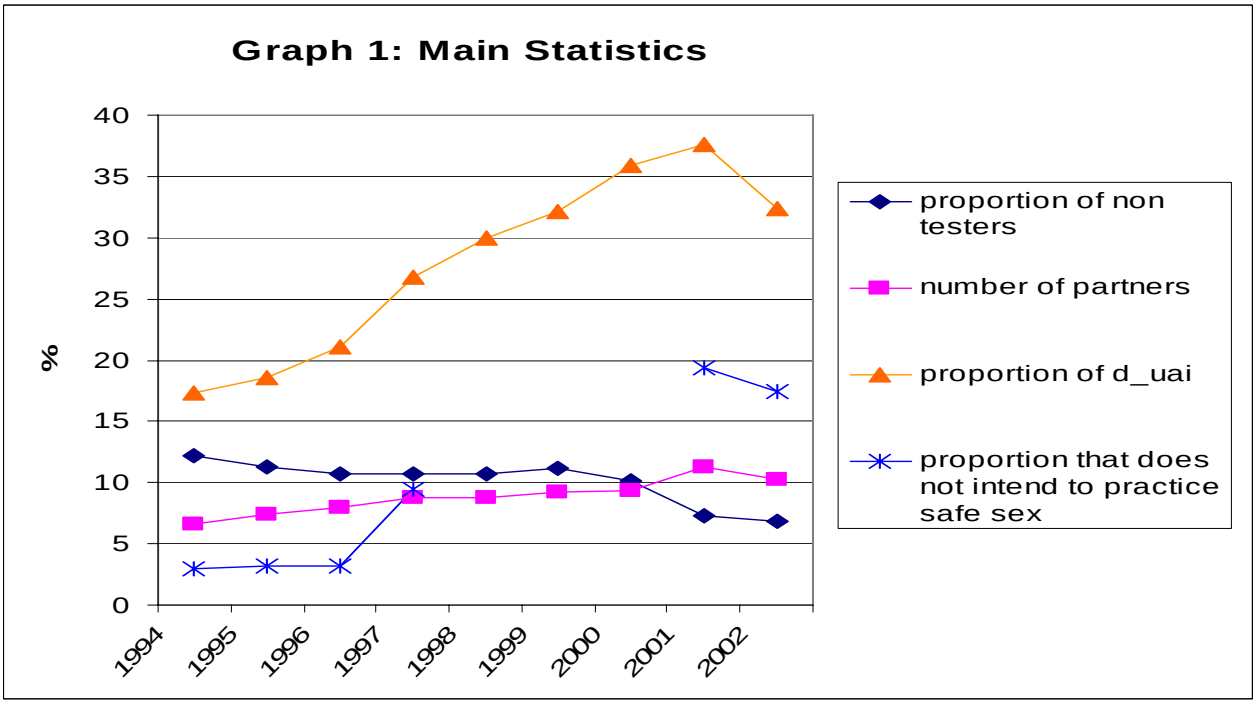
It is also useful to consider a synthetic indicator of risk that combines intensive and extensive margins. I construct d_Syn = d_{Numbers<2}×(1-d_UAI), where d_{Numbers<2} is an indicator for whether the individual declares fewer than two partners in the last six months. In other words, d_Syn = 1 if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

IV. Empirical Analysis

A. Data Patterns

The following diagrams illustrate some of the most interesting features conveyed by the raw data. Graph 1 summarizes the evolution of the main variables. The most striking patterns are

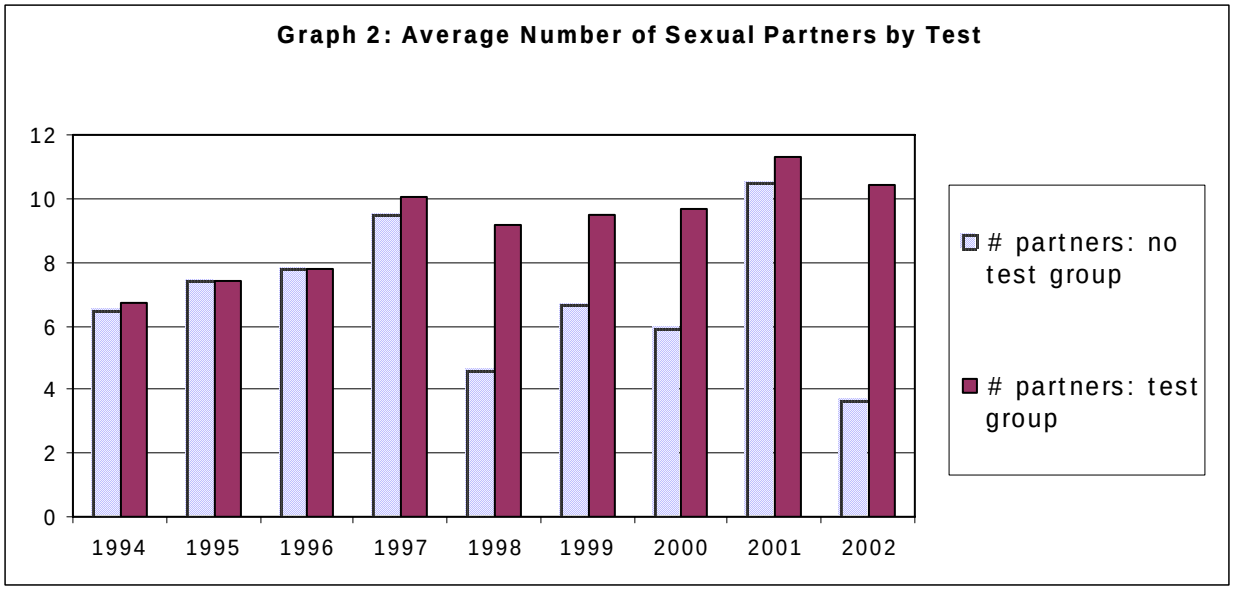
the increase in UAI over the period (see Page-Shafer *et al*., 1999; Chen *et al*., 2002), together with the increase in the proportion of individuals who do not intend to practice safe sex. These two features of the graph provide strong support for Hypothesis (1).[14] At the same time, testing odds remain fairly constant at a high level (88 to 93%). Recall that I only have the first semester of 2002. On top of seasonality issues (on average, I find that individuals take more risks in the second semester), the sample size is smaller than in previous years so any break observed that year — for this graph and the following ones — should be interpreted with caution.



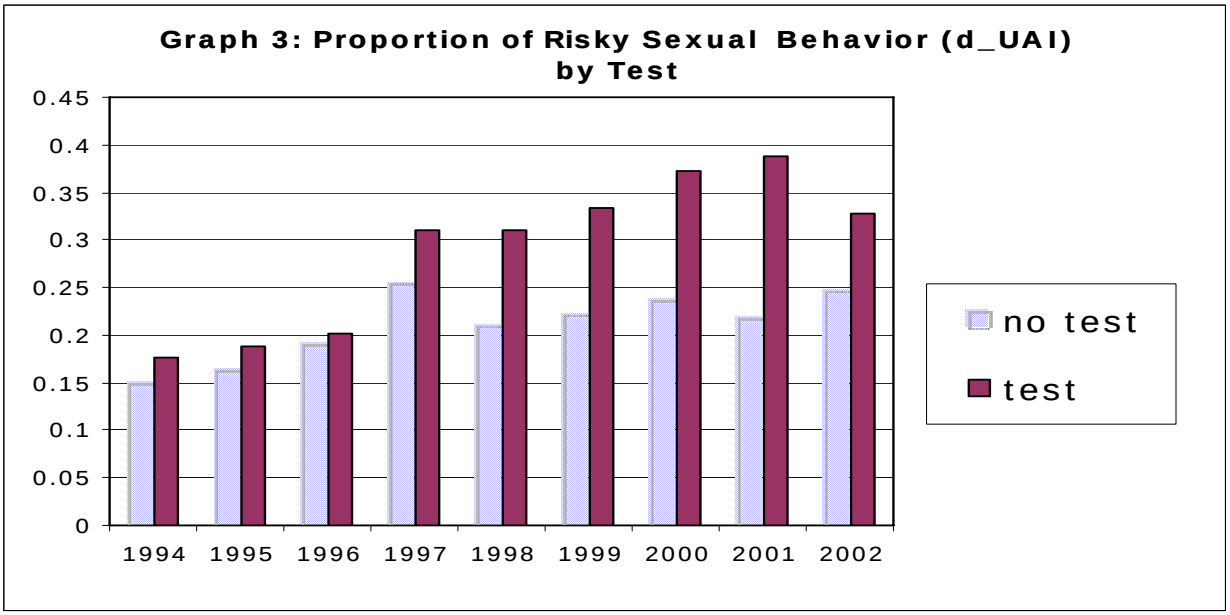
I start with the analysis of risk taking along the two dimensions of number of partners and UAI. I first compare the evolution of the number of partners for those who test and those who do not. Graph 2 illustrates a key element: the difference in number of partners between those who

[14] 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).

test and those who do not test increases sharply after 1996. The feature supports at least Hypothesis (2) and to a lesser extent Hypothesis (3).

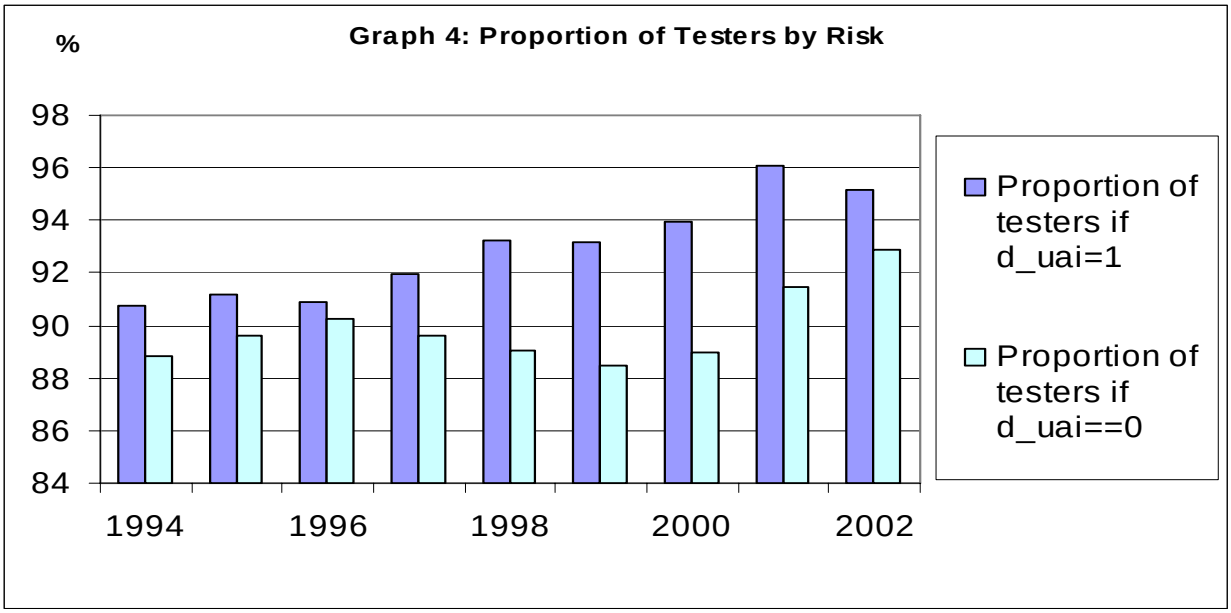


When looking at the proportion of UAI by testing category (Graph 3), I note that from 1997 on, the difference in the probability of UAI increases discretely between testers and non-testers. This significant and dramatic increase in UAI, specifically after 1996 among those who test, further supports Hypothesis (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.

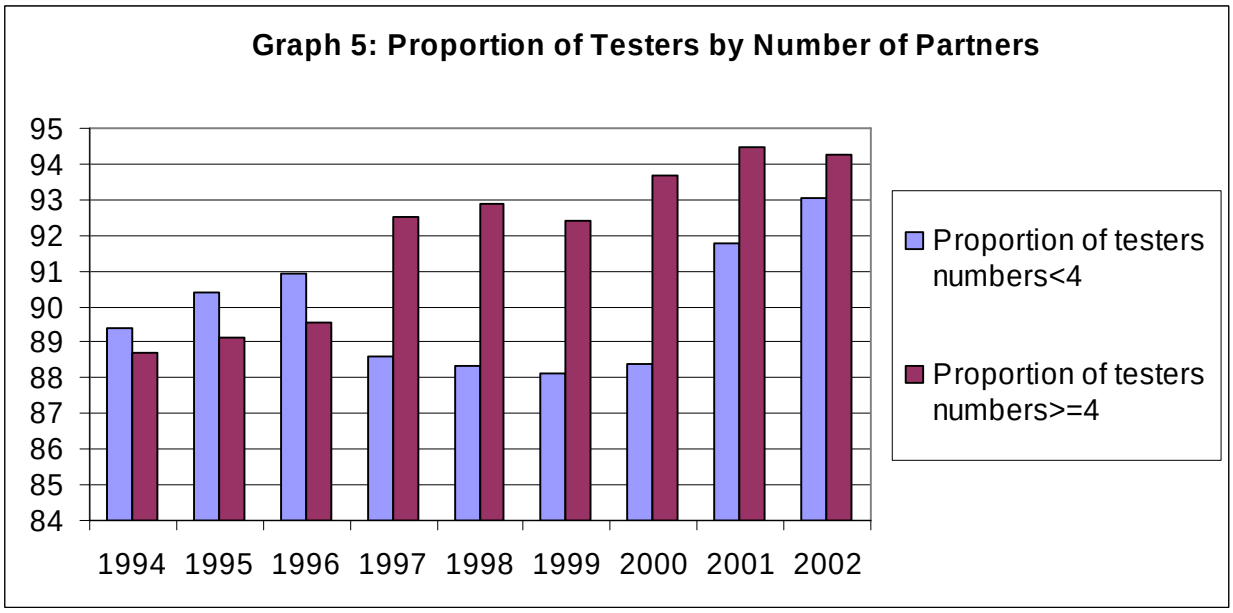


As for the proportion of testers, most noticeably I have just shown that it increases overall only in the last two years of the survey.[15] Note that the difference in the proportions of testers between the two groups UAI=0 and UAI=1 which was decreasing between 1994 and 1996 widens significantly afterwards (Graph 4). However, one should bear in mind the scale used: the changes in the proportion of testers are small.

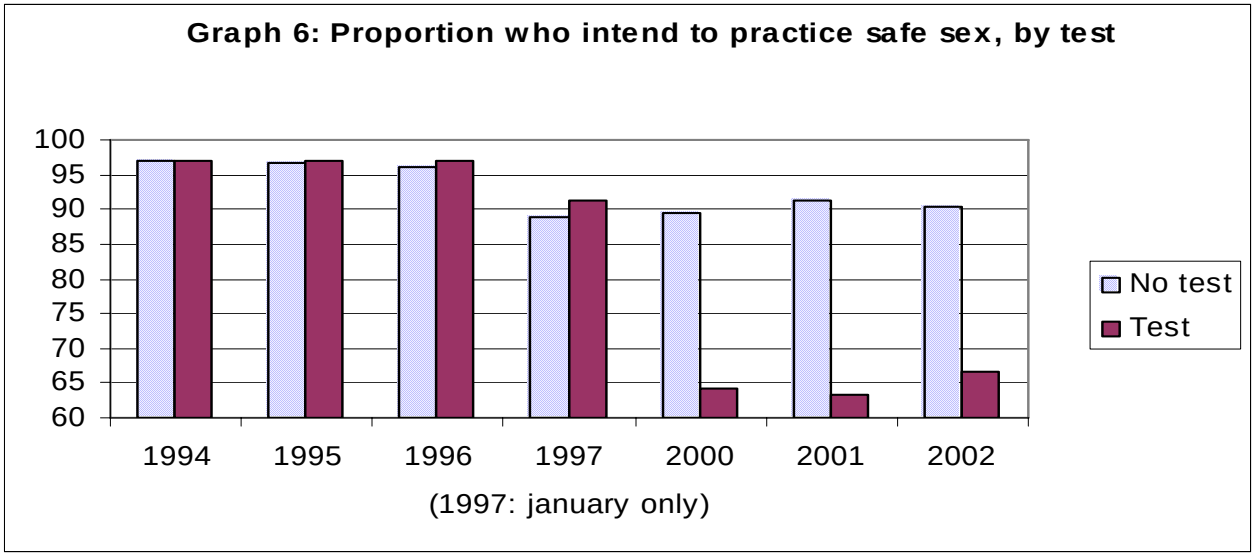
[15] This discontinuity is puzzling. However, I note that a similar increase in the proportion of testers is observed in Ontario in those two years (Report on HIV/AIDS in Ontario 2003) and is unexplained by the principal investigator responsible for the data collection. A sharp increase has been signaled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



The proportion of testers by number of partners (Graph 5) 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 as well) 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 6), one can also see a dramatic 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. Note that the decrease in intention to practice safe sex appears much more pronounced among those who declare testing. This is another argument in support of Hypotheses (1) and (2).



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

B. Econometric Analysis

Before testing my three hypotheses, I first verify whether HAART significantly influences the testing decision. The demand for testing stems for individuals with a sufficiently large probability to be infected. Given a certain risk level, there is a critical value of treatment quality such that testing is optimal if and only if the quality of treatment is greater than that quality cutoff. The introduction of HAART should then lead to an increase in testing. However, because changes in prevalence affect both risk taking and the incentives to test, a possible offsetting effect of treatments is, as mentioned earlier, the more risk-averse fraction of the

population decreasing its risk level and therefore its incentives to test. This could occur if the increase in prevalence sufficiently reduces the level of risk, so that testing is no longer worth its costs.

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.

$$\Pr(d\_test = 1) = \Phi(\alpha + \beta d\_after + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\_Characteristics_i) \quad (4)$$

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 increase again after 2000. The different specifications presented in Table 2 confirm this initial view. Table 2 tells us that the coefficient of d_after, if anything, is actually negative and always non-significant.[16]

A closer look at the stable proportion of testers following the introduction of HAART reveals that it is driven down by those with no UAI, which was already expected from Graph 4. This pattern is consistent with the proposition of complementarities between risk and testing. What is more puzzling is the non-significance of d_after when regressing d_test over the sub sample where d_UAI=1 exclusively: I expect risk takers to be more likely to test if it becomes more valuable. A possible explanation is that so many of them are testers in the first place. According to the preceding discussion, another explanation for the stability of testing after 1996 would be that the prevalence effect offsets the direct incentive to get tested.

[16] The coefficients reported here and for all other Probit regressions, are the discrete probability changes $d\Phi/dx$ computed at the means of the data, where Φ is the cumulative distribution function of N(0,1). For each coefficient, the t-test reported is that of the underlying Probit coefficient being zero.

However, I do not observe a significant increase in prevalence in the sample from the time this information becomes available on. 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 HAART. For example, testing may allow people to obtain unprotected sex with a higher probability and 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, biased 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 the non-testing group as a control group. Table 3 presents an analysis of d_UAI within the testing and non-testing groups where I estimate the following model:

$$\Pr(d\_UAI = 1) = \Phi(\alpha + \beta d\_after + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\_Characteristics_i) \qquad (5)$$

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 Probit on the subsample of testers, with observable HIV+ removed (from July 1997 on). Column (4) does the same on the counterfactual subsample 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. To summarize, the increase in UAI occurs within the

testing group (as expected) but not in the non-testing group (where all I can say is that the effect should be smaller). This finding constitutes a validation of Hypothesis (2). Yet, if Table 2 is to present conclusive evidence of the causal impact of HAART 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.[17] Such a concern, however, can be reasonably qualified. A flexible time trend buys considerable identification power in this context because the usual factors of sexual risk among MSM are known to show a high level of persistence. 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 among MSM vary slowly over time. Consequently, a flexible time trend should capture most of the secular changes in risk-taking. It is especially important here because the increase in risk predates the introduction of HAART. If the significance of a break in the data "survives" the

[17] To my knowledge, the only other element that meets those criteria and may have affected sexual behavior is a significant expansion of the Californian AIS Drug Assistance Program (ADAP) — itself a consequence of HAART. 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.

inclusion of such controls, it should intuitively correspond to 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 Goldman, Lakdawalla and Sood (2004). 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.[18] The medical literature further reveals that, for genetic reasons, Blacks are less responsive to HAART.[19] Consequently, among Black and Hispanic men, the death rate from AIDS is significantly greater than for Whites. The behavioral response of Whites should 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).[20] I thus form the hypothesis that gay men will be taking relatively more UAI risk following HAART compared to bisexuals. 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 HAART than those who do. The proportion of respondents who have sex with women is stable at 9%

[18] 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).

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

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

before and after 1996 which supports the idea that sex with women is exogenous with respect to HAART.[21]

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: d_caucasian × d_after, and d_vaginalsex × d_after. The idea behind the interaction term d_caucasian × d_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: d_vaginalsex × d_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-HAART 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 d_caucasian × d_after becomes then the pure differential effect of HAART between Whites and non-Whites. The coefficient of the variable d_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 d_caucasian × Time and d_caucasian × Time^2:

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

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 in the form of linear probability models.[22] For example, the preferred specification for the first model is of the form:

$$d\_UAI = \alpha + \beta d\_after + + \phi d\_after \times d\_Caucasian + \gamma Time + \delta Time^2 + \eta Time \times d\_Caucasian + \kappa Time^2 \times d\_Caucasian + \sum_i \lambda_i Individual\_Characteristics_i + \sum_i \mu_i \{Individual\_Characteristics_i \times d\_Caucasian\} + \varepsilon$$

where ε is normally distributed. (6)

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.[23] 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

[22] See Ai and Norton (2003) who discuss problems with interaction terms in nonlinear models.

[23] See Meyer (1995) for a comprehensive discussion. However, note that even if d_Caucasian × d_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 HAART, they should take more risks.

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 HAART 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 Numbers and d_Syn within the testing and non-testing groups where I estimate the following models:

$$Numbers = \alpha + \beta d\_after + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\_Characteristics_i + \varepsilon$$

where ε is normally distributed. (7)

$$\Pr(d\_Syn = 1) = \Phi(\alpha + \beta d\_after + \gamma Time + \delta Time^2 + \sum_i \lambda_i Individual\_Characteristics_i) \quad (8)$$

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 d_Syn. Models (1) and (5) support the idea that non-testers adopt safer attitudes following HAART.[24] From models (2)-(4), I can safely 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

[24] 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. Note that these results are driven by the White respondents.

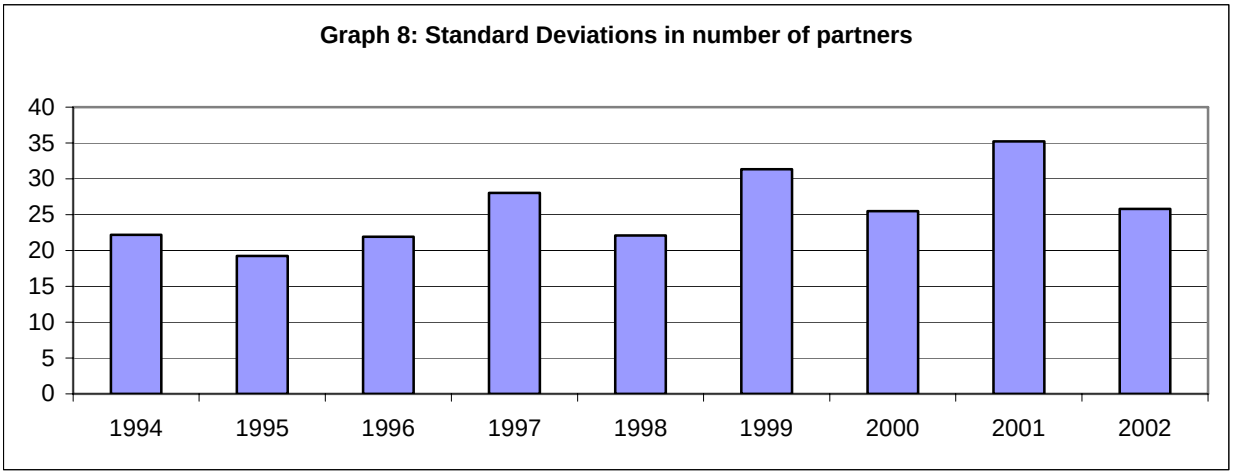
changes within that group: since there are 88-90% of testers initially and the proportion of testers remains stable (except for a slight change in 2001-2), at the very most I only have 11-13% of new members in the test group after 1996 and I would need an equal number of former testers becoming non-testers, 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, 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 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 fact, I find some characteristics of the no-test group, before and after 1996, that differ from the test group. 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. The proportion of California residents is 70% before and after 1996 in the no-test group; it is 78% in the test group before 1996, and 80% afterwards. The proportion of San Francisco residents is 60% before 1996, in the no-test group, and 62% afterwards; it is 68% in the test group before 1996 and 70% afterwards (and in both cases the testers have lived in San Francisco for a longer period of time). Respondents in the no-test group are younger in both periods — by two years before 1996, and by four years

afterwards. Respondents in the no-test group are less educated, by less than a month before 1996, and by less than three months afterwards. They are also less likely to be White — by 4% before 1996, and by 8% afterwards.[25] The cumulative evidence points toward viewing the composition of the populations of testers and non-testers as roughly stable throughout the period. Consequently, the most satisfying interpretation is the third one. In other words, Table 5 (combined with Table 3) does support Hypothesis (3). Again, even if prevalence does not increase, the perceptions of a higher risk level may have convinced the non-testers to decrease their exposures.

The argument of polarization is also consistent with the evolution of the variance in the number of partners (Graph 8): from an average (and quasi-constant) standard deviation of 21 in 1994-96, it moves to an average standard deviation of 28 in 1997-2002. Between 1996 and 1997 alone, the increase is from 22 to 28, that is, +27%.



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

[25] Comparisons of median Household Incomes are inconclusive.

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.[26]

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 risky sexual behavior and testing. In this paper, I develop and empirically test a simple 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 to take 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

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

through UAI while non-testers decrease their number of partners. I performed an Anderson (2004) polarization test to confirm this result non-parametrically. The composition of the two groups seems unchanged over time.

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 HAART 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, and given the little change in the composition of the testing and non-testing groups, 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- HAART 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 HAART-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 HAART 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 epidemiological response to therapeutic improvements. If the unintended consequence of better treatments is to increase the spread of the epidemic, as suggested by Katz *et al*. (2002) and Goldman, Lakdawalla and Sood (2004), the answer may be as a shift of R&D resources towards effective vaccine rather than treatments, or increased prevention programs. Alternatively, if HAART has 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). Finally, Philipson, Mechoulan and Jena (2006) show that the direct and external welfare benefits obtained through HAART dominate the cost of the R&D investment that led to it. This suggests that if HAART is prevalence neutral, which remains to be established, its positive welfare effects are unambiguous.

Bibliography

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 Sally M., Hayley B. Gershengorn and Robert 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) "ARV 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 HAART," *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 ARV 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 ARV 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 Economique* 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

Goldman, D., D. Lakdawalla and N. Sood (2004) “HIV Breakthroughs and Risky Sexual Behavior,” NBER Working Paper 10516

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 ARV 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 ARV 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.

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 ARV 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.

Table 1: Data Description and Summary Statistics

Variables	# Obs	Mean	Std. Dev	Min	Max
Dependent Variables					
d_test (=1 if Ever tested in last 6 months)	39,884	0.904	-	0	1
d_uai (=1 if Ever practiced UAI in last 6 months)	48,888	0.257	-	0	1
numbers (Number of partners in last six months)	48,060	8.414	25.13	0	999
d_intentions (=1 if Intention to practice safe sex)	22,161	0.93	-	0	1
d_vaginalsex (=1 if Sex with a woman in the last six months)	48,189	0.092	-	0	1
Independent Variables					
Age	48,598	33.478	9.334	15	99
zipcode	44,210	-	-	00600	99941
Sanfran (How long have you lived in San Francisco?)	26,439	8	9.6	0	81
d_answ (=1 if answered a STOP AIDS questionnaire before)	48,112	0.264	-	0	1
Time (Time trend)	48,888	0.396	-	0	1
d_caucas (= 1 if White)	48,888	0.666	-	0	1
d_doneit (=1 if ever attended a STOP AIDS meeting)	39,677	0.179	-	0	1
d_after (= 1 after 1996, 0 before 1997)	48,888	0.57	-	0	1
d_bay (= 1 if resident of San Francisco Bay Area)	48,880	0.693	-	0	1
d_cal (= 1 if resident of California)	48,880	0.791	-	0	1
med_edu (Median education in Zip Code)	42,368	15,837	1.589	9	20
med_HHI (Median Houshold Income in Zip Code)	42,352	34,450	10,898	4,999	129,654
d_result (=1 if HIV+, July 1997 onward; sample of testers)	19,512	0.14	-	0	1

TABLE 2: PROBIT REGRESSION WITH DEPENDENT VARIABLE: d_TEST

	(1)	(2)	(3)	(4)	(5)	(6)
d_after	**-0.009**	**-0.002**	**0.002**	**-0.006**	**-0.013**	**-0.004**
	(0.007)	**(0.007)**	**(0.008)**	**(0.007)**	**(0.007)**	**(0.009)**
Pseudo R^2	0.002	0.051	0.054	0.052	0.045	0.026
# observations	39,884	33,755	28,634	33,755	31,310	28,892

Samples control for: Time, Time^2, Age, Age^2, d_caucasian, d_bay, d_cal, d_answ, Median Household Income, Median Household Education

Model (1) controls for Time, Time^2

Model (2) controls for Time, Time^2 Age, Age^2, d_bay, d_cal, d_caucasian, d_answ, Med_Edu, Med_HHI

Model (3) controls for Time, Time^2, Age, Age^2, d_bay, d_cal, d_caucasian, d_answ, Med_Edu, Med_HHI, d_doneit

Model (4): same as Model (2) with d_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: PROBIT REGRESSIONS WITH DEPENDENT VARIABLE: d_UAI

	non-testers	testers		
	(1)	(2)	(3)	(4)
d_after	**0.016**	**0.08**	**0.058**	**0.055**
	(0.035)	**(0.01)*** **	**(0.012)*** **	**(0.013)*** **
Pseudo R^2	0.011	0.034	0.03	0.027
Adjusted R^2				
# observations	2,951	30,804	28,359	25,941

All samples control for: Time, Time^2, Age, Age^2, d_caucasian, d_bay, d_cal, d_answ, Median Household Income, Median Household Education

Sample for models (1): **non-testers**

Samples for models (2)-(4) : **testers**

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: d_UAI

	(1)	(2)	(3)	(4)	(5)	(6)
d_caucasian × d_after		0.037	0.038	0.036	0.042	0.03
		(0.009)***	(0.017)**	(0.017)**	(0.017)**	(0.018)*
d_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)
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, d_caucasian, d_bay, d_cal, d_answ, Median Household Income, Median Household Education

Model (3) contains controls for: d_caucasian × Time, d_caucasian × Time^2

Model (4) contains controls for: d_caucasian × d_answ, d_caucasian × age, d_caucasian × age^2, d_caucasian × d_bay, d_caucasian × d_cal, d_caucasian × median education, d_caucasian × Time, d_caucasian × 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 d_vaginalsex, d_vaginalsex × d_caucasian does not affect results

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

	(1)	(2)	(3)	(4)	(5)	(6)
d_vaginalsex × d_after		-0.054	-0.097	-0.095	-0.088	-0.086
		(0.015)***	(0.029)***	(0.029)***	(0.029)***	(0.03)***
d_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)***
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, d_caucasian, d_bay, d_cal, d_answ, d_vaginalsex, Median Household Income, Median Household Education

Model (3) contains controls for: d_vaginalsex × Time, d_vaginalsex × Time^2.

Model (4) contains controls for: d_vaginalsex × d_answ, d_vaginalsex × age, d_vaginalsex × age^2, d_vaginalsex × d_bay, d_vaginalsex × d_cal, d_vaginalsex × d_caucasian, d_vaginalsex × Time, d_vaginalsex × 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 AND PROBIT REGRESSIONS WITH DEPENDENT VARIABLE:

	Numbers				**d_Syn = d_{Numbers<2}×(1-d_UAI)**			
	Non-testers	testers			Non-testers	testers		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS				**Probit**			
d_after	-4.82	0.769	-0.194	-0.327	0.092	-0.011	-0.001	0.008
	(1.57)***	(0.69)	(0.7)	(0.732)	(0.04)**	(0.012)	(0.12)	(0.013)
Pseudo R^2					0.016	0.018	0.017	0.016
R^2	0.022	0.008	0.007	0.008				
# observations	2,905	30,432	28,027	25,643	2,905	30,432	28,027	25,643

All regressions contain controls for: Time, Time^2, Age, Age^2, d_caucasian, d_bay, d_cal, d_answ, Median Household Income, Median Household 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.