Working Paper Series
Health Economics Series No. 2016-03

Labor Markets in Statistics: The Subject Supply Effect in Medical R&D

Anup Malani, Tomas J. Philipson

February 29, 2016

Keywords: Medical R&D, Research & Development, Health Economics

JEL Codes: I1, I11

Becker Friedman Institute
for Research in Economics

Contact:
773.702.5599
bfi@uchicago.edu
bfi.uchicago.edu

Labor Markets in Statistics:

The Subject Supply Effect in Medical R&D

by

Anup Malani

and

Tomas J. Philipson[1]

March 1, 2016

[1] The University of Chicago Law School and The University of Chicago Harris School of Public Policy. We thank Leila Agha, Gary Becker, Tatyana Deryugina, Einer Elhauge, Jose Fernandez, Dana Goldman, Eric Helland, William Hubbard, Darius Lakdawalla, Richard Miller, Kevin Murphy, Julian Reif, Seth Seabury, Neeraj Sood, Tony Tse, Heidi Williams; workshop and conference participants at the University of Southern California, University of Chicago, Harvard Law School, NBER Summer Institute, the Association Lecture at The Southern Economic Association, HEC Montreal, AHEC 2012 Meeting for helpful comments; and Ilya Beylin, Mete Karakaya, Nate Reid, and Christopher Whaley for research assistance. Malani acknowledges financial support from the Samuel J. Kersten Faculty Fund and the Microsoft Fund at the University of Chicago Law School and Philipson acknowledges support from the Biotechnology Industry Organization and the George Stigler Center for the Study of the Economy and the State at the University of Chicago.

ABSTRACT

Medical research and development (R&D) differs from other R&D because of a unique linkage between output and input markets for medical products: potential consumers of existing medical products are also potential subjects in clinical trials required to develop new products. Therefore, an increase in the quality or reduction in the price of an existing treatment reduces the incentive of patients to participate in trials of new treatments. We provide evidence of this linkage, which we label the "subject supply effect," by showing that a breakthrough HIV/AIDs treatment led to a sharp drop in the supply of trial subjects after the introduction of the treatment in 1996. The subject supply effect has important positive implications for how policies such as recent insurance expansions affect the rate of medical R&D and normative implications for whether subjects ought to be compensated for enrolling in clinical trials, an ethically controversial practice.

Medical research and development (R&D) plays a central role in health care economics. On the one hand, it is an important driver of the benefits of health care. Improvements in health have been a major contributor to gains in overall economic welfare during the last century (Murphy & Topel 2006) as well to the reduction in global inequality (Becker et al 2005). A significant part of these gains has been attributed to medical R&D, including improvements in medical knowledge, procedures, drugs, biologics, and devices (e.g., Cutler & Kadiyala 2003, Lichtenberg 2003, Cutler et al. 2007). On the other hand, innovation is an important driver of health care costs. The medical consumer price index (CPI) has recently been 1.5% higher than the consumer CPI (BLS 2012) and health expenditures increased from 8% of GDP in 1990 to 16% in 2008. Newhouse (1992) and others have attributed much of this growth in costs to technology.

Given the impact of medical innovation on the health care system, there is considerable attention given to the high and growing costs of developing medical R&D. DiMasi et al. (2003) estimate that the cost of bringing a new drug to market is over $800 million. More recent estimates suggest these development costs may be as high as $1.7 billion per drug (Adams and Brantner 2006, Brichet and Cohen 2007). An important gap in this literature, however, is the absence of positive analysis of what drives the costs of medical R&D.

Under the traditional economic view, the private benefits of R&D are driven by future variable profits in the output market and the costs of R&D are driven by input supply, for example, the quantity and location of research and development talent. We argue in Section I that an important difference between medical and non-medical R&D, however, is a unique linkage between medical output markets and medical R&D inputs. This linkage is driven by the twin facts that medical R&D requires clinical trials on human subjects[2] and that individuals who can serve as human subjects are also potential consumers of medical products as patients in the output market.[3] In other words, input supply for medical R&D overlaps with output demand for medical products. The central implication of this overlap, which we term the "subject supply effect", is that improvements in output markets, e.g., an increase in quality or

[2] Regulatory approval requires clinical trials. However, it is possible that medical companies would conduct trials even in their absence, e.g., to demonstrate to consumers their experience products work. Thus, we do not assert that regulatory requirements are the reason there is a link between development and output markets for medical products.

[3] Clinical trials do not allow human subjects to participate if they are also consuming conventional care on their own outside of trials. Consumption of outside treatment confounds causal inferences between treatment in the trial and health outcomes. Sometimes clinical trials use conventional treatment as a control. Even in these cases, subjects assigned to experimental treatment are not allowed to use conventional treatment outside the trial, again to protect causal inferences.

reduction in price of conventional medical care, make patients more reluctant to enroll as subjects in clinical trials.

In Section II, we provide evidence of the subject supply effect in the context of HIV/AIDS, specifically the introduction of highly active antiretroviral therapy ("HAART") in 1996. HIV is a virus that, when it reaches a critical level of concentration in the bloodstream, triggers AIDS, a condition that cripples the immune system, rendering a person defenseless against secondary infections by bacteria and fungi. HAART is a cocktail of so-called antiviral medications that effectively slowed the reproduction of the HIV virus and prevented AIDS. Using longitudinal data from the Multicenter AIDS Cohort Study on over 1000 gay and bisexual men in 4 U.S. cities, we show that HAART dramatically reduced participation in trials for other, potentially innovative antiviral medications. Our basic result is summarized by the cross-tabulation in Table 1, which reports the fraction of sample HIV patients participating in clinical trials before and after the introduction of HAART in 1996. In Column A we find that, among all HIV positive subjects, participation in trials of antiviral drugs (labelled "primary drugs") fell from 11.5 to 7.3% after introduction of HAART. Among the subset of HIV-positive patients actually using such primary drugs the reduction was more dramatic: from 20.1 to 9.1%.[4]

To control for unrelated changes in trial participation during the sample period, we use time trends and a control group of HIV patients who consume non-antiviral medications. Whereas anti-viral or primary drugs target replication of the HIV virus, aiming to prevent AIDS, non-antiviral drugs (labelled secondary drugs) target the secondary infections that take advantage of AIDS.[5] HAART was a dramatically more effective primary drug. HIV patients often consume both primary and secondary drugs, though the trials for the two drugs are often different. While HAART reduced the number of HIV patients with AIDS and thus overall demand for secondary drugs, it did not change the relative quality of secondary drugs in the output market versus in trials and thus should not affect enrollment rates in secondary drug trials. Column B of Table 1 shows that, among HIV-positive patients, participation in secondary drug trials fell only 2 percentage points and, among secondary drug users, participation fell 5.5 percentage points after HAART. This yields a difference in difference estimate of a reduction in trial

[4] The change in the participation rate among HIV positive subjects captures both changes at the margin of consuming a primary drug and the margin of consuming that drug in a trial due to HAART. The change in the rate among primary drug users only captures the latter margin. The former sample includes subjects in the denominator who are never eligible for trials because they would not consume drugs even with HAART. Changes in the latter sample conflate changes at the trial participation margin among a fixed group of subjects and changes in the composition of subjects. Neither sample is perfect, but both are informative.

[5] In short, primary drugs address the cause of AIDS while secondary drugs tackle the symptoms of AIDS.

participation of 2.2 percentage points among HIV-positive subjects and a 6.3 percentage points among primary drug users.[6]

An important question is whether the observed decline in trial participation was driven by a contraction in the supply of subjects, a dynamic that this paper highlights, or by the fact that HAART reduced demand for further R&D and thus for subjects. We demonstrate that supply was the more important factor in three ways. First, we document that the decline in trial participation after HAART was largely driven by exit from ongoing trials. Because ethical rules prohibit the closure of ongoing trials by drug companies, an increase in exits signal a reduction in the supply of subjects rather than in the demand for subjects. Second, we show that funding for R&D in HIV/AIDS did not diminish after the introduction of HAART. Third, we observe that, because ethical rules also cap payment of trial subjects, equilibrium quantity is dictated by subject supply rather than demand.

The subject supply effect has both important positive and normative implications, as we explain in Sections **Error! Reference source not found.** and **Error! Reference source not found.**, respectively. From a positive perspective, failure to account for the subject supply effect means that one will incorrectly estimate the rate and thus level of medical innovation over time. The subject supply effect implies that an improvement in conventional care does not merely reduce demand for innovation, it actually increases the cost of innovation. Because ethical rules cap payment of trial subjects, this cost manifests as an increase in the time it takes to complete medical trials. The resulting delay in development has two effects on level of innovation. First, the delay increases the opportunity cost of developing medical products. Second, the delay reduces the return to medical innovation. Prolonging development time delays the start of product sales and thus profits. Moreover, because medical products are patented prior to development, development delays shorten effective patent life. The latter consequence of delay has been overlooked by prior estimates of the costs of drug development (e.g., DiMasi et al. 2003), which equated the cost of delay solely with the opportunity cost of capital.

The subject supply effect also implies that health care policies that operate on output markets have dual effects on the net returns to medical innovation. Conventional analysis suggest that health care reforms that expand insurance coverage increase the demand for innovation and thus innovative returns, while policies that impose price controls reduce demand for innovation and thus innovative

[6] In Table 1, the primary (secondary) drug trial participation rate is calculated from a sample of HIV-positive individuals who were consuming primary (secondary) drugs either in the output market or trials. Among all HIV-positive individuals, whether they consumed primary (secondary) drugs or not, trial participation in primary (secondary) drug trials fell 4.1 (2) percentage points, also highly significant. The difference in difference estimate is 2.1 percentage points ($p = 0.0177$).

returns. The subject supply effect suggests both policies also improve patients' prospects in the output market and reduce their incentive to participate in clinical trials. The resulting delay in development increases the cost of and reduces the return to innovation. Thus the subject supply effect mitigates the positive innovation effects from insurance expansion, while it compounds the negative innovation effects from price controls.

Previous estimates of the effect of health care policies on product introductions (Acemoglu et al. 2006, Finkelstein 2004) conflate these effects, underestimating the demand-side effect on innovation[7] of the policies that increase healthcare demand because they do not filter out the development delay effects from those policies.[8] Moreover, papers that examine patenting or trials rather than product introductions (Blume-Kohout and Sood 2008 and Clemens 2012) may capture only the innovation demand effects of policies and ignore the subject supply effects.

The subject supply effect also has important welfare implications. As we alluded above, an important feature that distinguishes the labor market for human research subjects from other labor markets is that wages are capped by ethical rules. Specifically, bioethicists – and thus Institutional Review Boards that regulate trials – frown upon research compensation because it may encourage subjects to enroll in trials for non-medical reasons.[9] Because of these ethical wage caps, we argue that observed delays in development due to the subject supply effect are likely to be inefficient. According to conventional economic analysis, improvements in output markets, such as an improvement in conventional treatment, may reduce innovation, but this is likely to be efficient because those improvements reduced the demand for innovation. Likewise, one could argue that improvements in output markets that affect innovation through the subject supply effect are also efficient because they reflect an increase in the opportunity cost to patients of enrolling in trials. However, as a result of ethical wage caps, subjects willing to participate in trials are rationed among trials by queuing rather than price.

[7] Consistent with the demand side analysis, Acemoglu and Linn (2004) find that diseases with larger markets, as measured by incidence, is associated with higher innovation.

[8] Finkelstein (2004) finds that vaccine promotion policies increased vaccine approvals 7+ years later. This is not inconsistent with the development delay effect we highlight because she does not examine whether trials took longer. Acemoglu and Linn (2004) find larger markets as measured by disease prevalence are associated with greater innovation – but that conflates a positive demand for innovation and a larger supply of subjects, both of which increase innovation.

[9] As a result, Institutional Review Boards (IRB) limit compensation for anything beyond incidental expenses, such as the cost of transportation to a trial or medical treatment for side effects suffered during the trial. Compensation for time is strongly discouraged. As a result, the most comprehensive survey to date found that the median payment is under $200 per subject and the maximum is $2000 (Grady et al. 2004). Indeed, surveys suggest that, in practice, IRB's do not even allow trials to fully compensate for non-time expenditures (Ripley et al. 2010).

Thus, a component of the development delay induced by the subject supply effect is the misallocation of subjects.

This inefficiency is compounded by the fact that research subjects confer positive external effects on future patients by providing them information on treatment effects. Medical producers can help internalize this externality by paying research subjects higher wages financed by expected revenue from future sales. Because these firms do not capture the full surplus from those sales, however, private wages cannot fully solve the externality. There is a welfare argument for public subsidies to research participants, but those too are barred by ethical wage caps.

The remainder of the paper may be outlined as follows. Section I presents a simple model of how changes in the quality or price of conventional care affect development times when subjects are rationed among trials pro rata rather than by price. Section II presents evidence consistent with our model showing that new HIV/AIDS treatments introduced in the mid-1990s significantly reduced trial participation and this effect is driven substantially by an increase in quality of conventional care that reduced the supply of subjects to trials. Section III discusses the effect of delays in development on innovative returns and the equilibrium level of development and development times when free entry dissipates profits. Section III discusses optimal subject compensation and the efficiency effect of ethical wage caps. Section V concludes with a discussion other positive and normative implications of the subject supply effect for medical products, including the implications for location of innovation and a re-examination of Pareto optimal sample sizes (Philipson 1999).

I. Development and changes in output markets

We begin with a simple theoretical model of the tradeoff that patients face when choosing between consumption of conventional care or participation in the trial of an experimental therapy. From this, we derive the supply of subjects. We then examine how output market conditions affects development times. Because the contribution of this paper is to highlight subject supply effects, we initially hold demand for subjects constant and assume that, because of binding ethical wage caps, subjects are rationed among trials pro rata. We relax these conditions over the course of the paper.

A. The supply of subjects to clinical trials

The recruitment of trial participants comes from a stock (prevalence) of a disease in a given period, N_t. This stock rises with the entry of new cases of the disease (incidence), b, and falls with the exit of existing cases due to recovery or death at rate m:

$$N_{t+1} = b + (1 - m)N_t \quad (1)$$

This implies that the steady state level of the stock of the disease rises with disease incidence and falls with the cure or mortality rate: $N = b/m$.

Only a fraction of patients with the disease will be willing to participate in clinical research. The decision to participate depends on whether the utility from access to the experimental therapy is greater than that from conventional treatment. The conventional treatment offers per-period implicit utility, $U(q, p)$, that increases in the quality of care, q (synonomous with the treatment effect, health outcome, or "effectiveness" of care), and decreases in the price, p (the full price, including premium and copay). If a patient enrolls in a clinical trial of experimental therapy, he obtains an uncertain implicit utility, $U(q^e, p^e)$. The quality of experimental care, q^e, is a random variable unknown at the time of entering the trial[10]. The uncertain quality of experimental care may be the product of uncertainty about the effects of the experimental therapy being studied or of the experimental design, e.g., the random assignment of treatments. The price of experimental care, p^e, is potentially zero if treatment is fully subsidized by the trial. We assume p^E embeds any subsidy provided for the control treatment, even if the control is conventional treatment, i.e., the trial uses an active control. Importantly, we assume there are binding ethical regulations that prohibit monetary compensation of subjects and thus a negative price for experimental care.

A patient participates in a trial if the expected utility from the trial exceeds that of being on conventional care,

$$E_F[U(q^e, p^e)] \geq U(q, p) \quad (2)$$

where the expectation is taken over the cumulative distribution function, $F(q^e)$, describing a patient's beliefs about the quality of the experimental therapy.[11] We assume heterogeneity among patients, e.g., different beliefs about the quality of experimental care or degrees of risk-aversion, gives rise to a differentiable supply function, $s(q, p)$, that is decreasing in the quality of conventional care and

[10] Conventional care may also, from the patient's perspective, of course be of uncertain quality.

[11] This selection model is similar to that in Philipson (1997), Chan & Hamilton (2006), and Malani (2008). A more general version may be found in Chassang, Miquel & Snowberg (2012). A less formal version is presented by Dunn & Gordon (2005).

increasing in its price: $s_q \leq 0, s_p \geq 0$.[12] This function depicts the fraction of the stock of patients willing to participate in the trial, given existing output market conditions.[13]

B. Changes in output markets and development times

Given a stock of patients with a disease, N, and a steady state a fraction, s, that participates in trials, we can derive the supply of subjects Ns. We assume that, because of binding ethical wage caps, clinical trial participants are allocated evenly across existing trials. For simplicity, we also assume trials are of identical sample size and design. Given M trials and a flow of subjects, f, into each particular trial, demand for subjects is Mf and the equilibrium condition in the subject market is that demand equal supply:

$$Mf = Ns \leftrightarrow Mf/N = s \tag{3}$$

Normalizing each side by the stock of patients, the condition can also be expressed in rates.

Suppose a trial seeks to recruit a sample size of n patients, assumed to be determined by considerations of statistical power. The development time or duration, T, it takes to complete this trial is the number of periods required to recruit n subjects given a flow of patients, f, enrolling each period:

$$Tf = n \rightarrow T = \frac{n}{f} = \frac{nM}{Ns} = \frac{m}{b}\frac{nM}{s(q,p)} \tag{4}$$

Plugging in equation (3) and the steady state number of patients yields mechanical relationship between various factors and development times. Increases in required sample size or the number of competing trials increase development times, as more time is needed to attain the desired sample size for a given trial.[14] Trial durations fall in the prevalence of a disease since it implies that a larger number of subjects are eligible for recruitment each period. As a consequence, development times fall with the incidence of the disease and rise with the rate or exist or mortality rate out of the disease.[15]

---

[12] Participants may also include not just those who have high expectations of experimental treatment, but also individuals who did not respond to conventional treatment. Let $r(q)$ denote the response rate on conventional care as a function of quality of that care, where $r'(q) > 0$. Then total participation is given by $(1-s)(1-r) + s$. The total participation rate still falls in quality, as we shall show below. Higher quality not only reduces direct participation, but also the ability to recruit from non-responders to conventional care.

[13] As an illustrative example, consider a standard discrete choice framework in which utility is determined by observed or studied efficacy or health outcomes q, price p, and unobserved or non-studied outcomes ϵ such as, unmeasured side effects or costs of compliance: $U + \epsilon = \alpha q - \beta p + \epsilon$. Under the standard extreme value distribution for ϵ, this would give rise to a participation rate $s(p,q) = \exp\{E_F[U(q^E, p^E)]\}/(\exp\{U(q,p)\} + \exp\{E_F[U(q^E, p^E)]\})$.

[14] A number of services have emerged to help match supply (patients) and demand (trials). Many of these are web-based, such as the NIH's *www.clinicaltrials.gov* database.

[15] A number of predictions flow from this relationship. For example, some oncology trials may take long to complete because patients die quickly and are only available for observation during a short time-window.

Changes in output market, i.e., the value of conventional care, will also affect development times. Holding constant demand for subjects, i.e., the number of trials M, an increase in the quality or fall in the price of conventional care lengthens the time required to recruit subjects into trials for new innovations:

$$T_q = T_s s_q > 0,\ T_p = T_s s_p < 0 \tag{5}$$

Thus policy changes that alter the quality or price of available care will affect development incentives. For example, coverage expansions that lower the out-of-pocket price of conventional care, price controls that reduce overall prices, and comparative effectiveness regulations that improve the quality of conventional care each reduces patients' incentives to enroll in trials and tend to delay innovation.

C. Incorporating changes in demand

In reality, changes in conventional care can also shift the demand for subjects. Because innovation competes with conventional care, demand for research is decreasing in the quality ($M_q < 0$) and increasing in the price ($M_p > 0$) of that care. The full effects of changes in conventional care on development times are ambiguous because supply and demand effects may flow in opposite directions:

$$T_q = T_s s_q + T_M M_q,\ T_p = T_s s_p + T_M M_p \tag{6}$$

For example, an increase in the quality of conventional care can both discourage patients from participating in trials, increasing trial duration, and also decrease the number of trials, reducing trial duration. The net effect of quality and price on development times depends on the relative magnitude of effects on supply versus demand.

Even if improvements in conventional do not induce changes in the demand for R&D, they might induce investigators, who are barred by ethical rules from raising subject wages, to modify the design of experiments to make them more attractive to patients. Because these are second-order effects, we explore them only in the appendix.

II. Evidence of the subject supply effect for AIDS innovations

In this section we examine empirical evidence for the subject supply effect in the case of AIDS treatment. We employ the introduction of Highly Active Antiretroviral Treatment (HAART) – a breakthrough AIDS treatment – in 1996 as a change in conventional treatment. Our aim is to identify, or

Moreover, the relationship suggests that so-called "me-too" innovations, for which experimental care is a close substitute to conventional care, may face recruitment difficulties and longer development times because they do not offer substantial gains over the conventional treatment outside of the trial.

at least sign, the effect of the implied change in output quality and price on total supply of subjects in HIV/AIDS clinical trials, i.e., validate the predictions $\partial(Ns)/\partial q < 0$ and $\partial(Ns)/\partial p > 0$, and participation rates, i.e., validate the predictions $s_q < 0$ and $s_p > 0$. We do this in three steps

First, we present unconditional and reduced-form estimates of how total enrollment and participation rates in HIV/AIDS trials changed after the introduction of HAART. These estimates measure changes in equilibrium participation levels or rates. Second, we argue that the decline in participation after HAART reflected an inward shift in the supply curve of subjects (Ns) and not just a downward shift in demand curve (Mf) for further HIV/AIDS research. Third, because HAART increased both the quality and price of conventional care, we use patients' insured status, which regulates the price effects of HAART, to decompose changes in trial participation from quality and price. Because the price effect of HAART offset the reduction in participation due to the quality effect of HAART, we likely underestimate the purely quality-driven effects of HAART on subject supply.[16]

A. Background on HIV/AIDS and HAART

HIV is a virus that infects and disables CD4 T-cells, an important component of the human immune system. When the viral replication reduces a patient's CD4 count below 200/mm^3, the patient's immune system is compromised, a condition called AIDS, and the patient becomes acutely vulnerable to non-HIV, secondary infections (U.S. Centers for Disease Control & Prevention 1993).[17]

AIDS is treated in two ways. First, the patient is given antiviral medications to suppress replication of the HIV virus. Effective antiviral therapy can prevent an HIV patient from progressing to AIDS but does not eliminate HIV infection. Therefore, we refer to antiviral therapy as a *primary* AIDS drug.

Second, the patient is given various non-antiviral medications, such as antibiotics, steroids and antifungals, to either treat or prevent secondary infections. These may be administered either to patients with HIV or those who have progressed to AIDS, but are more critical for AIDS patients. Because these drugs target the side-effects of AIDS, rather than AIDS itself, and are more intensely used by AIDS patients, we shall refer to them as *secondary* AIDS drugs. Note that patients may be on both primary

[16] In the appendix, we examine how HAART affected the design of clinical trials. One way trials can compete for subjects in the presence of pecuniary wage caps is to modify trial design to make trials more attractive to subjects. We show that, while HAART reduced enrollment in active-controlled trials by one-half, it eliminated all enrollment in placebo-controlled trials. Because the design effect of HAART also offsets the reduction in participation due to the quality effect of HAART, it is another reason we likely underestimate the purely quality-driven effects of HAART on subject supply.

[17] CD4 count is not an exclusive indicator for AIDS progression. Evidence of a compromised immune system, specifically the presence of secondary infections, is also used to diagnose an HIV patient as having AIDS.

and secondary AIDS drugs at the same time; indeed this is true for the vast majority of AIDS drugs users in the data we used. For these patients, secondary drugs are a safety net in case they are non-responsive to primary drugs.

HAART is the label for a regimen of antiviral medications that, together, substantially slows the progression of HIV into AIDS. This regimen typically includes two nucleoside reverse transcriptase inhibitors (NRTI) and either one protease inhibitor or one non-nucleoside reverse transcriptase inhibitors (non-NRTIs). HAART was first introduced to the market in 1996. Although the first NRTI – zidovudine (AZT) – was approved by the U.S. Food and Drug Administration (FDA) in 1987, NRTIs on their own were unable to affect progression from HIV to AIDS. In December 1995, the FDA approved the first protease inhibitor and, in July 1996, scientists demonstrated that the combination of NRTIs and a protease inhibitor was highly effective at controlling AIDS progression (Gulick et al. 1997; Hammer et al. 1997, Bartlett 2006). Later in 1996, the FDA approved a new class of antivirals, the non-NRTI. These also were shown to be highly effective at controlling the HIV virus when used with NRTI's. Together, these combinations of primary AIDS drugs, which quickly acquired the name "HAART," reduced AIDS deaths by a half before the end of the decade (Palella et al. 1998, Chaisson & Moore 1999). In 1997, HAART was pronounced the standard of care by the USA panel of the International AIDS Society, the guideline setting body for HIV treatment (Carpenter et al. 1997).

At the same time HAART was approved there was a second, smaller innovation that also improved conventional care for HIV/AIDS patients: real-time monitoring of viral load (Bartlett 2006, Murray 2011). Our analysis is formally a test of both innovations on trial participation. Howeer, real-time monitoring has not been shown to have a substantial impact on viral loads as it only tracked those loads.[18] To simplify our exposition, we refer to both innovations as HAART because it is the more

[18] In 1997, there was a technical change in FDA regulation of HIV drugs trials related to viral load monitoring, but it did not affect the way HIV trials were conducted before or after HAART (Murray 2011). Before 1997, viral loads (HIV-RNA counts) could not be used as a trial outcome to support application for ordinary drug approval. However, they could be used to support applications for accelerated approval. The hitch was that, after the drug was approved, the drug company had to conduct studies to demonstrate that the surrogate endpoint was valid, i.e., that viral loads were indeed correlated with progression from HIV to AIDS as measured by CD4 counts. Despite this hitch, all HIV drugs that received accelerated approval eventually received regular approval. By 1997, a number of studies, including the one using the data from our data set, MACS, validated HIV-RNA counts as an endpoint. Therefore, the FDA allowed HIV-RNA trials to support applications for ordinary drug approval, eliminating the requirement that the drug company do any further post-approval studied on HIV-RNA. However, this does not affect our predictions or empirical analysis because trials for primary drugs used the same endpoints – HIV-RNA and CD4 counts – before and after HAART was approved in 1996. Thus, there was no change in the ease of conducting development phase trials during our sample.

significant of the two innovations. However, we will attempt to distinguish the impact of the HAART drug regimen and viral monitoring when we examine trial exit and entry.

Our test of the effect of output markets on trial participation is motivated by the fact that the approval of HAART in 1996 affected the quality and price of conventional care. Specifically, the prediction we test is that HAART, by improving quality of conventional care, reduced trial participation; moreover, for uninsured persons, HAART may also have increased the price of conventional care, lowering trial participation for insured persons relative to uninsured persons. In much of our empirical analysis we define the pre-HAART period as ending in 1995 and the post-HAART period beginning in 1997. We drop the year 1996 because HAART was introduced and spread gradually over the course of that year.

B. Data

Our empirical analysis employs data from the Multicenter Aids Cohort Study (MACS). Given the relatively high prevalence of HIV in the gay community, this study tracked 6,972 homosexual and bisexual men in four cities (Baltimore, Chicago, Pittsburgh and Los Angeles) longitudinally during the period 1984 to 2005.[19] We first isolate a subsample of 1,725 unique subjects with HIV observed in the period 1990-2005.[20] This is a sub-sample that was observed both before and after the introduction of HAART in 1996 and for which we have reliable data on HIV-positive status and trial participation.

Subjects in the study were asked to visit a study site two times each year. Each visit typically included a series of medical exams, a survey of the subject's medication use, and a survey of his employment and health insurance status. The medication survey asked whether the subject took any primary or secondary AIDS drugs and whether he obtained that drug in a clinical trial.

We conduct our empirical analysis on either the entire sample of HIV-positive patients or a subset of patients who used HIV drugs. In the HIV-positive sample, we only include individuals the years after they are diagnosed with HIV. In the HIV drug sample, we only include individuals the years they report using a primary drug or a secondary drug, as appropriate. Among HIV-positive patients, 1,435 individuals took primary drugs and 1,267 took secondary drugs at some point during the sample period,

[19] More information on the MACS study is available at http://www.statepi.jhsph.edu/macs/macs.html. Not all 6,972 subjects were observed each year. Instead, subjects were enrolled in 3 large waves: in 1984-1985, in 1987-1991, and in 2001-2003. Only subjects enrolled in the first two waves are observed prior to HAART. Moreover, half of the confirmed HIV-negative subjects were administratively censored in 1993. We omitted these subjects because they are not observed after HAART.

[20] We begin our analysis in 1990 rather than 1984 because MACS did not start asking about trial participation until 1987. Moreover, the trial participation question changed twice between 1987 and 1990, but remained the same between 1990 and 2005.

whether the source of drug supply was inside or outside a trial. Of all HIV-positive individuals, 216 took only primary drugs, 48 took only secondary drugs, and 1,219 took both a primary drug and a secondary drug at some point.

We primarily focus on participation in trials that test primary drugs; however, we will also examine participation in secondary drug trials as a control for unobserved trends in participation. We determine participation in a primary (secondary) drug from survey questions that ask whether the primary (secondary) drug that an individual consumed was obtained in a trial.

C. Descriptive statistics

Table 2 provides descriptive statistics for all subjects with HIV. Statistics are calculated at the subject-year level and weighted so that each subject has an identical weight. The data are separately described for the years prior to HAART (1990-1995) and the years after HAART (1997-2005) to provide an unconditional assessment of the effect of HAART. Roughly the same fraction – roughly 75% – of subjects has some medical insurance before and after HAART. However, there is a substantial change in the sources of insurance following HAART. While private insurance coverage falls from 67% to 49% after HAART, government coverage rises from 14% to 33% after HAART. An important program for government insurance after HAART is, surprisingly, Medicare. This is because Medicare covers not just the elderly, but also the long-term disabled population. Our sample has relatively low income. Only 57% are employed after HAART and the median income is between $20,000 and $29,000.

D. Evidence on the unconditional impact of HAART

Before we present our regression results, we offer some basic graphical evidence of the effect of HAART on the quality of conventional treatment and on trial participation. Figure 1 documents the changing pattern of primary AIDS drug use over time. Panel A gives a breakdown of primary drug use by class of antiviral drugs. While use of NRTIs such as AZT are constant at nearly 100% throughout the 1990s, there is a sharp rise in use of protease inhibitors (PI) and non-NRTIs (NNRTI) after they are introduced in December 1995 and the end of 1996, respectively. At the same time we see a sharp drop in use of other classes of antiviral drugs.[21] For context, Table 2 notes that the fraction of HIV-positive subjects using primary drugs increased from 44% to 64% after HAART, while the fraction of these subjects using secondary drugs is constant at roughly 32% over time. Panel B of Figure 1 gives a breakdown of how the usage of primary drug regimens or combinations changed. Note that, although an individual may be on multiple primary drugs in panel A, he can only be on one of the three primary

[21] These are usually antiviral drugs that have proven effective against other sexually transmitted diseases, such as hepatitis. None have proven effective against HIV/AIDS.

drug regimens (NRTI only, HAART, and other non-HAART regimen) in Panel B. The main takeaway is that HAART use skyrocketed after 1996, while the use of other regimens fell.[22]

The dramatic rise in usage of HAART after 1996 led to a dramatic improvement in subjects' health. According to Table 2, after HAART the average CD4 count in our sample of HIV positive subjects increased from 466 to 563/mm^3. A higher CD4 count implies better health: an HIV patient with a CD4 count below 200 is considered to have AIDS. Moreover, average viral loads fell from roughly 114,547 to 29,121 copies/ml and a patient is considered to have AIDS with a load above 100,000 copies/ml. Figure 2, which plots CD4 counts and viral loads for subjects on primary drugs, provides even more detail on the trend. According to Panel A, CD4 counts for the median subject on primary drugs declined steadily until 1995, right before the introduction of HAART, after which they improved dramatically. We illustrate the effects for AIDS progression by plotting the CD4 counts for the 25th percentile subject. Until 1995, this subject had AIDS (CD4 count < 200/mm^3), but after HAART, their CD4 counts improved to the point where, by 1997, the 25th percentile subject did not have AIDS. Likewise, Panel B shows that viral loads for the median patient fell dramatically starting in 1996. The effect was even more pronounced for the 75th percentile subject in our sample. Until 1995, this individual had loads above the AIDS threshold. Within a year after HAART, his load fell so that he was no longer at risk.

Finally, we turn to trial participation. According to Panel A (right y-axis) of Figure 3, the total number of subjects in trials hovers around 120 until 1995. After HAART the total falls dramatically to around 70 in 1997 and trickles down to 50 by 2005. These totals correspond to the equilibrium number of subjects in trials, $Mf = Ns$. We can also look at equilibrium participation rates, i.e., $Mf/N = s$. According to Table 2, 9.2% of HIV positive subjects were in clinical trials for primary AIDS drugs prior to HAART. However, after HAART was introduced, only 5.1% were participating in trials.[23] Panel B of Figure 3 illustrates that, behind these changes, there is an even more dramatic change in the trend in trial participation over time. Trial participation increases to nearly 20% among HIV positive subjects in 1995, and then trial participation plummets and eventually settles at 5% by 2005, roughly half the level in

[22] Although the use of NRTI-only regimens had been falling since 1991, much of the decline prior to 1996 was offset by growing use of non-HAART cocktails on the market. These regimens typically combined a NRTI with one of the "other" antiviral drugs depicted in Panel A.

[23] At first blush, the level of trial participation may seem low for both types of drug users. In fact, it is very high relative to the 3% rate of trial participation for patients with diseases other than HIV (e.g., Lara 2001).

1990.[24] The change is even more dramatic among primary drug users where the participation rises to over 30% by 1995 and falls to a quarter of the 1990 rate by 2005.

We can use Figure 3 to unpack the causes of the observed changes in trial participation rates. From Panel A of Figure 3 we can rule out an increase in demand for research subjects, Mf, as the cause of the dramatic rise in trial participation rates before 1996. An increase in demand should increase total subjects in trials, but that is steady before 1996. As between the alternative explanations, an increase in the supply rate s or a decrease in the number of subjects with HIV, Panel B of Figure 3 favors the latter: the number of subjects with HIV fell from over 1400 to just 800 by 1996. The drop is largely due to high AIDS mortality rates prior to 1996. While primary drug utilization rates – a precursor to participation in trials – also fell, the fall was less dramatic, from 700 to 500, largely to due to evidence emerging that AZT, the foremost primary drug prior to 1996, was ineffective at controlling HIV.

The fall in the trial participation rate after HAART could either be due to a reduction in aggregate demand for subjects, Mf, or a reduction in the supply rate, s, of subjects. It cannot be attributed to an increase in the number of subjects with HIV or using primary drugs, N. To rule out changes in aggregate demand, Figure 4 examines the rate of entry and exit into trials. Entry (exit) rates are calculated by taking the number of subjects who enter (exit) trials in a given period and diving by the number of subjects not in (in) trials the previous period. A reduction in demand can only take the form of a decrease in the number of new trials started because ethical rules prevent researchers from closing down ongoing trials and kicking out or "firing" enrolled subjects. Thus changes in demand will be manifest only as decreases in entry. By contrast, a decrease in the supply rate can cause either subjects to exit trials or subjects to enter trials at a lower rate. Figure 4 shows a sharp (short-term) spike in exits in 1996 but only a mild reduction in entrances after HAART. Thus, it is more likely that the decline in the equilibrium trial participation rate immediately after HAART is due to a reduction in the supply rate of subjects, consistent with the prediction of our theory.

E. Regression analysis

1. Empirical strategy

Our graphical evidence on HAART provides only impressionistic evidence on the effect of HAART on participation. Our primary regression strategy for estimating the effect of HAART on the equilibrium, total number of subjects in clinical trials for primary drugs is:

[24] The pre-HAART average participation rate in Table 2 in is somewhat lower than the pre-HAART average in Figure 3 because the two use different weighting schemes. The former gives each subject equal weight.

$$P_t = \beta_1 + \beta_2 POST\_HAART_t + \beta_3 TREND_t + u_t \quad (7)$$

The dependent variable P_t is calculated by summing the total number of subjects reporting they participated in a primary drug trial in year t. The treatment effect is captured by the coefficient on $POST\_HAART_t$, an indicator for the period 1996-2005 or, in some specifications, 1997-2005. In the latter case, we drop the year 1996 from the sample. From Panel A of Figure 3 we see that the total number of subjects in trials is more variable before 1996 than after; therefore we calculate robust standards errors when conducting the regression above. Standard errors are unbiased only if the regression errors are not serially correlated; we report the p-value from Durbin's alternative test to check this condition.

The aggregate numbers conflate composition effects and changes due to individual-level circumstances. Therefore, we supplement with the following individual-level regression analysis:

$$p_{it} = \beta_1 PRE\_HAART_t + \beta_2 POST\_HAART_t + \beta_3 TREND_t + [controls] + e_{it} \quad (8)$$

The dependent variable is an indicator variable p_{it} for whether subject i in year t participated in a clinical trial. The treatment effect is captured by the difference between the coefficient on an indicator for some pre-period before the introduction of HAART and the coefficient on an indicator for some post period after the introduction of HAART. We employ three specifications of the pre- and post-periods to capture short and long term changes in participation rates. One specification, which we label short-run, uses a pre-window of just 1995 and a post window of just 1997. A second, which we label medium-run, uses a pre-window of 1993-1995 and a post window of 1997-1999. The last, which we label long-run, uses a pre-window of 1990-1995 and a post-window of 1997-2005. In all but one regression, we omit 1996 from pre- or post-windows, though we do not drop that year from the data.[25] The one exception is our analysis of trial exits. There we will focus on 1996 because that is year in which we observe a spike in exits in Figure 4.[26] We include a linear trend in the regression above to capture unobserved trends in participation. Thus the treatment effect is, to some extent, identified off a linear trend.[27]

[25] Dropping 1996 from the analysis does not change our results. It mainly changes the coefficient on the constant. That is only used to identify treatment effects when we use a post-window of 1997-2005.

[26] Our regression sample includes data for years other than those that fall in the pre-period or the post-period. The reason is that some subjects may not appear in the data in both the pre-period and in the post-period. However, they will appear in either of those periods plus another year. When we employ individual fixed effects, we will be able to identify the coefficient on the pre-period and the post-period even from data on such subjects. The additional sample size allows us to estimate more coefficients on the pre- and post-periods more precisely.

[27] Although the regression analysis reported in thus paper focuses on the level of trial participation, much of the effect of HAART was manifest through changes in the trend of participation. The level of trial participation depends on subjects' knowledge of the quality of experimental treatments and newly approved treatments. However, it may take time for subjects to learn this in the case of HAART. Although trials of protease inhibitors and NNRTI's began in last 1980s and early 1990s, it took a few years before patients learned about the value of those drugs.

We employ multiple specifications of controls. In specification (1), we included no controls. In specification (2), we included individual fixed effects, so that treatment effects are additionally identified from within-subject changes in participation. In specification (3), we also included time-varying controls for income and CD4 count. The CD4 controls are five indicators for whether cell counts lie in the range 0-199, 200-399, 400-599, and 600-799, respectively. (The above-800 indicator is omitted.) In all specifications, we include either level indicators or trends for years outside the specified pre and post windows depending on whether we are examining levels or trends in participation.

In most cases we treat the regression equation as a linear probability model and estimate it using OLS. However, in specification (4), we verify the results from the third specification using a logit regression. In either scenario, we allow the error term to be clustered at the subject level to account for serial correlation in trial participation. Finally, observations are weighted so that each individual has equal weight in the analysis regardless of how many years they appear in the data.

We conduct our participation rate regressions on two sample: one including all patients with HIV and one including only subjects on primary drugs. The change in the participation rate among HIV positive subjects captures both effects due to changes in the composition of individuals using a drug and changes at the margin of consuming that drug in a trial due to HAART. The change in the rate among primary drug users only captures the latter margin. The HIV sample includes subjects in the denominator who are never eligible for trials because they would not consume drugs even with HAART. Changes in the drug user sample conflate changes at the trial participation margin among a fixed group of subjects and changes in the composition of subjects. Neither sample is perfect, but both are informative.

2. Effect of HAART on equilibrium participation

Our initial results concerning the equilibrium quantity of subjects in primary drug trials are presented in Panel A of Table 3. (We will discuss Panels B to D, which examine specifically placebo- or active- controlled trials, later.) We find that the number of subjects in trials fell by roughly 90 from a base of 128, i.e., a drop of roughly three-quarters. We obtain a slightly larger estimate when drop 1996,

---

Likewise, even though the first protease inhibitor was approved in December 1995, it took a number of years to gather decisive evidence of the benefit of drug cocktails. For example, the first findings demonstrating HAART's effect on CD4 counts, a biomarker, were published in 1997 (Gulick et al. 1997; Hammer et al. 1997). Evidence of the reduction in death rates, the key endpoint, did not emerge until the end of the decade (Chaisson & Moore 1999). This may explain the sharp reversal in trial participation trends after 1996 in Panel B of Figure 1. In a previous version of this paper, we included regressions in which the pre- and post- windows were interacted with the trend term:

$$p_{it} = \beta_1 PRE\_HAART\_TREND_t + \beta_2 POST\_HAART\_TREND_t + [controls] + e_{it}$$

The difference in coefficients on the two interactions identified changes in trend. Those results, not included here, confirm a significant trend break and reversal in 1996.

our transition year, from the sample. When we added linear drift to the regression, the baseline rises to 143 participants and the HAART-induced reduction falls to 67 participants, i.e., a reduction of roughly one-half. In all specifications, we cannot reject that the errors are not serially correlated, suggesting our OLS estimates are adequate.

Table 4 presents our basic results on the equilibrium rate of participation in primary drug trials. Panel A reports results for regression samples including all HIV patients. Panel B reports results for samples including only subjects taking primary drugs. The bottom panel describes the specification of each regression, though the number at the top of each column is a quick reference to the nature of controls and estimation method employed. The first four columns of the table report short run specifications with one-year treatment and control windows, the second four columns report medium-run specifications with 3 year windows, and the last four columns report long-run specifications. At the bottom of Panels A and B we report treatment effects (usually post-window minus pre-window coefficients) and the p-value from a Wald test that the difference is zero. We use this template for many of the remaining tables.

HAART appears to lower trial participation rates among HIV patients by 8 to 9 percentage points in the short-run and 3 percentage points in the long run. The results are more dramatic if we confine our attention to primary drug users. Participation falls 17% in the short-run and more than 9% in the long-run. In general, logit results are at least as strong as OLS results, suggesting our model of errors is not responsible for our estimates.

3. Robustness of equilibrium participation results

a. Unrelated trends in trial participation

A concern with simple pre-post comparisons is that there may be unobserved trends in trial participation that are unrelated to the effect of interest, the introduction of HAART. For example, insurance companies may have changed their willingness to cover medical treatment for adverse events suffered due to trial participation. If this reduced participation, we may overestimate the reduction in participation after HAART.

One way we address is this to include a linear time trend as a control in our basic regressions. An alternative approach we use is to employ subjects using secondary AIDS drugs as a control group. If there are secular changes in trial participation incentives, they should affect individuals' willingness to participate in both primary AIDS drug trials and secondary AIDS drug trials. If we assume identical secular changes in both drug categories, then we can employ changes in participation rates in secondary drug trials as a control for secular changes in participation rates in primary drug trials. The argument in

favor of this assumption is that the same people enroll in both primary and secondary drug trials. The majority of subjects on primary AIDS drugs also took secondary AIDS drugs at the same time because secondary drugs serve as a safety net in case primary drugs do not work.[28] Moreover, our analysis excludes any secondary drug users that did not have HIV or that did not use primary drugs, depending the sample we employ.

A potential concern with using an individual's participation in secondary drug trials as a control for participation in primary drug trials is that, according to Table 2, the introduction of HAART also led to a reduction in secondary drug use. This is not fatal to our difference-in-difference identification strategy, however, because we are using as a control the *rate* at which users of secondary drugs participate in trials rather than the *level* of secondary drug use overall. While HAART may have reduced the demand for secondary drugs, that would not affect the rate at which secondary drugs were procured in trials conditional on positive demand for such drugs.

A better reason why secondary drug trials may not provide good controls for secular changes in participation rates is that many subjects who obtained secondary drugs in trials also obtained primary drugs in those same trials. In our sample, out of 226 person-years we observe in secondary drug trials, fully 142 person-years are also in primary drug trials. Subtracting secondary drug trial participation rates from primary drug trial participation rates could over-correct for secular trends since participation in the secondary drug trial might have also fallen because subjects did not want to obtain primary drugs in (the same) trials rather than secular trends. Thus our difference-in-difference estimate may underestimate the effect of HAART on participation rates in primary drug trials.

Cognizant of this risk, we implement a difference-in-difference identification strategy with the following variant of regressions (8):

$$p_{ijt} = \beta_1 + \beta_2 POST\_HAART_t + \beta_3 PRIM_{ijt} + \beta_4\left(PRE\_HAART_t \times PRIM_{ijt}\right) + \beta_5\left(POST\_HAART_t \times PRIM_{ijt}\right) + [controls] + e_{it} \quad (9)$$

The sample includes either all HIV patients at time t or individuals who used either a primary drug or a secondary drug at time t. Because individuals may use both types of drugs at the same time, observations are at the person x drug type x year level. The dependent variable is an indicator for whether individual i participated in trial for a particular drug type j. We added the indicator $PRIM_{ijt}$ for

[28] Although a large fraction of subjects were on therapeutic and palliative AIDS drugs at the same time, very few people (50) were participating in concurrent therapeutic drug and palliative drug *trials*. It is likely that these individuals were in trials that offered them both therapeutic and palliative drugs because. In general, researchers tend to exclude study subjects on other drugs since they may affect the outcomes measured inside the trial.

whether the drug type is a primary AIDS drug and an interaction between this indicator and our pre- and post-HAART windows. Aside from these changes, we estimate this regression model just as we did equation (8).

The results from the difference-in-difference specification are presented in Table 5. HAART had a small and insignificant effect in the short- and medium-run among all HIV patients and in the short-run among drug users. In the long-run, however, HAART had a small but significant effect on trial participation in some specifications among HIV patients. The same is true for in the medium- and long-run trial in the sample of drug users. At best there is a 1-2 percentage point reduction in participation due to HAART. It should be noted, however, that our (unreported) regressions that examine the effect of HAART on trends in participation rates find more significant effects in a difference-in-difference specification. In the medium-run (long-run), participation rates among drug users change from annual increases of 1.3 (2) percentage points to annual reductions of 4 (0) percentage points per year after HAART.

Because of the difference-in-difference analysis may over-correct for secular trends, we use regression (8) and identification off linear trends and rather than the difference-in-difference approach in subsequent analyses.

b. Changes in sample composition

A limitation of the results above is that they may be driven in part by the changing composition of subjects in the data over time. Subjects enter and exit our sample after the initial enrollment. A subject may show up for visits in one year but not another. Moreover, subjects may drop out permanently due to death. To the extent that death or drop out selects subjects in a manner that is correlated with trial participation decisions, this will bias our estimated treatment effect.

We address this concern in three ways. First, we check temporary and permanent attrition rates to rule out peculiar jumps around 1996. Figure 5 plots temporary dropout and permanent dropout from our sample over time. Temporary drop out includes individuals who do not have data in our sample for a specified year but do have data in the sample for a subsequent year. Permanent dropout includes individuals for the year they are last seen in the sample data and for every subsequent year. The flattening of trends in permanent dropout after 1996 likely represents the drop in death rates following the introduction of HAART. Importantly, there are no jumps in 1996 which would suggest that the spike observed in Figure 3 and documented in the first four columns of Table 4 is driven by attrition.

Second, because AIDS is an important source of attrition, we conduct our analysis controlling for CD4 counts and omitting individuals with AIDS. Figure 6 plots temporary and permanent attrition by CD4

count. It is clear that permanent attrition is driven by individuals with CD4 counts below 200, i.e., individuals with AIDS as opposed to merely HIV. Two of our regression specifications already include nuanced controls for CD4 counts (indicators for binned CD4 counts); these should reduce selection bias due to attrition. To be sure, we re-run the analysis reported in Table 4 but on a subsample of subjects whose CD4 counts never dip below 200 in our sample frame. The first three columns of Table 4 report the results of specification (3) with short-, medium- and long-run windows. We find that the effects of HAART among non-AIDS patients are significant in the short-run among HIV patients and in all durations among primary drug users, though the treatment effects are smaller.

Third, to address changes permanent dropout trends after HAART, we also re-run the analysis reported in Table 4 but on a subsample of subjects that we confirmed were alive as of 2000 and a subset of subjects confirmed alive as of 2005.[29] For the alive-in-2000 subsample, we can estimate a treatment-on-treated effect without survival bias in the short- and medium-run. We can do the same even in the long-run for the alive-in-2005 sample. The results are presented in the last six columns of Table 6. We find the effects of HAART is significant in nearly all durations and in both the HIV and primary drug user samples, though the magnitude of the treatment effects is somewhat smaller. This implies that, even without survival bias, HAART has a significant effect on equilibrium participation rates.

Aside from bias, the main implication of drop out is that we lose observations that can be used for identification of individual fixed effects in our specifications. Given that we find significant effects of HAART on participation, despite the small sample size, we are not concerned about loss of power due to drop out.

F. Explaining the observed changes in the participation rate

The previous subsections identified the reduced-form relationship between HAART and trial participation to illustrate a robust relationship between innovation and equilibrium trial participation. We now examine what the change in equilibrium participation reveals about changes in the supply of research subjects, the mechanism behind the subject supply effect, and try to decompose the effect of HAART into a positive quality effect, which should increase supply, and a positive price effect, which should decrease supply. Finally, we examine the effect of HAART on the design of trials, an alternative way for researchers to compete for subjects when ethical rules cap monetary wages.

1. Supply versus demand response

[29] We cannot address temporary drop out with this test. Multiple imputation methods cannot solve this temporary drop out problem because we would be imputing the dependent variable (participation) with other dependent or independent variables, which would cause us to underestimate our standard errors. Thus, an assumption required for identification is that the probability of observing the dependent variable is independent of HAART.

Recall from equation (3) that the equilibrium participation satisfies $Mf = Ns$ or, in rates, $Mf/N = s$, where M measures demand, f reflects pro-rata allocation of subjects across trials, N is the number of eligible subjects (either HIV-positive individuals or primary drug users), and s is the supply function for individuals. For a number of reasons, we think that the relationship between HAART and the equilibrium trial participation rate is driven by changes in the supply of subjects, s, and not merely changes in demand for subjects, i.e., the number of trials M.[30]

First, ethical rules cap monetary wages in the market for human research subjects. As a result, there is likely to be excess demand for subjects in each period. Because quantity is the minimum of supply and demand when prices are capped, quantity will be set by supply. In equilibrium, the supply of subjects is rationed across trials by queuing, which is reflected in f in our model. Therefore, our estimates of the effect of HAART on the per-period trial participation rate identify the supply behavior of subjects. Even if the introduction of HAART reduced demand by decreasing trials, M, we predict it would not cause the observed decline in trial participation per period.[31] In other words, a sufficient assumption to identify the effect of HAART on supply is that the equilibrium wage is determined by a binding cap.

Second, we investigate whether HAART affected exits from existing clinical trials, a response that is likely to be driven only by the supply rate. In theory, patients may have exited trials after HAART because they no longer wanted to participate – reduced supply – or because drug companies cancelled trials – due to reduced demand for innovation.[32] In standard labor markets "quits" versus "fires" are hard to distinguish empirically, but this is less of an issue for clinical trial recruitment. In practice, we can rule out "fires" because it is standard ethical practice not to cancel ongoing trials unless the treatment drug *in that trial* clearly works or the experimental drug has a severe side effect (Cannistra 2004).[33] Ethics rules do not permit cancellation because some other drug outside the trial, in our case HAART, works better. Empirical support for this view comes from Figure 7 which plots data on trial terminations from the AIDS Clinical Trials Group (ACTG), the main organizing body for HIV/AIDS trials during the period. According to Figure 7, while the number of trials terminated in 1996 were higher than average, it was less than the number terminated in 1994, prior to HAART. Moreover, average annual trial

[30] We can rule out that the decline in participation is due to an increase in the number of eligible patients N because Panel A of Figure 1 shows that the number of patients with HIV or on primary drugs was roughly the same after 1996 as in 1995.

[31] It might explain the decline if demand fell below supply, since equilibrium quantity is the minimum of supply and demand. However, that would imply a reduction in the explicit wage for subjects – or even payment by subjects to trial. To our knowledge, no declines in wage or payments by subjects have been observed.

[32] It should be noted that a drug company may cancel a trial because of subjects dropping out. Thus, drug company terminations may also capture some supply side factors.

[33] This policy may, of course, also be a result of the excess demand for subjects, making "firing" of them unlikely.

terminations in the years after 1996 did not exceed those prior to 1996. If reduced demand for innovation after HAART was not causing trial terminations, then the exit rate from trials is a reasonable measure of the supply-side of the trial participant market.

Raw data on exit rates after HAART can be seen in Figure 4: there was a sharp increase in the exit rate in 1996, after which exits declined gradually, just as they had been prior to HAART. The regression analyses reported in Table 7 confirm this finding. Observations are at the year level and the dependent variable is the number of exits (first two columns) or entrances (last two columns) in a year. The results suggest that exits increased by 80% (36 subjects) in 1996 and then actually fell by 45% (19 subjects) after 1996 relative to the pre-HAART period. While the decline in exits vanishes once we insert a linear drift term, it is clear there is an isolated spike in exits in 1996, which is only consistent with a supply effect. By contrast, there is no significant change in entrances in 1996. After 1996, entrances fall by 45% (11 subjects) or 20% (6 subjects) net of trend. This last effect is consistent with either a supply or demand effect.

Recall that 1997 not only saw the introduction of the HAART drug regimen but also of real-time viral load monitoring as the standard of care. We can use data on exits and entrances to distinguish between the effects of these two changes. First, whereas HAART would induce exits from ongoing trials, the introduction of viral load monitoring would not affect existing trials because the FDA and good trial practice does not allow researchers to change trial protocols midstream; such changes make treatment effects before and after the protocol change less comparable. Thus, the 1996 spike in exits, which only affected ongoing trials, can only be the effect of the introduction of HAART that year.[34] Second, the introduction of viral load monitoring in trials should increase both entrances as well as exits from new trials after 1996. With real time viral monitoring, enrolled subjects can exit as soon as they judge their group assignment – whether treatment or control – is not reducing their viral load as desired. Moreover, a more effective option to exit should make subjects more willing to enroll in trials. Yet we find a reduction in exits in the long run, unless we include a trend in the exit regression. Moreover, we find a reduction rather than an increase in entrances after 1996. These two finding lead us to conclude that viral load monitoring cannot explain the changes we see in trial participation 1996 onwards.

Finally, we present suggestive evidence that the amount of money spent on AIDS-related R&D, a proxy for demand for subjects, did not decline after HAART. Figure 8 plots NIH spending on R&D for HIV/AIDS between 1995 and 2010. It shows that federal spending continued to rise dramatically even

[34] Moreover, whereas HAART was partly introduced in December 1995 and the benefits partly reported in January 1996, information on the benefits of real time viral load monitoring was not published until 1997.

after the introduction of HAART. Indeed, the data suggests that federal spending nearly doubled between 1996 and 2005, the last year of our regression sample. Although we do not observe private R&D spending on HIV/AIDS, we know that government agencies were the sponsor of more than half of all HIV/AIDS trials in the 1990s. Of the 1,121 ACTG trials, 581 were sponsored by NIH. Moreover, we have indirect data on private R&D that suggests that R&D did not plummet after HAART. Specifically, the number of AIDS drugs that private drug companies were testing in clinical trials increased steadily from 25 in 1987 to 125 in 1997 (Neumann & Sandberg 1998). By 2001, the number had fallen to 98 (PhRMA 2002), and stayed constant at 100 until the 2010 (PhRMA 2010).

2. Quality versus price response

Even if HAART largely operated through supply effects, to what extent did HAART operate through the quality versus the price of conventional treatment? Prior literature and the graphical evidence in Figure 2 suggest HAART clearly improved the quality of conventional treatment (as measured by health outcomes). But HAART also increased the price of such treatment for uninsured patients. Whereas AZT (conventional care prior to HAART) cost up to $8000/year in 1989 (NYT Aug. 28, 1989), a HAART regimen cost $12,000-$15,000 per year in 2000 (Steinbrook 2001). Our theoretical analysis implies that the quality improvement would reduce the supply of trial participants, while a price increase would increase supply. Thus our estimates of the effect of HAART on supply may underestimate the effect of improvement in quality on supply because of the offsetting effect of the price increase.

We tackle this issue by comparing subjects who had insurance to those who did not. Individuals without insurance experienced a change in both the quality and price of conventional treatment after HAART. Individual with insurance, however, faced low co-pays both before and after HAART. Thus, they saw HAART mainly as a change in quality of conventional treatment. Table 8 provides evidence consistent with this discrepancy. It shows average out of pocket (OOP) expenditure on drugs among subjects in our sample. Individuals without insurance who enrolled in clinical trials did experience a sharp reduction in OOP drug spending. For individuals with insurance, however, enrollment in clinical trials afforded no reduction in OOP spending.[35]

The specific regression model we employ to separate price and quality effects is

[35] Individuals with insurance, whether private or public, do not have significantly lower OOP drug expenditure. This is not necessarily because they did not have good coverage for drugs; it is likely due to adverse selection into insurance. For the same reason, we do not observe that individuals without insurance have higher rates of trial participation than individuals with insurance.

$$p_{ijt} = \beta_1 + \beta_2 POST\_HAART_t + \beta_3 UNINS_{ijt} + \beta_4 \left(PRE\_HAART_t \times UNINS_{ijt}\right) \quad (10)$$
$$+ \beta_5 \left(POST\_HAART_t \times UNINS_{ijt}\right) + [controls] + e_{it}$$

$UNINS_{ijt}$ is an indicator for not having insurance. The coefficient on the uninteracted post-HAART indicator captures the common effect across the insured and uninsured groups and thus the effect of a change in quality after HAART. The coefficient on the post-HAART indicator interacted with uninsured status captures the differential impact on the uninsured and thus the effect of a change in price after HAART.

From Table 8 we see that Medicare or Medicaid did a better job of insulating members from OOP drug spending than private insurance: trial participation yielded a smaller reduction in OOP spending for individuals without Medicare or Medicaid than without private insurance. Therefore, Table 9 presents results where lack of insurance is an indicator for lack of Medicare or Medicaid insurance in particular. (We obtain qualitatively similar results when we use an indicator for lack of private insurance.) We find that that HAART reduced the trial participation rate among insured populations up to 7 percentage points among HIV patients and 10 percentage points among primary drug users. Relative to this group, uninsured populations experienced an increase in participation rate of 7 percentage points among HIV patients and 8 percentage points among primary drug users. These results are significant in the short-run, but not always so in the medium- and long-runs. Yet they imply that the quality improvement due to HAART significantly reduced the supply of participants, while the price effect offset much of this. In the aggregate, because most individuals have insurance, we observe the quality effect dominating.

III. Positive implications of the subject supply effect

Due to the subject supply effect, changes in the output market have non-standard impacts on the incentive to innovate. Specifically, changes in the output market affect the time it takes to complete R&D, which in turn alters the present value of profits from R&D. Here we elaborate on these effects, and discuss equilibrium development time.

A. Innovative returns

We assume the innovation has a finite period, T^P, during which it is under patent. After this period, the market becomes competitive and profits fall to zero. Let $A(x)$ be the value of a standard annuity paying one dollar for x periods. The net present value of the overall innovative return can be written

$$NPV(q,p) = -F + \beta^{T(p,q)} A(T^P - T(p,q))\Pi(q,p) \quad (11)$$

Here, F is the fixed cost of development, $\Pi(q,p)$ is the reduced form of per-period expected variable profits after development, given the distribution of experimental outcomes, the probability of market approval, and the quality and price of the conventional care. The quality and price of conventional care are assumed to be exogenously given, as would be the case in a competitive market for the conventional treatment.

Delays in development have two negative effects on the present value of innovative returns. First, they delay when the firm begins earning positive variable profits. This effect operate through the term β^T. Second, delays in development eat up patent life, i.e., the period during which firms have exclusive rights to sell their new product, reducing the number of periods the firm earns positive profits. This effect operates through the term $A(T^P - T)$. Prior research suggests that the cost of development delays can be economically significant. For example, DiMasi, Hansen & Grabowski (2003) find that that financial opportunity costs attributable to the duration of clinical trials account for half the total cost of clinical research.[36] Even that is likely to be an underestimate because it does not address the effect of trial duration on remaining patent life.

We can now distinguish our argument from previous arguments for how output markets affect innovation. To simplify notation, let $V(p,q) = \beta^T A(T^P - T)$ denote the value of a dollar of annual profits during a product's patent life. For the two reasons just given, this value falls with development time: $V_T < 0$. Changes in output markets have the following effects on innovative returns:

$$\frac{dNPV}{dx} = V\Pi_x + V_T T_x \Pi, x \in \{q,p\} \quad (12)$$

The first term captures the standard argument for how changes in output markets affect innovation: they change expected, per-period variable profits. The second term captures our insight: through subject supply effects, reforms alter development times and thus the present value of profits.

For example, an improvement in the quality of conventional care reduces the nominal per-period patent profits a firm can expect from a future innovation. The same improvement in the quality of conventional care also reduces the propensity of patients to enroll in trials. The resulting delay reduces the present value of per-period patent profits. If we instead consider changes in the price of conventional care, standard arguments suggest the first term is positive: the profits from a new product

[36] Philipson and Sun (2010) offer estimates of the magnitudes of delay costs for a number of specific classes of drugs.

increase in the price of conventional care. Subject supply effects suggest the second term is also positive because of faster development. Thus, a higher price for conventional care would increase the return to innovation.[37]

When extrapolating from this theory to predict the effect of health care reforms on innovation, it is important to be careful about the definition of price. In some cases it will be straightforward. Reforms that reduce the reimbursements that health care providers receive clearly reduce the price of conventional care, which will reduce innovative returns. Other cases may not be so simple. For example, reforms that expand insurance coverage increase the price at which innovative firms can expect to sell their products. However, it also lowers the price at which patients are able to purchase conventional care, so-call out-of-pocket payments. This reduces their willingness to enroll in clinical trials. Thus, the firm price effect controls the sign of the first term in (**Error! Reference source not found.**), while the patient price effect controls the signs of the second term in (**Error! Reference source not found.**). The net effect is ambiguous: insurance expansions may increase or decrease innovation, depending on which effect is larger. More generally, standard effects are driven by the price of conventional treatment faced by health care producers, while the non-standard effects we highlight – the subject supply effects – are driven by the price faced by consumers.

B. Free entry and the equilibrium duration of development with a wage cap

As a result of the ethical cap on wages of research subjects, even when the number of new ideas is high, the pace of innovation may be slowed down because subjects cannot be compensated to participate in the development process. Wage caps limit the supply of subjects. An increase in ideas, captured in our model by the number of trials, increases the demand for these subjects. Instead of being rationed by price, they are rationed by queuing. As a result, an increase in ideas may, paradoxically, delay innovation. In this section we examine how the price and quality of conventional care affect the development time under free entry of trials and rationing by delay.

Rationing of subjects by queuing suggests that, holding the supply of subjects constant, an increase in trials increases development times: $T_M > 0$. In the preceding section, we saw that an

[37] The preceding discussion on the impact of conventional care on innovative returns is very general and extends well beyond the specific analysis or model discussed here. Indeed, if higher quality and lower price of conventional care drive up development times, the innovative return must fall. This is because, regardless of the economic environment, the innovator is always free to choose a longer, but not shorter, time of development. Therefore, shortening development times always makes the innovator weakly better off since he can always still choose the original longer time of development.

increase in development time reduces the present value of profits: $NPV_T < 0$. Together these facts imply that the present value of innovation falls in the number of trials.

Under free entry, the equilibrium number of trials dissipates the profits from entering,

$$NPV(T(M, q, p), q, p) = 0$$

The equilibrium number of trials will be affected by output market conditions, i.e., $M(q, p)$. While development time, $T(M, q, p)$, increases in the number of trials, it also increases in the quality of conventional care and falls in its price.[38]

Applying the implicit function theorem, we find that the equilibrium number of trials respond to output markets as follows:

$$\frac{dM}{dq} = \frac{1}{T_M}\left[\frac{NPV_q}{-NPV_T} - T_q\right] \leq 0 \quad \frac{dM}{dp} = \frac{1}{T_M}\left[\frac{NPV_p}{-NPV_T} - T_p\right] \geq 0$$

There are two reinforcing effects of higher quality of conventional care on the number of trials. The first is that development time rises ($T_q > 0$,), so that fewer trials are needed before profits are dissipated. The second is that the future profits are lowered ($NPV_q < 0$,), also suggesting fewer trials in equilibrium. Analogous arguments imply that increasing the price of conventional care raises the number of trials in equilibrium.

Given this relationship between output markets and the number of trials, we can derive the connection between output markets on development times in equilibrium:

$$\frac{dT}{dq} = T_q + T_M\left(\frac{dM}{dq}\right) \quad \frac{dT}{dp} = T_p + T_N\left(\frac{dN}{dp}\right)$$

Higher quality conventional care has the direct effect of increasing development times as previously discussed, but also an offsetting indirect negative effect of reducing the number of trials. A higher price has a negative direct effect but a positive indirect effect by increasing the number of trials.

Note that, when development times ration trials, the flow of subjects recruited *per period* into trials, s, is independent of demand, N, because price controls do not allow wages to rise when demand increases. However, the number of periods or length of development is not independent of demand; it rises with N, i.e., $T_N > 0$, thereby rationing further demand. This is important when empirically

[38] For example, if development has only a fixed cost, F, and no variable costs, then equilibrium development time is derived from the free entry condition $\beta^T A(T^P - T)\Pi(q, p) = F$. Development time is an implicit function $T(\beta, T^P, q, p)$ of exogenous parameters. The associated number of trials is defined by $TNs/M = n$ or $M = TNs/n$.

investigating changes in the participation rates (quantities *per period*), which will be driven by supply. Put another way, under a maximum wage policy, the market quantity observed is the lower amount supplied at the constrained price, not the higher amount demanded at that constrained price.

IV. Normative implications of the subject supply effect

Our analysis of the economics of medical R&D also has important normative implications for medical R&D and policy. These implications are driven by the ethical cap on subject wages. Here we derive the effect of wage caps on the equilibrium level of medical innovation and contrast the privately and socially optimal level of subject compensation.

A. Free entry and inefficiently long development times with wage caps

When it is not possible to compensate subjects, there are negative external effects on development from entry of trials that resembles the standard commons problem. One can describe this as "over-fishing" for subjects by investigators. Equilibrium development time will be inefficiently long in the sense that industry profits will be not maximized.

More precisely, the efficient level of entry maximizes aggregate profits, $M * NPV(T(M))$. The necessary first-order condition has the additional profits from entry just making up the reduced profits of existing entrants

$$NPV + M * NPV_T T_M = 0$$

This directly implies profits are positive, $NPV > 0$. By contrast, in the presence of wage caps, there is free entry until $NPV = 0$. The analysis in the previous section implies development takes inefficiently long relative to what maximizes industry profits. The industry can earn positive profits by restricting entry into development. This will speed up development by reducing the competition for subjects.

B. Privately and socially optimal wages

Here we discuss socially optimal subject compensation in the absence of ethical wage caps and show that even unregulated private wages may be suboptimal. A research subject generates information about the experimental drug by participating in a trial, information that is valued by future patients. Those future patients would like to compensate current research subjects for the positive external effect of generating that information, but there are two obstacles they face. First, information is a public good. Second, transactions costs, including the fact that future patients may not even be sick yet, prevent bargains directly between those patients and current subjects. However, innovative firms can serve as middle men internalizing this externality. By financing subject compensation from future patient sales they allow compensating subjects for the benefit they provide to those future patients.

Specifically, a representative firm that is free to compensate subjects chooses a pecuniary wage to maximize

$$NPV(w) = -nw + D\big(T(w)\big)\Pi \tag{13}$$

where $T(w)$ is how duration responds to higher wages and $D\big(T(w)\big) = \beta^{T(p,q)}A(T^P - T(p,q))$ is the present value of a dollar received throughout the patent window.[39] The necessary first order condition equates the marginal increase in present value of expected profits as wages speed development with the marginal cost of a higher wage bill:

$$D_T T_w \Pi = n$$

When D_T and T_w are both decreasing firms with more profitable drugs, through a larger demand of future patients benefitting from learning, should offer higher trial wages and get to market faster.

Of course private profits do not capture the entirety of social welfare. The shortfall has two components.[40]

$$S(w) = -nw + D\big(T(w)\big)[\Pi + CS_{patent}] + \beta^{T(p)}A(\infty)[CS_{generic}]$$

Even when medical products are under patent, consumers earn per period surplus CS_{patent}, though less than first best from a static perspective. Moreover, consumers earn larger surplus $CS_{generic}$ when products go off patent. Even without wage caps, a firm's compensation to research subjects will not capture the full value of consumer surplus under a patent. If development times are so long that the net present value of product development (**Error! Reference source not found.**) at the privately optimal wage becomes negative, the firm will not develop the product. Yet if the firm internalized the loss of surplus when the product went off patent it might have reversed course.

The failure of firms to fully internalize the benefit to future consumers of clinical trials suggests the need for Pigouvian subsidies to encourage research subjects to enroll in clinical trials. Therefore, the ethical wage cap has a double cost. It not only prevents the firm from efficiently capturing future

[39] Recall that a positive wage w for trial subjects implies a negative price p^E for experimental treatment in equation (2), i.e., a positive wage implies $p^E = -w$.

[40] We acknowledge, however, that the wage cap may satisfy some ethical criteria that may matter to welfare. For example, it may serve some deontological aims such as preventing patients from being coerced by payment into dehumanizing action; preventing patients from making medical decisions motivated by "impure" aims, i.e., something other than their own health or altruism; or it may challenge the professional ethos that elevates care of patients in medicine (e.g., Reiser 2005). Alternatively, the constraint may satisfy a consequentialist aim, such as preventing irrational individuals from making decisions that do not promote their own utility (e.g., Fry et al. 2006). Some bioethicists argue, however, that more robust informed consent and counseling can address both the deontological and consequentialist concerns with subject compensation (e.g., Grady 2005).

producer surplus. It also prevents it from capturing future consumer surplus. The result is yet longer development times in a Pareto sense.[41]

To some extent excessive development times can be addressed by changes in non-pecuniary wage, i.e., modifications to the design of clinical trials that provide more benefit the subject. However, changes in the design of trials is a less efficient way of rationing subjects than wages. Whereas trials for higher value medical innovations are able to finance higher wages, they are no more capable of modifying design than trials for lower value medical innovations. Moreover, changes in design may also reduce the value of the public information on treatment effects generated by trials, information that is more valuable for trials for more valuable innovations. Finally, if regulators set sample size to ensure adequate quality information is generated by trials, e.g., 95% confidence and 80% power, then regulators may offset the recruitment benefits of design changes by raising sample size requirements.

It might seem intuitive to address cap-induced delays in development by extending patent-life. This would not be a Pareto-efficient solution. First, additional patent length generates additional static deadweight loss without speeding up recruitment; hurting future patients in two ways. Second, wage caps would continue to prevent the firm from using the additional producer surplus to compensate subjects. A better use of the extended patent term might be to enable the firm to internalize more of the social surplus from innovation in a regime without a wage cap. The question of whether direct Pigouvian subsidies or extended patent term is a superior tool for this problem depends on the same considerations that determine whether it is better to encourage innovation with patents or rewards: whether the government or the firm has a better sense of the social surplus from their innovation.

Our analysis of wage caps also reveals a second flaw in the existing methodology for estimating the cost of medical R&D (e.g., Dimasi et al. 2003). According to that methodology, if a firm spends more on subject compensation, it would raise medical R&D costs. Our analysis suggests a more optimistic conclusion. Greater spending on subject compensation also reduces development time and thus opportunity costs. Moreover, the overall returns from innovation may be positive because the present value of profits rise. From a social perspective, spending more on subject compensation – relative to

[41] To be clear, we advocate for subject compensation with informed consent, not the elimination of informed consent. Without informed consent, it is possible that trades result in an inefficient transfer from subjects to future patients because some involuntary subjects may have opportunity costs higher than the value of the information they produce for future patients. Our point is analogous to the argument for a volunteer army rather than a compulsory draft; the public wants the public good a military provides, but a volunteer army ensures that good is provided at lower social cost.

virtually zero compensation today – efficiently raises innovative returns by internalizing the key externality involved in medical development.

V. Concluding remarks and implications for future research

In this paper we argued that changes in the price and quality of conventional care have non-standard effects on development mediated by the market for research subjects, and consequently on the health benefits from and future cost growth induced by medical innovation. We illustrated such subject supply effects by exploring the introduction of Highly Active Antiretroviral Treatment (HAART) in 1996, a dramatic increase in the quality of conventional treatment for HIV/AIDs. We documented a substantial reduction in the supply of subjects due to this new innovation. In our conclusion, we explore a number of positive and normative issues that would be productive avenues for future research.

A. Location of medical R&D

Subject supply effects have implications for the spatial location of medical R&D. In most other product markets, R&D performed in one country can be used to create a product sold in any other country. Therefore, it is efficient for the location of R&D spending to be determined independently of output market conditions in different regions. Under this view, a Swedish firm does not innovate for its own 9 million people, but for the world market. As a result, the output market in Sweden should not affect the nature of the R&D that firms perform in Sweden.

Medical R&D does not appear to conform to this wisdom. The locus of medical R&D has shifted dramatically in the last two decades from Europe towards the US. In 1990, Europe conducting 15% more R&D than the U.S., but by 2010 it only conducted 72% of U.S. R&D (European Federation of Pharmaceutical Industries and Associations 2011). A common explanation in medical policy circles for this change is that European price controls are responsible for the shift towards the US (Gambardella et al. 2000; U.S. Department of Commerce 2004). This argument is dismissed by economists who use the logic in the preceding paragraph to argue that R&D does not track output markets. However, our analysis of subject markets implies that the naive policy argument may have more credit than it is given.

The competition among trials for research subjects suggests that the amount of research across locations depends on the relative disease prevalence, subject participation rates, and the number of competing trials in each location. This implies that there may be a connection between output markets and development across regions, contrary to standard arguments. Here are three examples. First, trials of treatments for obesity may be easier to conduct in the U.S. where those treatments will disproportionately be sold. The high prevalence of obesity there speeds up recruitment through two

channels: a direct effect from grater prevalence and an indirect effect from high prices of conventional care. Second, price controls in Europe may be responsible for the decline of clinical trials there. Price controls not only lower future profitability in local output markets, they also dissuade patients from participating in local trials. Third, developing countries with low quality conventional care or high relative prices for that care are good candidates for clinical trials. These conditions speed up trial recruitment. This may explain the recent growth in clinical research in those countries (Thiers, Sinskey & Berndt 2008). It may explain why few industries other than medical products spend so much on development in the continents of Africa and South America.

B. Optimal sample size determination

Current regulation of clinical research for drug approval dictates sample size based on statistical power criteria (see, e.g., Food & Drug Administration 1998). Prior economic analysis has expanded the analysis of sample size to balance the immediate financial cost of enrolling more subjects with the benefit of avoiding incorrect treatment decisions (see, e.g., Claxton & Posnett 1996).

Pareto optimal experimental design, however, is inherently an inter-temporal consumption problem of delaying treatment for most consumers until tomorrow in order to learn about product quality amongst a small number of consumers today. Subject supply effects will affect this balance because recruitment of larger samples, n, delays development. Holding other considerations constant, this suggests output market considerations will tend to lower optimal sample sizes. Counterintuitively, it also suggests that optimal sample sizes may depend on both the quality and the price of conventional treatment.

C. First mover advantage in medical R&D

Subject supply effects have direct implications for the differential R&D costs associated with first-in-class entry versus subsequent or so-called me-too innovation. Our analysis suggests that first-in-class drugs have a first mover advantage for two reasons. First, firms have lower R&D costs when they alone are recruiting subjects in a disease class. Second, first-in-class innovations, by raising quality of conventional care in the output market, slow down development of subsequent development. In other words, medical R&D may have a self-limiting effect, with earlier innovations raising the cost of subsequent innovations. This first-mover advantage may have important market structure implications (see Tirole 1988). It also has important implications for the optimal speed of creative destruction for medical products.

D. Innovations that affect (known) disease prevalence

Subject supply effects may differ for curative treatments and new diagnostic tests. Development time falls with the steady-state prevalence of a disease, which, in turn, increases with its incidence, b, and decreases with mortality or the exit rate, m. For some diseases the exit rate may be affected by the price and quality of conventional care, $m(q,p)$. The cheaper or more effective that conventional care is the more patients utilize it. Therefore, its quality and price not only affects participation rates but also the stock of subjects available to participate.[42] For example, if a newly approved conventional treatment can cure a disease, improving its quality or reducing its price will lower the prevalence of that disease, reducing trial participation and increasing development times, compounding the usual subject market effect. Conversely, if a newly approved conventional treatment lowers mortality of a disease without curing it, improving it will increase the prevalence of that disease, offsetting the standard subject market effect of from new treatments.[43]

There may also be indirect effects of price and quality of conventional care that operate through the incidence (flow into the disease stock), rather than through cure or mortality (flow out of the disease stock).[44] This is particularly relevant for understanding how diagnostic technologies affect development times. While improved diagnostic capabilities do not increase the actual stock of disease, they may improve recruitment and shorten development times because only diagnosed patients can enter trials.[45]

E. Comorbidities and Cross-Diseases Effects

Future analysis should also consider the interaction between R&D activities for different diseases. Our analysis focused on a single disease for which experimental and conventional care were gross substitutes in the sense that participation in trials rose with output prices. When there are multiple diseases that have separate trials recruiting subjects, the price and quality of conventional care for a given disease not only affect development times and innovative returns for that disease, but also for other diseases through multiple channels.

---

[42] Recovery or mortality may have also have a direct effect on development times through the participation rate $s(p,q)$. For example, lower mortality will increase the number of future periods the patient will enjoy positive utility: $U(q,p) = (1/m(q,p))u(q,p)$, where $1/m$ is life expectancy under conventional care and u is annual utility under such care. Thus, if conventional care lowers mortality, it will reduce willingness to participate in trials.

[43] In particular, if quality affects exits out of the stock of the disease by $m_E(p_E, q_E)$ for those in the experiment and $m_C(p,q)$ for those on conventional care, then the overall exit of the stock of disease is given by $m(p,q) = s(p,q)E[m_E(p_E,q_E)] + [1-s(p,q)]m_C(p,q)$.

[44] For example, Lakdawalla, Goldman, Sood (2006) find that the introduction of a new HIV/AIDS treatment in 1996 was associated with an increase in sexual risk taking, increasing the incidence of HIV.

[45] By contrast, Williams (2007) stressed the "pull" effect of diagnostics. By identifying new patients for treatment, diagnostics increase the returns to treatment innovations.

For example, conventional care for one disease may affect the prevalence of a second disease. For example, if conventional care for treating heart disease improves, there will be a larger population facing co-morbid risks, such as Alzheimer's disease. This in turn lowers development times for Alzheimer's research. As a result, while innovation may be self-limiting within a disease, it may increase innovative returns for competing diseases (e.g., Geoffard and Philipson 1999).[46] This has important implications for the debate over whether and why there has been a slowdown in aggregate productivity of medical research (e.g., Cockburn 2007). The productivity of medical R&D is negatively affected by the self-limiting effects of R&D within a disease, but positively affected by cross-disease effects.

More generally, we have endeavored to provide support for the claim that there is a unique relationship between output markets and development in health care. This relationship suggests that standard positive and normative analysis of R&D may not necessarily apply to medical product markets. Thus reforms to the quality and price of conventional care may have non-standard implications for the cost of R&D and thus the future spending growth fueled by such R&D. Given the enormous economic benefits of previous medical advances and the potentially large gains that medical progress may deliver in the future (Murphy & Topel 2006), a better understanding of the unique dynamics of medical R&D seems warranted.

[46] We do not assert that cross-disease effects are welfare reducing. Recruitment for trials in each disease class may be inefficient given ethical wage caps. However, shifting research from one disease class to another need not increase the inefficiency from those caps.

REFERENCES

1. Acemoglu, D.; D. Cutler; A. Finkelstein and J. Linn. 2006. "Did Medicare Induce Pharmaceutical Innovation?" American Economic Review, 96(2), 103-07.
2. Acemoglu, D. and J. Linn. 2004. “Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry.” The Quarterly Journal of Economics, 119(3): 1049-1090.
3. Adams, C. P. and V. V. Brantner. 2006. "Estimating the Cost of New Drug Development: Is It Really 802 Million Dollars?" Health Aff (Millwood), 25(2), 420-8.
4. Bentley, J. P. and P. G. Thacker. 2004. "The Influence of Risk and Monetary Payment on the Research Participation Decision Making Process." Journal of Medical Ethics, 30(3), 293-98.
5. Blume-Kohout, Margaret and Neeraj Sood. 2012. “Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development.” Journal of Public Economics, forth coming. Available at doi: 10.1016/j.jpubeco.2012.10.003.
6. Brichet, Stephanie and Nicole Cohen. 2007. “The Pursuit of High Performance through Research and Development Understanding Pharmaceutical Research and Development Cost Drivers”. Accenture White Paper.
7. Charles J. Carpenter, M.A. Fischl, S.M. Hammer, M.S. Hirsch, D.M. Jacobsen, D.A. Katzenstein, J.S.G. Montaner, D.D. Richman, M.S. Saag, R.T. Schooley, M.A. Thompson, S. Vella, P.G. Yeni, and P.A. Volberding. 1997. “Antiretroviral therapy for HIV infection in 1997: updated recommendations of the international AIDS society—USA panel.” Journal of the American Medical Association, 277, 1962–1969.
8. Chan, Tat Y. and Hamilton, Barton H. 2006. “Learning, Private Information, and the Economic Evaluation of Randomized Experiments.” Journal of Political Economy, 114(6), 997-1040.
9. Chandra, Amitabh and Douglas Staiger. 2007. Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. Journal of Political Economy. 115: 103-140.
10. Chassang, Sylvain, Gerard Padró i Miquel and Erik Snowberg. 2012. “Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments.” American Economic Review. 102: 1279-1309.
11. Claxton, K. and J. Posnett. 1996. "An Economic Approach to Clinical Trial Design and Research Priority-Setting." Health economics, 5(6), 513-24.
12. Clemens, Jeffrey. 2012. “The Effect of U.S. Health Insurance Expansions on Medical Innovation.” Working paper.

13. Cockburn, I. 2007. “Is the Pharmaceutical Industry in a Productivity Crisis?” Chapter in A. Jaffe, J. Lerner, and S. Stern (eds.) Innovation Policy and the Economy, Volume 7. MIT Press for the National Bureau of Economic Research, Cambridge MA. pp. 1-32.
14. Congressional Budget Office. 2006. Research and Devlopment in the Pharmaceutical Industry. Pub. No. 2589.
15. Cutler, David. 2004. “Your Money or Your Life: Strong Medicine for America’s Health Care System.” New York, NY: Oxford University Press.
16. Cutler, David M. and Srikanth Kadiyala. 2003. The Return to Biomedical Research: Treatment and Behavioral Effects. 110-162. In Kevin Murphy and Robert Topel, eds., Measuring the Gains from Medical Research. U. Chicago Press, Chicago, IL.
17. Cutler, D. M.; G. Long; E. R. Berndt; J. Royer; A. A. Fournier; A. Sasser and P. Cremieux. 2007. "The Value of Antihypertensive Drugs: A Perspective on Medical Innovation." Health Aff (Millwood), 26(1), 97-110.
18. DiMasi, J. A.; R. W. Hansen and H. G. Grabowski. 2003. "The Price of Innovation: New Estimates of Drug Development Costs." Journal of Health Economics, 22(2), 151-85.
19. Dunn, L. B. and N. E. Gordon. 2005. "Improving Informed Consent and Enhancing Recruitment for Research by Understanding Economic Behavior." Journal of the American Medical Association, 293(5), 609-12.
20. European Federation of Pharmaceutical Industries and Associations. 2011. “The Pharmaceutical Industry in Figures.” Available at: http://www.efpia.org/content/default.asp?PageID=608.
21. Finkelstein, A. 2004. “Static and dynamic effects of health policy: Evidence from the vaccine industry.” The Quarterly Journal of Economics, 119(2), 527–564.
22. Food and Drug Agency, International Conference on Harmonisation; Guidance on Statistical Principles for Clinical Trials, Availability, 63(179) Fed. Reg. 49583, 49593 (Sept. 16, 1998).
23. Gambardella, Alfonso; Luigi Orsenigo and Fabio Pammolli. 2000. “Global Competitiveness in Pharmaceuticals: A European Perspective. Report prepared for the Enterprise Directorate-General of the European Commission. Available at: http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/comprep_nov2000_en.pdf.
24. Grady, C.; N. Dickert; T. Jawetz; G. Gensler and E. Emanuel. 2005. "An Analysis of U.S. Practices of Paying Research Participants." Contemp Clin Trials, 26(3), 365-75.
25. Gulick, R. M.; J. W. Mellors; D. Havlir; J. J. Eron; C. Gonzalez; D. McMahon; D. D. Richman; F. T. Valentine; L. Jonas; A. Meibohm, et al. 1997. "Treatment with Indinavir, Zidovudine, and Lamivudine

in Adults with Human Immunodeficiency Virus Infection and Prior Antiretroviral Therapy." N Engl J Med, 337(11), 734-9.

26. Hammer, S. M.; K. E. Squires; M. D. Hughes; J. M. Grimes; L. M. Demeter; J. S. Currier; J. J. Eron, Jr.; J. E. Feinberg; H. H. Balfour, Jr.; L. R. Deyton, et al. 1997. "A Controlled Trial of Two Nucleoside Analogues Plus Indinavir in Persons with Human Immunodeficiency Virus Infection and Cd4 Cell Counts of 200 Per Cubic Millimeter or Less. Aids Clinical Trials Group 320 Study Team." N Engl J Med, 337(11), 725-33.
27. Hutt, Peter Barton; Richard A. Merrill and Lewis A. Grossman. 2007. Food and Drug Law: Cases and Materials. New York, NY: Foundation Press Thomson/West.
28. Lakdawalla, D.; N. Sood and D. Goldman. 2006. "HIV Breakthroughs and Risky Sexual Behavior." Quarterly Journal of Economics, 121(3), 1063-102.
29. Peter Landers. “Cost of Developing a Drug Increases to About $1.7 Billion.” Bain & Company, 2003 study.
30. Lichtenberg, Frank. Pharmaceutical Innovation, Mortality Reduction, and Economics Growth. 74-109. In Kevin Murphy and Robert Topel, eds., Measuring the Gains from Medical Research. Chicago: U. Chicago Press.
31. Lichtenberg, Frank. 2004. “Public Policy and Innovation in the U.S. Pharmaceutical Industry,” in Public Policy and Entrepreneurship Douglas Holtz-Eakin and Harvey Rosen (eds.) MIT Press, Cambridge, MA.
32. Malani, Anup. 2008. Patient enrollment in medical trials: Selection bias in a randomized experiment. Journal of Econometrics. 144: 341-351.
33. Moore, R. D. and R. E. Chaisson. 1999. "Natural History of Hiv Infection in the Era of Combination Antiretroviral Therapy." Aids, 13(14), 1933-42.
34. Murphy, Kevin, and Robert Topel. 2006. The Value of Health and Longevity. Journal of Political Economy, 114(5): 871-904.
35. Neumann, P. J. and E. A. Sandberg. 1998. "Trends in Health Care R&D and Technology Innovation." Health Aff (Millwood), 17(6), 111-9.
36. New York Times. AZT’s Inhuman Cost. August 28, 1989.
37. Pharmaceutical Research and Manufacturers of America. 2002. 2002 Industry Profile. PhRMA, Washington, DC.
38. Pharmaceutical Research and Manufacturers of America. 2010. Pharmaceutical Industry Profile 2010. PhRMA, Washington, DC.

39. Philipson, T. 1997. "The Evaluation of New Health Care Technology: The Labor Economics of Statistics." Journal of Econometrics, 76(1-2), 375-95.
40. Ripley, E.; F. Macrina; M. Markowitz and C. Gennings. 2010. "Who's Doing the Math? Are We Really Compensating Research Participants?" J Empir Res Hum Res Ethics, 5(3), 57-65.
41. Scherer, F. M. 1993. "Pricing, Profits, and Technological Progress in the Pharmaceutical Industry." Journal of Economic Perspectives. 7(3): 97-115.
42. Steinbrook, R. 2001. "Providing Antiretroviral Therapy for Hiv Infection." N Engl J Med, 344(11), 844-6.
43. Thiers, F. A.; A. J. Sinskey and E. R. Berndt. 2008. "Trends in the Globalization of Clinical Trials." Nature Reviews Drug Discovery, 7(1), 13-14.
44. Tirole, J. 1988. The Theory of Industrial Organization, Cambridge, MA: MIT Press.
45. U.S. Centers for Disease Control & Prevention. 1993. "1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults." Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm.
46. U.S Department of Commerce, International Trade Administration. 2004. "Pharmaceutical Price Controls in OECD Countries: Implications for U.S. Consumers, Pricing, Research and Development, and Innovation. Available at: www.ita.doc.gov/drugpricingstudy.
47. U.S General Accounting Office. 2006. New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts. GAO-07-49.
48. Welton, A. J.; M. R. Vickers; J. A. Cooper; T. W. Meade and T. M. Marteau. 1999. "Is Recruitment More Difficult with a Placebo Arm in Rcts? Reply." British Medical Journal, 319(7213), 854-54.
49. Heidi Williams. 2007. "Dynamic effects of medical technologies: Evidence from the prostate cancer market." Working paper.

APPENDIX: Changes in trial design

Price and quality changes can also have indirect effect on development due to their impact on the experimental design of clinical trials. For example, in order to compete with improvements in conventional care, researchers often modify the design of trials to include active controls[47] or an increased share of patients receiving experimental care.[48] It is extremely difficult to recruit for a placebo-controlled trial with a high quality treatment on the market. In essence, alteration of trial design is a form of non-price competition by those conducting experiments in response to improved output market conditions.

Consider two design parameters, represented by the vector $d = (d_1, d_2)$, that affect the value of trial participation. Together these parameters generate a non-pecuniary "wage" for trial participants. One of the design parameters is the probability of being assigned to the treatment group, d_1. A second parameter, d_2, is the quality of the control treatment, say placebo or active control.[49] For these two design features, the expected utility of trial participation is given by

$$d_1 E_F U(q^E, p^E) + (1 - d_1) U(d_2, p^E) \quad (14)$$

Trial participation will rise with the two forms of non-wage compensation.

Changes in output markets may induce changes not only in output markets but also in experimental designs. If $d(q, p)$ denotes the design under a given set of output market conditions, the total impact of a change in quality or price on participation is

$$ds/dx = s_x + s_d d_x, x \in \{q, p\} \quad (15)$$

An improvement in the quality of conventional care, for example, now has both a direct negative impact on participation and an offsetting indirect positive effect from inducing a trial design that is more favorable to subjects.[50]

---

[47] Welton et al. (1999) offer evidence that that using placebo as a control hinders recruitment.

[48] Malani (2008) offers evidence that increasing the probability of receiving experimental care improves recruitment.

[49] Active controls are only of value for those who respond well to conventional care. For non-responders, active controls have no benefit and may raise trial costs.

[50] Given the response of trial designs to output markets, one might reason that improvements in conventional quality may not reduce trial participation because trials can always use the current standard of care as the control treatment, i.e., $d_2 = q$. This argument is fallacious because an increase in the surplus from conventional care increases the expected utility outside the trial, $U(q, p)$, more than it raises the expected utility inside the trial, $d_1 U(q^E, p^E) + (1 - d_1) U(q, p^E)$. To illustrate, consider the case where prices are the same inside and outside trials, $p^E = p$, as may be true for fully insured populations facing uncompensated trials, and where conventional care is used as an active control. Replacing the left-hand side of (2) with (**Error! Reference source not found.**) and

Turning to our empirical application, in order to compete with HAART, researchers may have modified the design of primary drug trials to offer greater value to enrolled subjects. Specifically, they may have switched from placebo-controlled to conventional-controlled trials. This switch may have offset some of the reduction in participation that might otherwise have occurred. In order to provide estimates of the effect of HAART on design, we estimate equation (**Error! Reference source not found.**) separately for enrollment in active-controlled trials and enrollment in placebo-controlled trials. The results are presented at the bottom of Table 3. In Panel B we see that HAART is associated with at least a 50% reduction in enrollment in active-controlled trials. However, the effect on placebo controlled trials is more dramatic: according to Panel C, enrollment falls nearly 100% if we do not control for trends. Durbin's alternative test rejects zero serial correlation, so we must be careful with the standard errors in Panel C. According to Panel D, however, taking first differences of enrollment in placebo-controlled trials addresses serial correlation. Moreover, the coefficient on the constant, which measures the trend in participation is nearly the same as the coefficient on HAART, suggesting HAART wiped out the pre-HAART positive trend in placebo-controlled trial participation.

simplifying yields the condition $E_F[U(q^E, p^E)] > U(q, p)$. Therefore, an increase in the quality of care lowers trial participation even when conventional care is used as an active control.

TABLES AND FIGURES

Table 1. Trial participation rate before and after HAART.

	Participation in primary drug trials		Participation in secondary drug trials	
A: Subjects with HIV	Obs.	Rate	Obs.	Rate
Pre-HAART (1990-1995)	6669	0.1149	6669	0.0393
Post-HAART (1996-2005)	5464	0.0728	5464	0.0194
Difference		-0.042		-0.020
P-value		[0.000]		[0.000]
Difference in difference				-0.022
P-value				[0.001]
B: Subjects on relevant drug	Obs.	Rate	Obs.	Rate
Pre-HAART (1990-1995)	3645	0.2093	2651	0.0977
Post-HAART (1996-2005)	4366	0.0912	2495	0.0425
Difference		-0.118		-0.055
P-value		[0.000]		[0.000]
Difference in difference				-0.063
P-value				[0.000]

Notes. This table compares the fraction of primary and secondary drug users that enrolled in primary and secondary drug trials, respectively, before and after the introduction of HAART in 1996. Observations are at the individual-year level. The p-values for the differences are from a t-test of the difference in trial participation rate before and after HAART. The p-values for the difference in difference is from a t-test of equality in the post-period coefficients from unweighted OLS regressions of trial participation on a constant and post-period for samples of primary and secondary drug users.

Table 2. Summary statistics.

			Pre-HAART (1990-1995)			Post-HAART (1997-2005)		
Category	Variable	Units	Obs.	Mean	Std. Dev.	Obs	Mean	Std. Dev.
AIDS drug use	Primary treatment	0/1	9483	0.442	0.497	10269	0.644	0.479
	Secondary treatment	0/1	9483	0.326	0.469	10269	0.313	0.464
Trial participation	Primary drug trial	0/1	9483	0.092	0.288	10269	0.051	0.22
	Placebo controlled	0/1	9483	0.041	0.199	10269	0.009	0.096
	Secondary drug trial	0/1	9483	0.032	0.175	10269	0.011	0.102
Employment	Full time	0/1	7817	0.748	0.434	8369	0.574	0.494
	Part time	0/1	7817	0.117	0.321	8369	0.135	0.342
Income	<$10,000	0/1	7703	0.142	0.349	9272	0.206	0.404
	$10,000-$19,999	0/1	7703	0.173	0.378	9272	0.16	0.367
	$20,000-$29,999	0/1	7703	0.213	0.41	9272	0.14	0.347
	$30,000-$39,999	0/1	7703	0.213	0.41	9272	0.127	0.333
	$40,000-$49,999	0/1	7703	0.145	0.352	9272	0.126	0.331
	$50,000-$59,999	0/1	7703	0.232	0.422	9272	0.207	0.405
	$60,000-$69,999	0/1	7703	0.007	0.086	9272	0.133	0.34
	>$70,000	0/1	7703	0.015	0.122	9272	0.004	0.062
	Refused to answer	0/1	7703	0.105	0.306	9272	0.079	0.269
Insurance	Any	0/1	7907	0.758	0.429	9233	0.737	0.44
	Government	0/1	7801	0.136	0.343	9438	0.325	0.469
	-Medicare	0/1	7795	0.042	0.202	9209	0.17	0.375
	-Medicaid	0/1	7798	0.081	0.273	9208	0.171	0.377
	-VA coverage	0/1	7795	0.038	0.192	9208	0.027	0.162
	Private	0/1	7807	0.665	0.472	9219	0.487	0.5
	-Individual	0/1	7084	0.146	0.353	9215	0.092	0.288
	-Group	0/1	7092	0.531	0.499	9219	0.431	0.495
	None	0/1	7907	0.358	0.479	9233	0.36	0.48
Health	Age	Yrs	7856	40	7	8336	45	8
	CD4 count	cells/mm3	7718	466	357	8368	563	315
	Viral load	copies/ml	3294	114547	313340	7479	29121	146625

Notes. Sample includes all individuals with HIV. Each individual-year observation is weighted equally. Not all individuals are observed each year. The sample size for a variable varies within a period (per or post-HAART) because of missing variables. For example, financial variables are only available after mid-1990 and viral load is measured for only half of HIV positive subjects in the pre-HAART period.

Table 3. Effect of HAART on number of individuals enrolled in primary drug trials.

	(1)	(2)	(3)	(4)
Panel A: Total subjects enrolled in primary drug trials				
HAART	-87.9***	-39.1	-95.6***	-66.6***
	-11.6	-25.7	-8.8	-13.9
Trend		-6.1**		-3.4**
		-2.4		-1.2
Constant	127.7***	155.1***	127.7***	143.0***
	-7.5	-14.1	-7.5	-9.6
Observations	16	16	15	15
Durbin's Alt (p-value)	0.18	0.69	0.84	0.21
Panel B: Total subjects enrolled in active-controlled primary drug trials				
HAART	-54.4***	-25.2	-60.0***	-47.8**
	-13.4	-24.1	-12.3	-18.8
Trend		-3.6		-1.4
		-2.3		-1.9
Constant	89.7***	106.1***	89.7***	96.1***
	-11.5	-17.7	-11.6	-16.4
Observations	16	16	15	15
Durbin's Alt (p-value)	0.22	0.36	0.73	0.83
Panel C: Total subjects enrolled in placebo-controlled primary drug trials				
HAART	-49.9***	-14.8	-53.1***	-21.1*
	(9.6)	(10.1)	(9.1)	(10.6)
Trend		-4.4***		-3.8**
		(1.2)		(1.4)
Constant	58.7***	78.4***	58.7***	75.6***
	(8.9)	(9.6)	(9.0)	(10.8)
Observations	16	16	15	15
Durbin's Alt (p-value)	0.26	0.02	0.00	0.00
Panel D: Total subjects enrolled in placebo-controlled primary drug trials				
Lagged subjects	0.7***	0.6***	0.6***	0.6***
	(0.1)	(0.1)	(0.1)	(0.1)
HAART	-8.9	-8.4	-15.8***	-16.5***
	(7.6)	(8.4)	(4.6)	(4.6)
Trend		-0.3		0.3
		(0.6)		(0.4)
Constant	9.2	13.0*	15.9***	12.4
	(7.5)	(6.5)	(4.8)	(7.2)
Observations	15	15	14	14
Durbin Alt (p-value)	0.42	0.46	0.84	0.81
Include 1996?	Yes	Yes	No	No

Notes: Dependent variable is total subjects in primary drug trials indicated at top of each panel. Regression in panel D includes first lag of dependent variable as a control. Regressions in columns (3) and (4) omit the transition year 1996. Robust standard errors in parentheses. */**/*** indicate p<.10/.05/.01. Durbin's alternative test checks whether the null of no serial correlation in residuals can be rejected.

Table 4. Effect of HAART on trial participation rate: identification off time-trend.

Specification	(1)	(2)	(3)	(4)	(4)	(4)
Panel A: HIV patients						
Trend	-0.003***	-0.001	-0.010	-0.064*	-0.054*	-0.044
	0.000	-0.001	-0.009	-0.033	-0.033	-0.035
Pre-HAART	0.081***	0.086***	0.064***	0.650***	0.644***	
	-0.008	-0.016	-0.017	-0.105	-0.072	
Post-HAART	-0.012	-0.004	-0.011	-0.161	-0.030	-0.263**
	-0.009	-0.011	-0.012	-0.158	-0.108	-0.128
Treatment effect	-0.093	-0.089	-0.075	-0.811	-0.674	-0.263
Wald test (p)	0.000	0.000	0.000	0.000	0.000	0.040
Observations	12,133	12,133	10,862	4,187	4,187	4,187
Panel B: Primary drug users						
Trend	-0.010***	-0.012***	-0.025**	-0.221***	-0.216***	-0.155***
	(0.001)	(0.002)	(0.011)	(0.040)	(0.041)	(0.042)
Pre-HAART	0.135***	0.147***	0.119***	0.895***	0.857***	
	(0.012)	(0.026)	(0.026)	(0.119)	(0.081)	
Post-HAART	-0.037***	-0.037***	-0.042***	-0.438**	-0.288**	-0.780***
	(0.013)	(0.014)	(0.015)	(0.171)	(0.120)	(0.145)
Treatment effect	-0.171	-0.184	-0.160	-1.333	-1.145	-0.780
Wald test (p)	0.000	0.000	0.000	0.000	0.000	0.000
Observations	8,011	8,011	7,421	3,377	3,377	3,377
Pre-window	1995	1995	1995	1995	1993-95	1990-95
Post-window	1997	1997	1997	1997	1997-99	1997-2005
Identification	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	No	Yes	Yes	Yes	Yes	Yes
Add'l covariates	No	No	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	Logit	Logit	Logit

Notes: Dependent variable is whether a subject participated in a primary drug trial. Sample includes years 1990-1995 and 1997-2005 and individuals indicated at top of each panel. Observations are at the individual-year level. Pre-period and post-period variables are dummies for the years indicated in last panel. Other controls as well as estimation method--OLS or logit--are indicated in last panel. Additional covariates include binned income and CD4 count. Coefficients identify change in trial participation rate relative to a linear trend. Standard errors, clustered at the individual level, are reported below coefficients. */**/*** indicates p<0.10/0.05/0.01. Treatment effect is the coefficient on post-HAART minus the coefficient on pre-HAART in the first 5 columns. In the last column it is simply the coefficient on post-HAART. Last line of each panel reports p-value for Wald test of whether the treatment effect is zero.

Table 5. Effect of HAART on trial participation rate: identification off secondary drug controls.

Specification	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel A: HIV patients												
Pre-HAART x PD trial	-0.003	0.021	0.022	0.227	0.007	0.022	0.011	-0.061				
	(0.021)	(0.040)	(0.039)	(0.296)	(0.013)	(0.028)	(0.028)	(0.286)				
Post-HAART x PD trial	0.024	0.043	0.026	0.523	0.022	0.035	0.020	0.905	0.028**	0.020**	0.012	1.870**
	(0.035)	(0.032)	(0.026)	(0.840)	(0.022)	(0.024)	(0.023)	(0.821)	(0.012)	(0.009)	(0.011)	(0.734)
Treatment effect	0.028	0.022	0.003	0.296	0.014	0.013	0.009	0.966	0.028	0.020	0.012	1.870
Wald test (p)	0.484	0.667	0.942	0.733	0.538	0.661	0.764	0.227	0.016	0.037	0.282	0.011
Observations	16,712	16,712	14,975	7,572	16,712	16,712	14,975	7,572	16,712	16,712	14,975	7,572
Panel B: Primary or secondary drug users												
Pre-HAART x PD trial	0.036	0.084*	0.087*	0.441	0.032**	0.074**	0.068**	0.501*				
	(0.023)	(0.047)	(0.047)	(0.329)	(0.015)	(0.031)	(0.032)	(0.303)				
Post-HAART x PD trial	0.021	0.022	0.006	0.233	0.016	0.016	0.011	0.671	0.044***	0.014	0.008	2.129***
	(0.039)	(0.031)	(0.025)	(0.876)	(0.026)	(0.026)	(0.025)	(0.848)	(0.014)	(0.017)	(0.019)	(0.741)
Treatment effect	-0.015	-0.062	-0.081	-0.208	-0.016	-0.059	-0.057	0.170	0.044	0.014	0.008	2.129
Wald test (p)	0.745	0.254	0.123	0.820	0.551	0.068	0.090	0.838	0.002	0.426	0.696	0.004
Observations	14,139	14,139	12,833	6,841	14,139	14,139	12,833	6,841	14,139	14,139	12,833	6,841
Pre-window	1995	1995	1995	1995	1993-95	1993-95	1993-95	1993-95	1990-95	1990-95	1990-95	1990-95
Post-window	1997	1997	1997	1997	1997-99	1997-99	1997-99	1997-99	1997-2005	1997-2005	1997-2005	1997-2005
Identification	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD
Individual FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Add'l covariates	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Estimation	OLS	OLS	OLS	Logit	OLS	OLS	OLS	Logit	OLS	OLS	OLS	Logit

Notes: Dependent variable is whether a subject participated in a trial. Sample includes years 1990-1995 and 1997-2005 and individuals indicated at top of each panel. Observations are at the individual x drug type x year level. Pre-period and post-period variables are dummies for the years indicated in last panel. Other controls as well as estimation method--OLS or logit--are indicated in last panel. Additional covariates include binned income and CD4 count. Coefficients identify change in the participation rate in primary drug trials relative to the change in the participation rate in secondary drug trials. Standard errors, clustered at the individual level, are reported below coefficients. */**/*** indicates p<0.10/0.05/0.01. Treatment effect is coefficient on post-HAART x primary drug user minus the coefficient on pre-HAART x primary drug user in the first 8 columns. In the last 4 columns it is simply the coefficient on post-HAART x primary drug user. Last line of each panel reports p-value for Wald test of whether the treatment effect is zero.

Table 6. Effect of HAART on trial participation rate: analysis of subsamples more robust to attrition.

Specification	(3)	(3)	(3)	(3)	(3)	(3)	(3)	(3)	(3)
Panel A: HIV patients									
Trend	-0.011	-0.011	-0.013*	-0.010	-0.009	-0.010	-0.028***	-0.026***	-0.024***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
Pre-HAART	0.063***	0.023***		0.081***	0.041***		0.164***	0.096***	
	(0.017)	(0.008)		(0.015)	(0.010)		(0.026)	(0.018)	
Post-HAART	0.017	0.013	0.022	-0.010	0.000	0.002	-0.036**	-0.021*	-0.061***
	(0.015)	(0.011)	(0.017)	(0.012)	(0.009)	(0.014)	(0.016)	(0.012)	(0.021)
Treatment effect	-0.045	-0.009	0.022	-0.091	-0.041	0.002	-0.200	-0.117	-0.061
Wald test (p)	0.044	0.488	0.178	0.000	0.001	0.910	0.000	0.000	0.004
Observations	5,995	5,995	5,995	7,996	7,996	7,996	5,478	5,478	5,478
Panel B: Primary drug users									
Trend	-0.027***	-0.026***	-0.024***	-0.011	-0.010	-0.011	-0.031***	-0.029***	-0.027***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
Pre-HAART	0.156***	0.069***		0.083***	0.041***		0.174***	0.096***	
	(0.033)	(0.019)		(0.019)	(0.011)		(0.032)	(0.021)	
Post-HAART	-0.012	-0.016	-0.051**	-0.015	0.004	0.001	-0.038**	-0.016	-0.063**
	(0.020)	(0.015)	(0.026)	(0.014)	(0.011)	(0.017)	(0.018)	(0.014)	(0.025)
Treatment effect	-0.167	-0.085	-0.051	-0.098	-0.038	0.001	-0.213	-0.112	-0.063
Wald test (p)	0.000	0.000	0.048	0.000	0.010	0.966	0.000	0.000	0.011
Observations	3,603	3,603	3,603	5,916	5,916	5,916	4,077	4,077	4,077
Sample	CD4 always >= 200	CD4 always >= 200	CD4 always >= 200	Surv. until 2000	Surv. until 2000	Surv. until 2000	Surv. until 2005	Surv. until 2005	Surv. until 2005
Pre-window	1995	1993-95	1990-95	1995	1993-95	1990-95	1995	1993-95	1990-95
Post-window	1997	1997-99	1997-2005	1997	1997-99	1997-2005	1997	1997-99	1997-2005
Identification	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Add'l covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Notes: Dependent variable is whether a subject participated in a primary drug trial. Sample includes years 1990-2005, excluding 1996. Sample includes individuals who have HIV or use a primary drug, as indicated at top of each panel, and who survived until the year 2000 or 2005 or had a CD4 count above 200 for the entire sample period, as indicated in the bottom panel. Observations are at the individual-year level. Pre-period and post-period variables are dummies for the years indicated in last panel. Other controls as well as estimation method (OLS or logit) are indicated in last panel. Additional covariates include binned income and CD4 count. Coefficients identify change in trial participation rate relative to a linear trend. Standard errors, clustered at the individual level, are reported below coefficients. */**/*** indicates p<0.10/0.05/0.01. Treatment effect is the coefficient on post-HAART minus the coefficient on pre-HAART in the first 8 columns. In the last 4 columns it is simply the coefficient on post-HAART. Last line of each panel reports p-value for Wald test of whether the treatment effect is zero.

Table 7. Effect of HAART on supply: analysis of exits and entrances.

	(1)	(2)	(3)	(4)
Dependent variable	Exits	Exits	Entrances	Entrances
Year 1996	36.0***	47.7***	5.3	7.5
	-5.4	-5.2	-4	-5.2
Years 1997-2005	-19.4***	8.9	-10.9**	-5.7
	-6.4	-9.1	-4.2	-7.6
Trend		-3.3***		-0.6
		-0.8		-0.6
Constant	43.0***	51.3***	24.7***	26.2***
	-5.4	-5.2	-4	-4.1
Observations	16	16	16	16
Durbin Alt (p-value)	0.79	0	0.98	0.7

Dependent variable in first (last) four columns is total subjects exiting from (entering into) primary drug trials. Sample includes years 1990-2005. Constant picks up exits or entrances in years 1990-1995. Robust standard errors in parentheses. */**/*** indicate p<.10/.05/.01.

Table 8. Average out of pocket drug spending by insurance status and trial participation.

	Average OOP drug spending			
Private insurance	In PD trial	Not in PD trial	Differ-ence	Obs.
No	318	151	-167	12,341
Yes	479	396	-83	16,242
Difference in difference			-84	
Medicare or Medicaid				
No	413	289	-124	22,801
Yes	392	400	8	5,783
Difference in difference			-132	

Notes: Table presents average out of pocket spending on drugs among individuals indicated in rows and columns.

Table 9. Effect of HAART by insurance status: separating price and quality effects.

Specifications						
	(1)	(2)	(3)	(4)	(4)	(4)
Panel A: HIV patients						
Trend	-0.005***	-0.004**	-0.006	-0.077**	-0.074**	-0.116***
	(0.000)	(0.002)	(0.005)	(0.036)	(0.035)	(0.036)
Post	-0.071***	-0.072**	-0.071**	-1.561***	-0.634***	-0.081
	(0.013)	(0.032)	(0.033)	(0.305)	(0.151)	(0.158)
Post x uninsured	0.075***	0.077**	0.073*	1.661***	0.743***	0.489***
	(0.015)	(0.038)	(0.039)	(0.322)	(0.168)	(0.157)
Treatment effect	0.075	0.077	0.073	1.661	0.743	0.489
Wald test (p)	0.000	0.041	0.060	0.000	0.000	0.002
Observations	6,063	6,063	5,975	2,479	2,479	2,479
Panel B: Primary drug users						
Trend	-0.012***	-0.015***	-0.021***	-0.464***	-0.477***	-0.441***
	(0.001)	(0.003)	(0.007)	(0.161)	(0.163)	(0.166)
Post	-0.099***	-0.106***	-0.102**	-1.903***	-0.725***	-0.416**
	(0.016)	(0.038)	(0.040)	(0.327)	(0.159)	(0.170)
Post x uninsured	0.083***	0.073	0.069	1.603***	0.497***	0.175
	(0.018)	(0.045)	(0.046)	(0.347)	(0.179)	(0.170)
Treatment effect	0.083	0.073	0.069	1.603	0.497	0.175
Wald test (p)	0.000	0.102	0.135	0.000	0.005	0.305
Observations	4,683	4,683	4,617	2,029	2,029	2,029
Pre-window	1995	1995	1995	1995	1993-95	1990-95
Post-window	1997	1997	1997	1997	1997-99	1997-2005
Identification	Time trend	Time trend	Time trend	Time trend	Time trend	Time trend
Individual FE	No	Yes	Yes	Yes	Yes	Yes
Add'l covariates	No	No	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	Logit	Logit	Logit

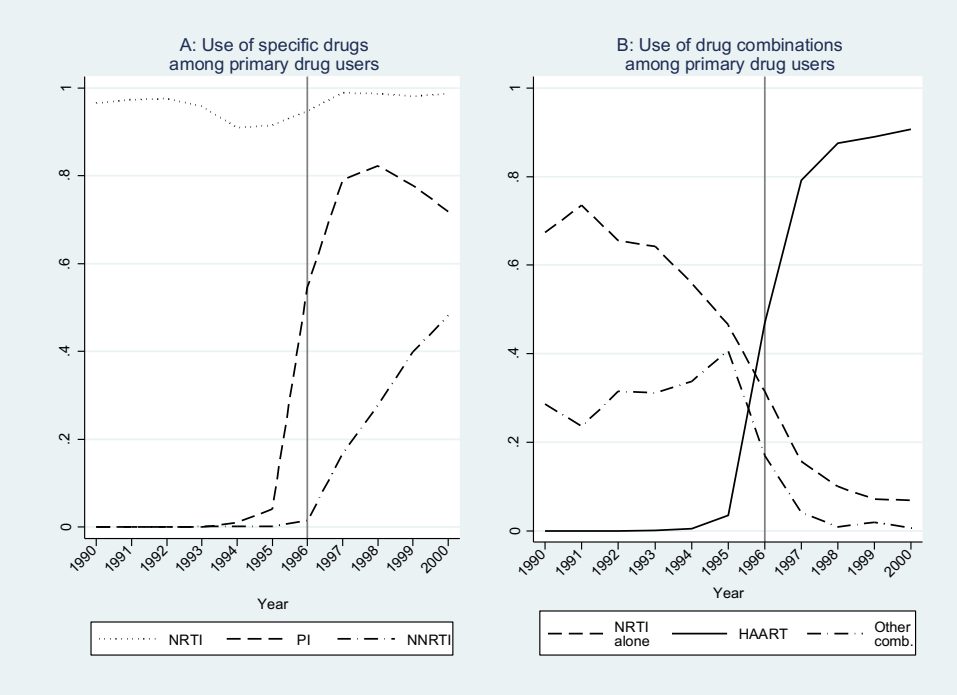

Figure 1. Effect of HAART on choice of AIDS therapy.

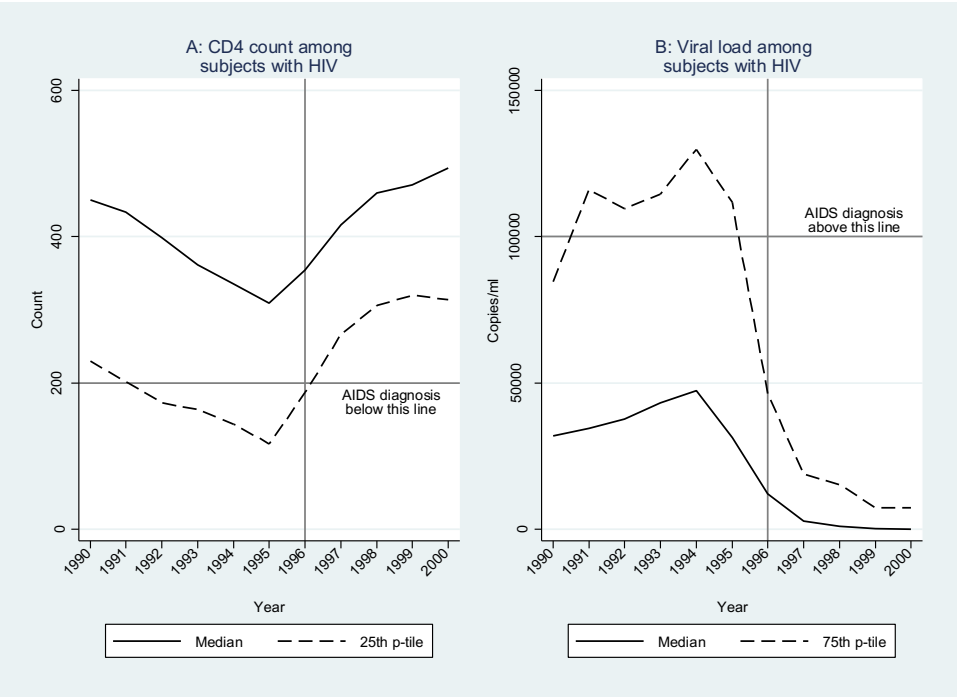

Figure 2. Effect of HAART on CD4 counts.

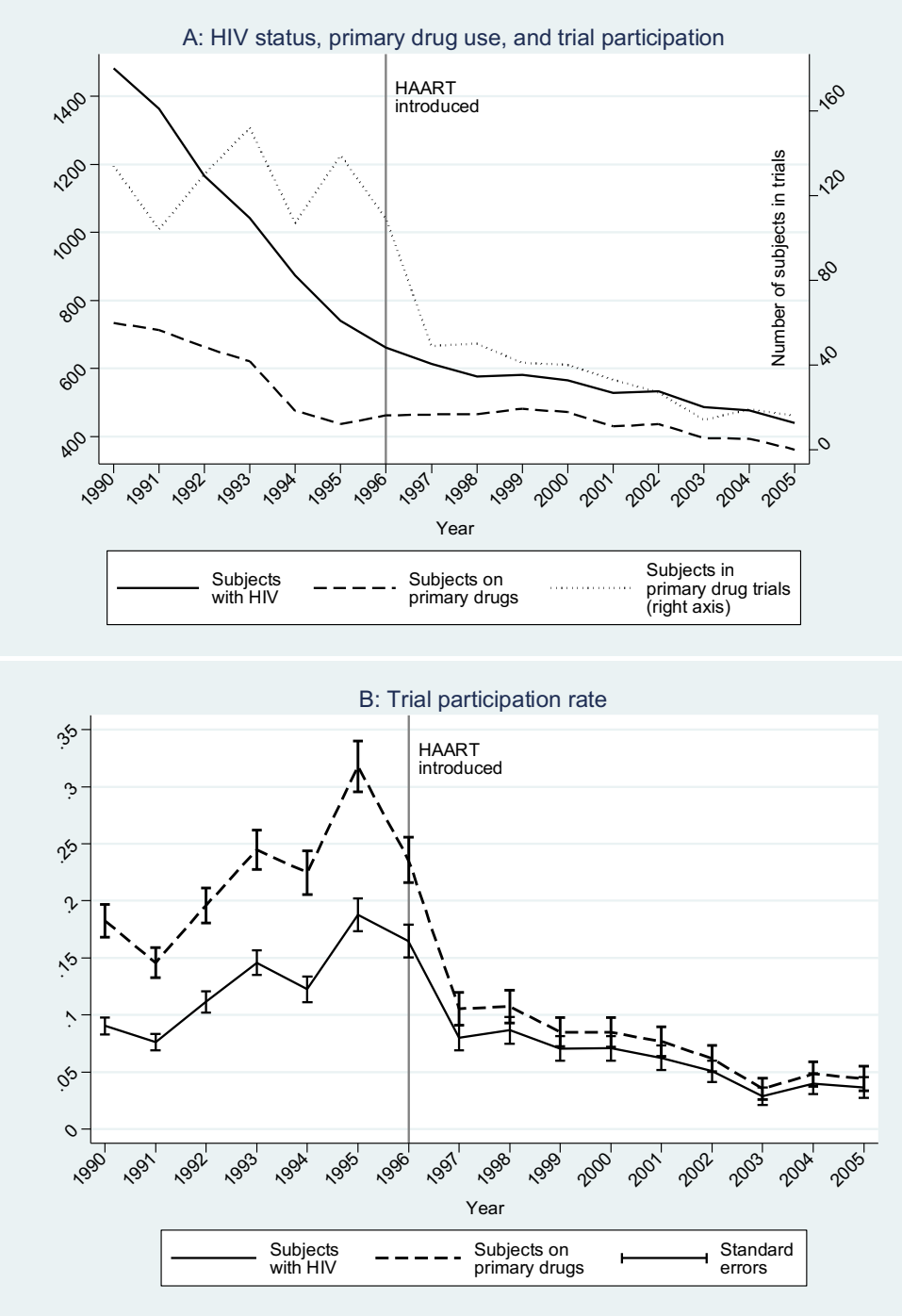

Figure 3. Effect of HAART on trial participation

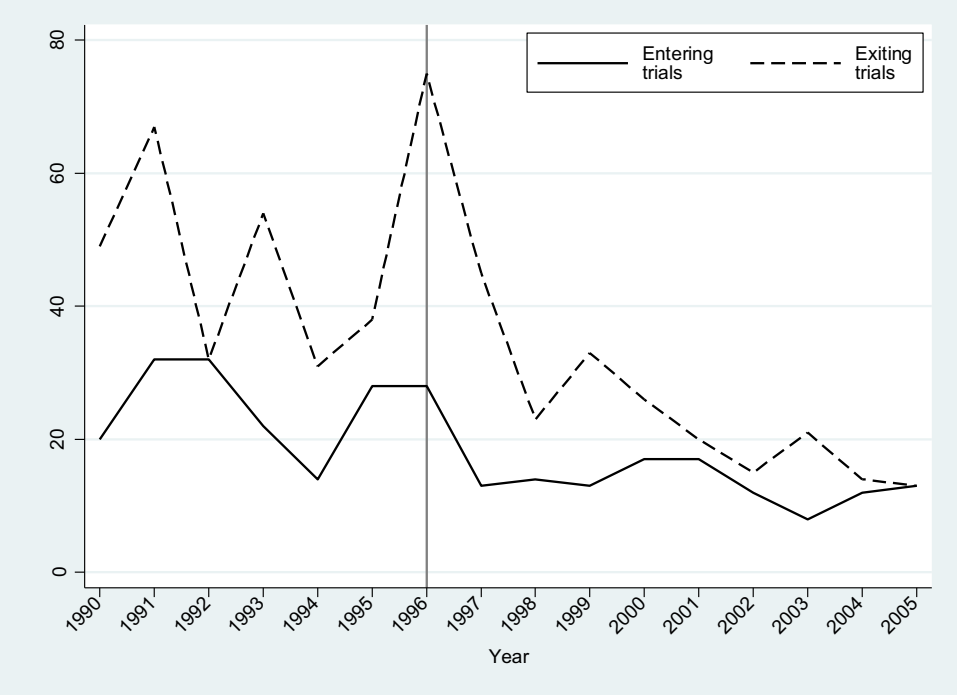

Figure 4. Effect of HAART on trial entry and exit.

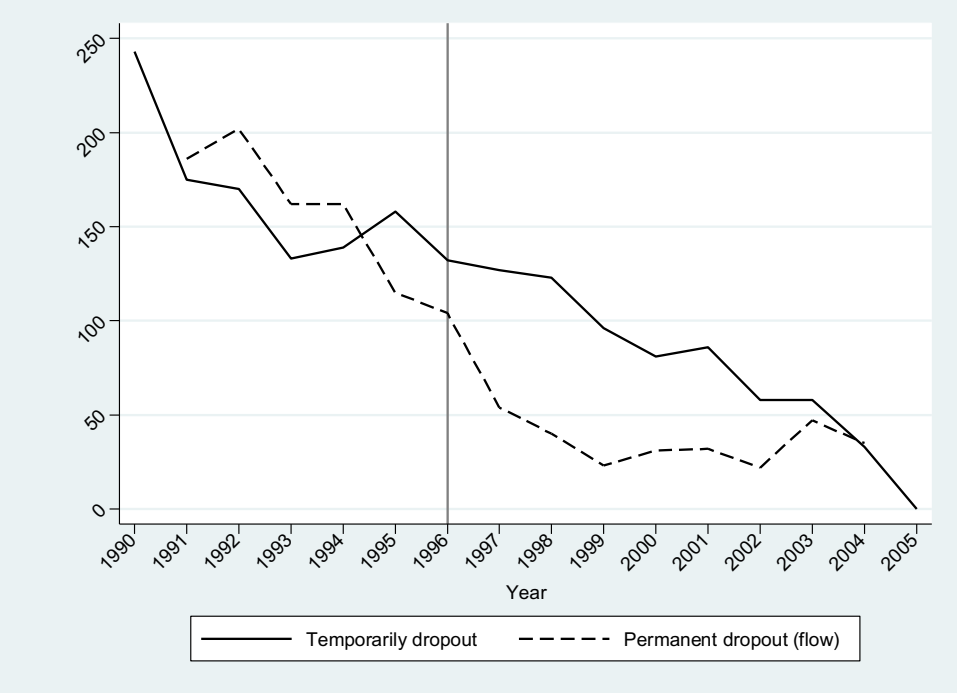

Figure 5. Temporary and permanent attrition over time.

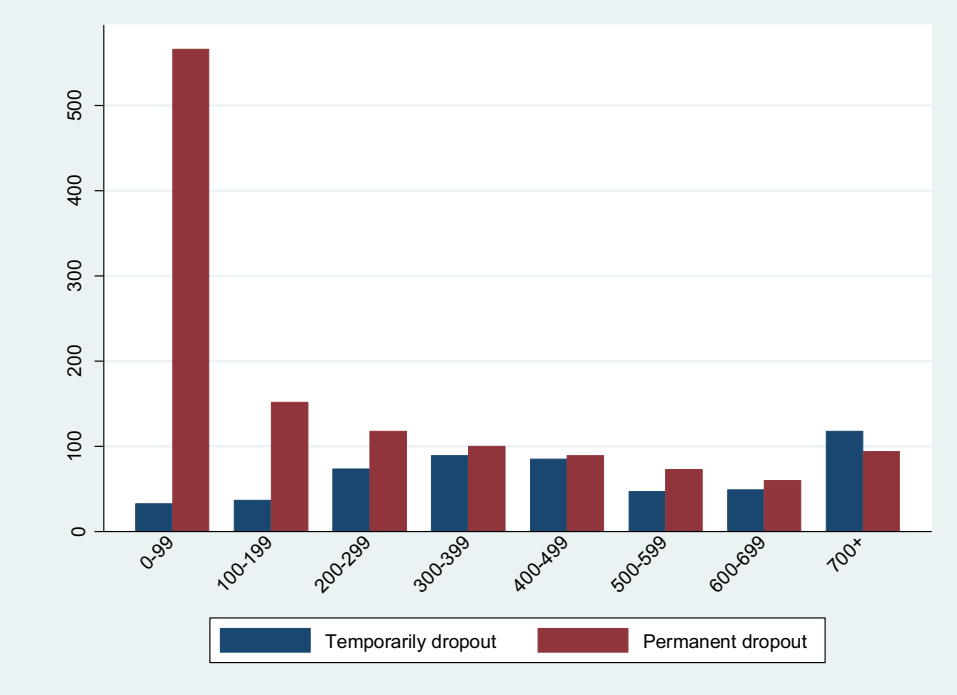

Figure 6. Temporary and permanent attrition, by CD4 count.

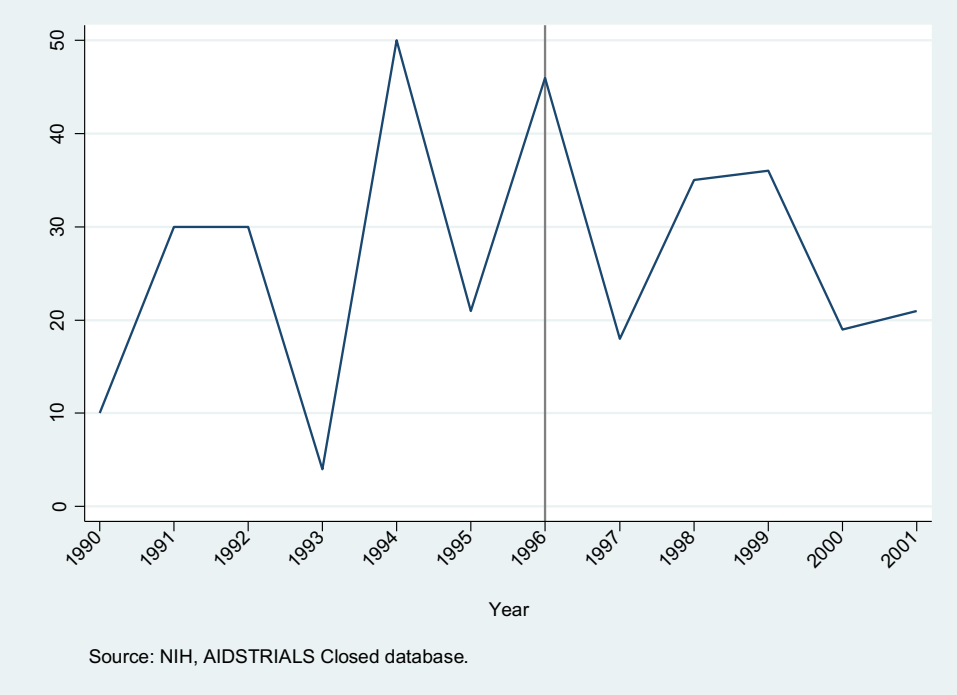

Figure 7. Termination of ACTG trials over time.

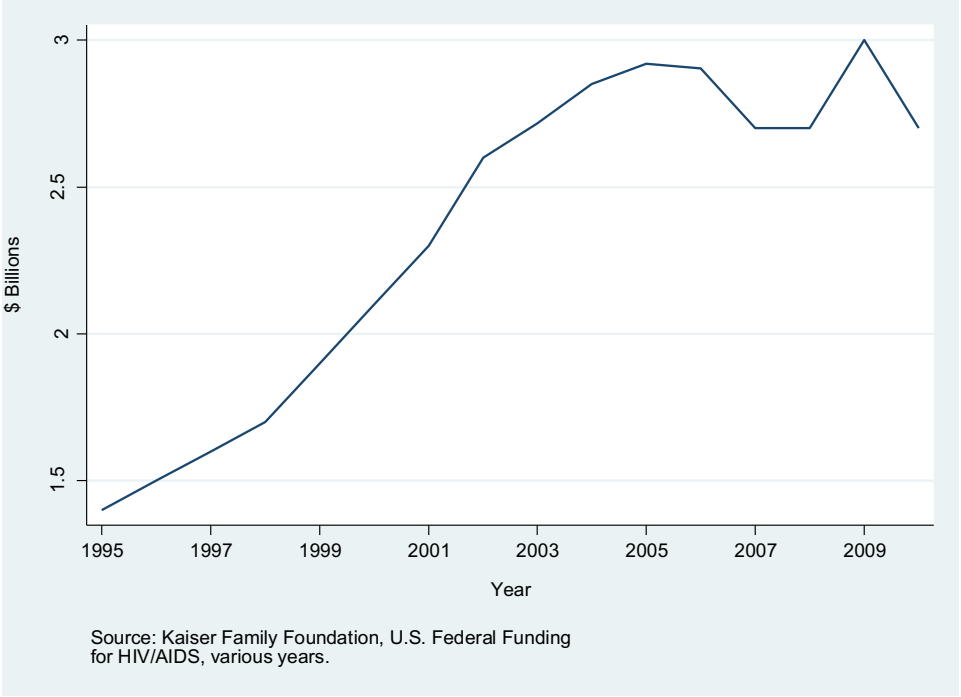

Figure 8. Federal government (NIH) spending on HIV/AIDS research.

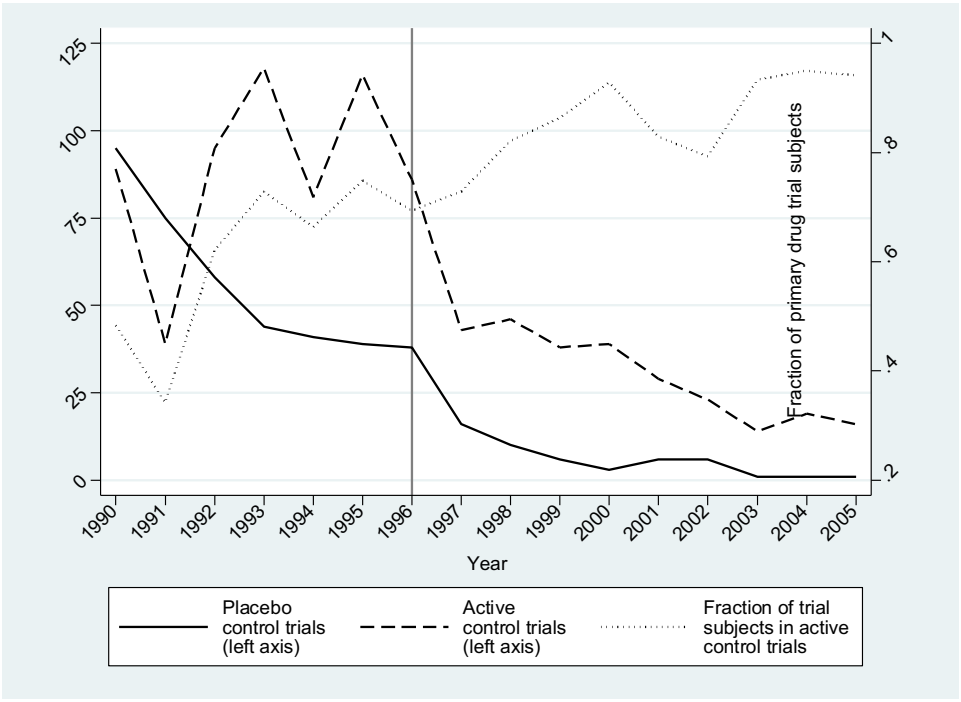

Figure 9. Enrollment in placebo and active control trials.