

NBER WORKING PAPER SERIES

CAN MEDICAL PROGRESS BE SUSTAINED? IMPLICATIONS OF THE LINK
BETWEEN DEVELOPMENT AND OUTPUT MARKETS

Anup Malani
Tomas J. Philipson

Working Paper 17011
<http://www.nber.org/papers/w17011>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2011

We thank 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, and HEC Montreal 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 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by Anup Malani and Tomas J. Philipson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Can Medical Progress be Sustained? Implications of the Link Between Development and Output Markets

Anup Malani and Tomas J. Philipson

NBER Working Paper No. 17011

May 2011, Revised September 2012

JEL No. I1,I11

ABSTRACT

There is considerable debate about the impact of health care reform on the growth in medical spending. Medical innovation is thought to be a central contributor to that growth. We argue that there is a unique linkage between reforms that affect output markets for medical care and medical R&D costs. This linkage is due to the fact that potential consumers of medical care are also potential participants in clinical trials that are required to develop new medical products. Therefore, reforms that increase the quality or reduce the price of already developed treatments reduce the incentive of patients to participate in trials of experimental treatments. This delays development and reduce the returns to innovation. We provide evidence of this “subject market effect” by considering the impact of changes in the quality of conventional care on development. We document a dramatic drop in trial recruitment following introduction of break-through HIV/AIDS therapies in 1996. We conclude by discussing additional positive and normative implications of the subject market effect that link input and output markets for medical products.

Anup Malani

University of Chicago Law School

1111 E. 60th Street

Chicago, IL 60637

and NBER

amalani@uchicago.edu

Tomas J. Philipson

Irving B. Harris Graduate School

of Public Policy Studies

University of Chicago

1155 E. 60th Street

Chicago, IL 60637

and NBER

t-philipson@uchicago.edu

There is considerable debate about the extent to which various health care reforms, including the new Patient Protection and Affordable Care Act in the United States, will affect the growth of health care spending. If there is any consensus on the causes of cost growth, it is that that growth is driven by medical innovation (Newhouse 1992). Thus, an important factor in determining how reforms affect cost growth is how they affect medical innovation.

Under the traditional economic view of innovation, the private benefits of R&D are driven by future variable profits in the output market. The price and quality of the existing and competing products in output markets affect these profits and thus these benefits of R&D. The costs of R&D are driven by input supply, e.g., the quantity and location of research talent.²

For medical innovation, however, a central input is research subjects for medical trials that are required before firms can market new medical products. Our central insight is that the subjects in trials for new medical products are also potential consumers of the existing, conventional care in output markets. An improvement in the quality or reduction in price of that conventional care may therefore reduce the incentive of patients to enroll in trials. We term this effect the “subject market effect”. As a result of this effect, the costs and duration of medical R&D are also driven by the quality and price of conventional care in output markets. This effect is in addition to the effect that price and quality have on R&D due to their impact on future variable profits.

This subject market effect has important, previously unrecognized implications for the relationship between health care reforms and spending growth. Examples of such reforms include comparative effectiveness requirements, expanded Medicaid eligibility, tax credits for the purchase of insurance, and price controls. If such reforms raise the quality or reduce the price of conventional care, they affect subject markets by lowering patients’ incentives to enroll in clinical trials. This, in turn, prolongs development and lowers innovative returns for medical products. Because development eats up the patent life, reforms may both delay and shorten the duration of patented profits. The bottom line is that, because most reforms aim to either raise the quality of care or lower its price, they may have non-standard effects on medical R&D and future cost growth due to their impact on subject markets.

We provide evidence of the subject market effect by considering how improvements in the quality of conventional care affect the development of new medical products. In particular, we examine trial recruitment before and after a 1996 innovation – highly active antiretroviral therapy (HAART) – that

² See Tirole (1991) for a general summary statement, and Weisbrod (1993) and Acemoglu et al (2006) for applications in health care.

affected the quality and price for treating HIV/AIDS. Using longitudinal data on over 1000 men in 4 U.S. cities, we find that this improvement in conventional care dramatically reduced participation in trials for HIV/AIDS treatments. This is most dramatically depicted by the solid line in Figure 1, which shows a sharp reduction in trial participation levels and a sharp reversal in participation trends starting in 1996 among users of HIV/AIDS drugs.³ We verified this impression from the raw data with regression analysis. To control for unobserved trends in trial participation we use a time trend or a control group of patients who did not experience a change in conventional care in 1996. We find that the trial participation rate fell 4-6% in absolute terms (50-75% in relative terms) immediately after HAART was marketed. This decline was largely driven by exit from ongoing trials, which implies that it was mainly due to a reduction in the supply of subjects to trials rather than a reduction in the demand for subjects through fewer trials. Our interpretation that the decline is a supply effect is supported by the presence of wage caps on research subjects imposed by ethical IRB boards. These caps imply excess demand for research subjects so that quantity changes are determined by changes in supply. Our overall findings suggest that improvement in the quality of conventional HIV/AIDS care due to HAART substantially increased the development times for HIV/AIDS-related therapies.

Finally, we discuss the many positive and normative implications of the subject market effect for medical products, such as the value of incremental innovation, how the linkage suggests a re-examination of Pareto optimal sample sizes (Philipson 1999), and the welfare effects of allowing firms to compensate patients for participating in clinical research.

The remainder of the paper may be outlined as follows. Section I presents a simple model of how reforms that raise the quality or lowers the price of conventional care affect the cost of developing new medical products. Section II discusses the equilibrium effects of reforms on development when free entry dissipates profits. Section III presents evidence suggesting that the new HIV/AIDS treatments introduced in the mid-1990s significantly reduced trial participations and slowed down development. Section IV concludes with further discussion of the normative implications of our analysis.

I. Development and Changes in Output Markets

³ HAART is an example of what we shall call a “primary” AIDS drug because it directly attacks the HIV virus. We use the term “secondary” AIDS drugs to describe medications that treat the consequences of AIDS, but not the HIV virus itself.

We begin by examining 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 economic determinants of development times and how they are affected by output market conditions.

A. The supply and demand for 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 utility, $U(q, p)$, that increases in the quality of care, q (synonymous 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 utility, $U(q^E, p^E)$. The quality of experimental care, q^E , is a random variable unknown at the time of entering the trial⁴ and p^E is the price of experimental care, potentially zero if treatment is subsidized by the trial.⁵ The uncertain quality of experimental care may be the product of experimental design, e.g., the random assignment of treatments, or uncertainty about the effects of the experimental therapy being studied. Importantly, ethical regulations banning compensation of subjects may prohibit paying for trial participation and thus too a negative price for experimental care, a feature to which we will return later.

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.⁶ We assume heterogeneity among patients, e.g.,

⁴ Conventional care may also, from the patient’s perspective, of course be of uncertain quality.

⁵ 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.

⁶ 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 (2010). A less formal version is presented by Dunn & Gordon (2005).

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 increasing in its price: $s_q \leq 0, s_p \geq 0$.⁷ This function depicts the fraction of the stock of patients willing to participate in the trial, given existing output market conditions.⁸

B. Changes in output markets and development times

Given a stock of patients with a disease, N , in steady state, a fraction, s , participates. We assume clinical trial participants are split evenly across M existing trials so that the flow of subjects, f , into a particular trial is

$$f = Ns/M \quad (3)$$

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:

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

This implies a 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.⁹ 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.¹⁰

1. Direct effects of reform-induced changes in output markets on development

⁷ 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.

⁸ 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 $s(p, q) = \frac{\exp\{E_F[U(q^E, p^E)]\}}{\exp\{U(q, p)\} + \exp\{E_F[U(q^E, p^E)]\}}$.

⁹ 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.

¹⁰ 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. 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.

Changes in output markets induced by, e.g., health care reforms, will also affect development times. The value of conventional care is an important driver of the trial participation rate. In particular, 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:

$$T_q = T_s s_q > 0 \quad T_p = T_s s_p < 0 \quad (5)$$

Because reforms alter the quality or price of available care, they will affect development incentives. For example, coverage expansions lower the out-of-pocket price of conventional care, price controls reduce overall prices, and comparative effectiveness regulations improve the quality of conventional care. Each therefore reduces patients' incentives to enroll in trials.

2. Indirect effects of reform-induced changes in output markets on development

Price and quality changes induced by reforms can also have indirect effect on development due to their impact on disease patterns or the experimental design of clinical trials.

a. Indirect effects through prevalence

Development time falls with the steady-state prevalence of a disease which in turn may be affected by reforms. Prevalence of a disease, 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.¹¹ For example, if conventional treatment can cure a disease, improving its quality or reducing its price will lower the prevalence of that disease. Conversely, if conventional treatment lowers mortality of a disease without curing it, improving it will increase the prevalence of that disease.¹²

Generally speaking, there are two effects of changes in output markets on development times:

$$dT/dx = T_s s_x + T_m m_x, x \in \{q, p\} \quad (6)$$

The first term captures the trial participation margin, holding prevalence constant. The second term captures the eligibility margin, which considers how prevalent the disease is and, thus, how many

¹¹ 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.

¹² 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 $m(p, q) = s(p, q)E[m_E(p_E, q_E)] + [1 - s(p, q)]m_C(p, q)$.

eligible patients there are.¹³ Holding prevalence constant, higher quality of conventional care, for example, discourages trial participation. For curative treatments, the eligibility effect of better conventional care reinforces the participation effect, implying even larger development costs if conventional care improves. For treatments that only reduce mortality rates, eligibility effects counteract participation effects. The net effect is uncertain.¹⁴

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).¹⁵ 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.¹⁶

b. Indirect effects through trial design

Another indirect channel through which reforms or other changes in output markets may affect development times is by forcing changes in the design of trials. For example, in order to compete with improvements in conventional care, researchers often modify the design of trials to include active controls¹⁷ or an increased share of patients receiving experimental care.¹⁸ 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 $w = (w_1, w_2)$, 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, w_1 . A second

¹³ The price of conventional has analogous effects on development times due to its effect on prevalence. An increase in price of a non-curative conventional care has offsetting effects on development times: it shortens the time by making recruitment easier for a given prevalence of eligible patients, but lengthens the recruitment time by lowering the prevalence through an increase in mortality.

¹⁴ In particular, the relative elasticities determine whether development time rises or falls. From (2), we see $dT/dx > 0 \Leftrightarrow |s_x x/s| > |m_x x/m|, x \in \{p, q\}$.

¹⁵ 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.

¹⁶ By contrast, Williams (2007) stressed the “pull” effect of diagnostics. By identifying new patients for treatment, diagnostics increase the returns to treatment innovations.

¹⁷ Welton et al. (1999) offer evidence that that using placebo as a control hinders recruitment.

¹⁸ Malani (2008) offers evidence that increasing the probability of receiving experimental care improves recruitment.

parameter, w_2 , is the quality of the control treatment, say placebo or active control.¹⁹ For these two design features, the expected utility of trial participation is given by

$$w_1 E_F U(q^E, p^E) + (1 - w_1) U(w_2, p^E) \quad (7)$$

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 $w(q, p)$ denotes the design under a given set of output market conditions, the total impact of a change in quality or price on participation is

$$ds/dx = s_x + s_w w_x, x \in \{q, p\} \quad (8)$$

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.²⁰

II. Innovative Returns and Equilibrium Development Times

Through subject market effects, changes in the output market affect the incentive to innovate by increasing the cost of R&D in a non-standard way. Specifically, changes in the output market affects 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, discuss equilibrium development time, and relate them to reforms that alter the price or quality of conventional care.

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

$$NPV(q, p) = -F + \beta^{T(p, q)} A(T^P - T(p, q)) \Pi(q, p) \quad (9)$$

¹⁹ 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.

²⁰ 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., $w_2 = q$. 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, $w_1 U(q^E, p^E) + (1 - w_1) U(q, p^E)$. 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 (7) and 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.

Here, F is the fixed cost of development, $\Pi(q, p)$ is the reduced form of per-period expected variable profits after market entry, given the quality and price of the conventional care and the distribution of experimental outcomes. 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 operates 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.²¹ 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:

$$\frac{dNPV}{dx} = V\Pi_x + V_T T_x \Pi, x \in \{q, p\} \quad (10)$$

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 market 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 increase in the price of conventional care.

²¹ Philipson and Sun (2010) offer estimates of the magnitudes of delay costs for a number of specific classes of drugs.

Subject market 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.²²

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-called 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 (10), while the patient price effect controls the signs of the second term in (10). 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 market effects – are driven by the price faced by consumers.

B. Free Entry and the Equilibrium Duration of Development

An important feature that distinguishes the labor market for human research subjects from other labor markets is that wages are capped by ethical rules. Bioethicists frown upon research compensation that encourages subjects to enroll in trials, even when it may not be medically prudent to do so. 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).

As a result of the 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,

²² 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.

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, paradoxically, delays 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 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.²³

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:

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

²³ 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 $M = TNs/n$.

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 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.

C. Free Entry and Inefficiently Long Development Times

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, $NPV > 0$, as opposed to zero as in the case of free entry, $NPV = 0$. In other words, 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.

III. Evidence of the link between development and output markets for AIDS innovations

The novel link we identify between output markets and medical R&D is premised on the idea that patients compare conventional treatment to experimental treatment when deciding whether to enroll in clinical trials. Thus, an improvement in the quality or price of the conventional treatment reduces the supply of patients into clinical trials. In this section we examine empirical evidence for this link in the case of an innovation in AIDS treatment, focusing on predictions about the effect of output quality and price on participation rates: $s_q < 0, s_p > 0$.

First, we present reduced-form estimates of how participation rates in HIV/AIDS trials changed after the approval of a breakthrough AIDS treatment – Highly Active Antiretroviral Treatment (HAART) – in 1996. We document that trial participation fell by 4-6 percentage points the year after HAART was introduced. Indeed, whereas trial participation was increasing by roughly one percentage point per year prior to HAART, it declined almost a half percentage point per year in the years after. This result is robust to the inclusion of various controls for unobservable trends on trial participation independent of HAART.

Our evidence suggests that the main cause of the decline in participation after HAART was a reduction in the supply of subjects rather than a reduction in demand for further HIV/AIDS research. HAART changed participation rates primarily by increasing exits in the short run from pre-existing trials rather than reducing entry into new trials. Those exits are a pure supply response since they hold demand (trials) constant. Finally, we show that the change in trial participation is explained largely by the improvement in the perceived quality of AIDS treatment rather than changes in price.

A. Background on AIDS and HAART

HIV is a virus that infects CD4 T-cells, an important component of the human immune system. When a patient's CD4 count falls below 200/mm³, the patient becomes acutely vulnerable to non-HIV infections, so-called secondary infections, and is classified as having AIDS (U.S. Centers for Disease Control & Prevention 1993).²⁴

HIV and AIDS are treated in two ways. First, the patient is given antiviral medications to suppress replication of the HIV virus. The impact of these medications is gauged by their effect on a patient's CD4 count. Effective antiviral therapy can prevent an HIV patient from progressing to AIDS but does not eliminate HIV infection. Therefore, we shall 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 HIV, rather than HIV 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 and secondary AIDS drugs at the same time; indeed this is true for the vast majority of AIDS drugs users

²⁴ 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.

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).

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. However, real-time monitoring has not been shown to have a substantial impact on viral loads as it only tracked those loads.²⁵ To simplify our exposition, we refer to both innovations as HAART because it is the more significant of the two innovations.

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;

²⁵ In 1997, there was a technical change in FDA regulation of HIV drugs related to viral load monitoring, but it did not affect the way HIV trials were conducted before or after HAART (Murray 2011). Before 1997, 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 HIV-RNA counts 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 studies 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.

moreover, for uninsured persons, HAART may also have increased the price of conventional care, lowering trial participation for insured persons relative to uninsured persons.

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 2005.²⁶ We first isolate a subsample of 2,266 unique subjects observed in the period 1990-2005.²⁷ 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 trial participation.²⁸

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.

Our empirical analysis will focus on the trial participation of the 1,465 individuals in our subsample who actually took primary or secondary drugs at some point during the sample period and the source (in trial or outside trial) of their drug regimens. Of these individuals, 215 took only primary drugs, 56 took only secondary drugs, and 1194 took both a primary drug and a secondary drug at some point.

C. Descriptive Statistics

Table 1 provides descriptive statistics for the subjects who took any AIDS drugs. 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 (1996-2005) to provide an unconditional assessment of the effect of HAART. Roughly 60% of subjects have some medical insurance before HAART and 70% after HAART. Moreover, there is a substantial change in the sources of insurance following HAART. The most common source of insurance is private coverage which increases nearly 6%, from 47% pre-HAART to 53% after. Government insurance also increase, covering 14% of the sample pre-HAART but 23% after. The most important program for government

²⁶ More information on the MACS study is available at <http://www.statepi.jhsph.edu/macs/mac.html>.

²⁷ We begin our analysis in 1990 rather than 1984 because MACS started asking about trial participation in 1990.

²⁸ 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.

insurance after HAART is, surprisingly, Medicare and not Medicaid. This is because Medicare covers not just the elderly, but also the long-term disabled population. That said, our sample is neither extremely wealthy nor impoverished. Roughly 70% are employed before and after HAART, with a median income between \$30,000 and \$39,000.

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 2 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.²⁹ These results are consistent with the summary statistics in Table 1, which show that the number subjects using primary drugs increased from 62% to 80% and the number of subjects using secondary drugs fell from 38% to 20% after HAART.

Panel B 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 drug regimens (NRTI only, HAART, and other non-HAART regimen) in Panel B. The main takeaway is that HAART use skyrocketed after 1995, while the use of other regimens fell.³⁰

The dramatic rise in usage of HAART after 1996 led to a dramatic improvement in subjects' health. According to Table 1, after HAART the average CD4 count in our sample increased from 291 to 494/mm³. (Recall that 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 140,305 to 22,465 copies/ml. Figure 3, 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/mm³), but after HAART, their CD4 counts improved to the point where, by 1997, the 25th percentile subject did not have AIDS. Likewise,

²⁹ These are usually antiviral drugs that have proven effective against other sexually transmitted diseases, such as hepatitis. None have proven effective against HIV/AIDS.

³⁰ 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.

Panel B shows that viral loads for the median patient fell dramatically after 1996. The effect was even more pronounced for the 75th percentile subject in our sample. Until 1995, this individual had loads in excess of the 100,000 copies/ml threshold used to identify at risk HIV patients. Within a year after HAART, his load fell so that he was no longer at risk.

Finally, we turn to trial participation. According to Table 1, 7.7% of subjects were in clinical trials for either primary or secondary AIDS drugs prior to HAART. However, after HAART was introduced, only 4.6% were participating in trials.³¹ Figure 1 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 30% in 1995, and then trial participation plummets and eventually settles at 5% by 2005, roughly half the level in 1990.³² Our interpretation of why trial participation for primary drugs rose dramatically before 1996 is that patients revised upwards their beliefs about the quality of experimental therapy as news leaked about health improvements among trial participants. Within our model, this means that $F(q_E)$ fell between 1992 and 1995. This likely led to a surge in the supply of subjects for trials just prior to approval of HAART.

The rate of trial participation is calculated by taking the number of subjects on a drug in a trial and dividing by the number of subjects on that drug. We want to ensure that the post-HAART drop in trial participation rates for primary drugs was driven by changes in trial participation (the numerator) rather than changes in the population of primary drug users (the denominator), which spiked after HAART according to Table 1. To do so, we take the first difference of a subject's trial participation to determine trial entries and exits over time. We calculate entry rates by dividing entries by the total number of subjects not in trials the prior period and the exit rates by dividing exits by subjects in trials the prior period. We can rule out a change in composition of users as a driver of changes in participation by focus in on exit rates because individuals previously in trials were already on primary drugs prior to 1996. As Figure 4 indicates, exit rates rose from 15% to over 20% in 1996 and 1997.

E. Regression analysis

1. Main results

Our primary strategy for estimating the effect of HAART on trial participation involves a regression that examines the level of participation pre and post HAART:

³¹ 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).

³² The pre-HAART average participation rate in Table 1 is somewhat lower than the pre-HAART average in Figure 2 because the two use different weighting schemes. The former gives each subject equal weight.

$$s_{it} = \beta_1 PRE_HAART_t + \beta_2 POST_HAART_t + \beta_3 TREND_t + [controls] + e_{it} \quad (11)$$

The dependent variable is an indicator variable 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. In specifications presented in this paper, the pre-period is 1995, the year before HAART is introduced and the post-period is 1996, the year after HAART was introduced. (HAART was introduced in December 1995-January 1996.)³³ We include a linear trend in the regression above to underlying unobserved trends in participation. Thus the treatment effect is, to some extent, identified off a linear trend.

Although our theory focuses on the level of trial participation, much of the effect of HAART will manifest through changes in the level 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. This was the case for 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. 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). We believe this explains the sharp reversal in trial participation trends after 1996 in Figure 1.

Our secondary strategy for estimating the effect of HAART on trial participation involves a regression that examines changes in participation levels pre and post HAART:

$$s_{it} = \beta_1 PRE_HAART_TREND_t + \beta_2 POST_HAART_TREND_t + [controls] + e_{it} \quad (12)$$

The dependent variable remains the level of trial participation. However, the treatment effect is captured by the difference between coefficients on a pre-HAART trend and a post-HAART trend. We employ two sets of windows for this comparison. The first set compares the pre-period trend in 1992-1994 with the post-period trend 1997-1999 to capture the medium term effects of HAART. The second compares the pre-period trend in 1990-1995 to the post-period trend in 1996-2005 to capture the long term effects on trends.

³³ In a previously posted version of this paper, we also report results with 3 year pre (1992-1994) and post (1997-1999) windows and longer pre (1990-1995) and post (1996-2005) windows that span the entire sample period. The results, also available from the authors, largely confirm the impression left by Figure 1.

For each of our two strategies 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 specification 3 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.

Our initial results are presented in Table 2. The three panels report results for the three different proxies for HAART. Focusing on linear specifications, HAART appears to lower trial participation rates by 4-6 percentage points in the short-run (Panel A) in linear specifications. To put that in perspective, recall from Table 1 that the trial participation rate was 7.7% prior to HAART. The immediate decline in participation in 1996 dropped level that by half to three-quarters.

Given the picture of trial participation painted in Figure 1, levels are not the best depiction of participation prior to HAART: participation rates were increasing dramatically until 1995. Comparing pre and post trends, we find that, in the medium term, HAART converted an upward trend of an additional 1-2% of all sample members participating each year into a decline of 5 percentage points in the participation rate each year (Panel B). In the long run, the trend reversal was more moderate, from an additional 2-3 percentage point per year to reduction of 3% per year (Panel C). The logit results suggest, if anything, stronger effects from HAART. The difference in measure pre and post levels and trends are all statistically significant.

These results suggest a reduced form relationship between innovation and equilibrium trial participation. The remainder of this section undertakes two tasks. First, we try to confirm this reduced form relationship is robust. Second, we determine what the reduced form relationship tells us about specific parameters in our theoretical model.

2. Robustness of reduced form 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 this is to include a linear time trend as a control in our basic regression in (11). 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 the effect on both prescription categories is the same, then we can employ trial participation in secondary drug trials as a control for those secular changes. One reason we are confident that the secular changes affect both groups identically is that the majority of subjects on primary AIDS drugs also took secondary AIDS drugs at the same time (see Table 1), with the latter serving as a safety net in case the former do not work.³⁴

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 1, the introduction of HAART also led to a reduction in secondary drug use. This may not be 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, perhaps that would affect the rate at which secondary drugs were procured in trials conditional on positive demand for such drugs.

We implement a difference-in-difference identification strategy with the following variant of regressions (11) and (12):

$$s_{ijt} = \beta_1 HAART_t + \beta_2 PRIM_{ijt} + \beta_3 (HAART_t \times PRIM_{ijt}) + [controls] + e_{it} \quad (13)$$

The sample includes individuals who used some primary drug or some secondary drug at time t . Because individuals may use both types of drugs at the same time, the sample sometimes contains two observations for an individual in a period, one for each drug type. The dependent variable is an indicator for whether individual i participated in trial for a particular drug type j . The $HAART_t$ variable represents

³⁴ 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.

the pre- and post-HAART indicator or trend, depending on whether we are modifying regression (11) or (12). We added the indicator $PRIM_{ijt}$ for whether the drug type is a primary AIDS drug and an interaction between this indicator and our proxy for whether HAART has been introduced. Aside from these changes, we estimate this regression model just as we did equation (11) and (12).

Regression results in which unrelated trends in participation are captured by time trends are presented in Table 3. The first two columns contain the results with a linear time trend and the third and fourth columns present results with a quadratic time trend.³⁵ In either case, HAART reduced participation by about 16% in the very short run (Panel C) and 18% or 53% in the long run (Panel A), though the effects are smaller than in the base specification from the previous subsection. In the medium run (Panel B), HAART has a positive effect on participation when linear time trends are employed and an insignificant or negative effect when quadratic trends are used. These medium-term results are probably revealing more about how well linear or quadratic trends capture unrelated participation trends than about the medium term effects of introducing HAART.

The results from the difference-in-difference specification are presented in Table 3. These suggest that HAART had a small and insignificant effect in the short-run. This is not surprising given the immediate decline in trial participation among secondary drug users depicted in Figure 1. In the medium- and long-run, however, HAART had a significant effect on trial participation trends. In the medium term, while participation was rising by 1.3 percentage points per year prior to HAART, they began to fall by 4 percentage points per year after HAART. In the long term, participation went from rising 2 percentage points per year to essentially remaining flat.

Because Figure 1 suggests a temporary response to HAART from secondary drug users and the difference-in-difference estimates are roughly the same as those in Table 2 for medium and long-run trends, we use regressions (11) and (12) 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

³⁵ The specifications presented always include individual fixed effects and time varying controls for income and CD4 count. The odd columns estimate an OLS model and the even columns use a logit model.

permanently due to death.³⁶ 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 around in 1996 which would suggest that the spike observed in Figure 1 and documented in Panel A of Table 2 are 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. Specifications (3) and (4) in all our analysis 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 2 but on a subsample of subjects whose CD4 counts never dip below 200 in our sample frame. The first two columns of Table 4 reports the results of specifications (3) and (4). We find that the effects of HAART among non-AIDS patients are if anything stronger in the short and long runs, and always statistically significant.³⁷

Third, to address changes permanent dropout trends after HAART, we also re-run the analysis reported in Table 2 but on a subsample of subjects that we confirmed were alive as of 2000 and a subset of subjects confirmed alive as of 2005.³⁸ 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 four columns of Table 4. As with the non-AIDS sample, we find the effects of HAART in each of the long-lived samples we examine

³⁶ Because of temporary drop out, our regression sample starts in 1990 with 693 subjects (667 primary drug users) rather than all 1,465 unique subjects (1,409 unique primary drug users). Because the rate of permanent drop out exceeds the rate of entry, the sample size falls monotonically to 372 subjects (363 primary drug users) by 2005. On average, each subject is observed for six years during the sample period.

³⁷ We conducted the same exercise with HIV viral load rather than CD4 count and obtained similarly strong results.

³⁸ 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 we make for identification is that the probability of observing the dependent variable is independent of HAART.

are somewhat stronger in short-run levels and long-run trends, and somewhat smaller medium run trends. However, HAART has a significant depressive effect on trial participation in all periods. This implies that there is no substantial survival bias in our estimates. Therefore, we use the entire sample – not just subjects alive in 2000 or 2005 – for our subsequent empirical analysis.

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.

3. Explaining the observed changes in the participation rate

The previous subsections identified the effect of HAART on trial participation to illustrate a robust reduced form relationship between innovation and equilibrium trial participation. Given our model of trial participation, one might wonder why observed participation fell after HAART. First, did HAART reduce the willingness of patients to participate in trials – the supply side of the market – or did it lower the number of trials that tried to recruit patients – the demand side of the market? Second, if HAART affected supply, how can the change be decomposed into effects due to changes in conventional care’s quality versus price? Third, since the implicit wage for a trial participant is affected by the design of the trial, how did HAART affect participation, holding the design of trials constant? We address these questions in turn.

a. Supply versus demand response

For a number of reasons, we think that the reduced form relationship between HAART and equilibrium trial participation is driven by changes in the supply of subjects rather than demand for subjects. First, Section II.C documented that prices are capped 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 manifests as development time. This sort of rationing implies that the *per period* participation rate, s , is affected by supply but the duration of development, T , is affected by demand. Therefore, our estimates of the effect of HAART on the per period trial participation rate identifies the supply behavior of subjects. Even if the introduction of HAART reduced demand by decreasing, N , we predict it would not cause the observed decline in trial

participation per period.³⁹ In other words, a sufficient identifying assumption for the effect of HAART on supply is that the equilibrium wage is determined by a binding cap.

Second, we investigate whether HAART affected the rate of exit from existing clinical trials, a response that is likely to be driven only by supply side factors. 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.⁴⁰ 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 the trial clearly works or that drug has a severe side effect (Cannistra 2004).⁴¹ Ethics rules do not permit cancellation because some other drug outside the trial works well. Empirical support for this view comes from 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, trial 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 good 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. The regression analyses reported in Table 5 also provide additional evidence of HAART’s impact. We conduct the same analysis as in Panel A of Table 2 except that we replaced the dependent variable with exit from trials and confined the sample to subjects who were enrolled in trials the previous year. Table 5 reports the resulting coefficients not only on the 1995 and 1996 indicators, but also on indicators for medium term pre (1992-1994) and post (1997-1999) indicators, in order to paint a picture of the full timeline of exit rates around HAART. The results suggest that exit rates increased by up to 24 percentage points the year after the introduction of HAART. The short run change is significant in all specifications except 3, where our Wald tests yield a p-value of 0.06. Moreover, the

³⁹ 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.

⁴⁰ 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.

⁴¹ This policy may, of course, also be a result of the excess demand for subjects, making “firing” of them unlikely.

medium term level response, comparing 1992-1994 with 1997-1999 is as large as the short-run effect and always significant.⁴²

While the exit rate results suggest that HAART reduced the supply of trial participants, they only reveal the behavior of subjects who were already in clinical trials, pre-HAART. They do not tell us about the entry of new subjects into trials, let alone whether that entry was driven by supply factors or demand factors. Although we cannot discriminate between the influences of supply and demand on entry rates, we can use the results from analysis of all subjects and of the subset of previous trial participants to bound the supply-side effect of HAART.

We begin by decomposing the total change in trial participation into the change in participation among the subjects who were not already in trials and those who were: $ds = \sigma_{NP}ds_{NP} + \sigma_P ds_P$, where σ_{NP} and $\sigma_P = 1 - \sigma_{NP}$ are population shares of non-participants and participants, respectively. To compare the relative roles of exit and entry on total participation, we estimate the regression used in the previous subsection, except that we replace the dependent variable with entry into clinical trials and confine the sample to subjects who were not enrolled in a trial the previous year. The results are presented in Table 6. Surprisingly, we find that entry rose in the short run by 14 percentage points, which is consistent with the graphical data in Figure 7. From these results we infer that the reduction in total trial participation in the short run can only be explained by an increase in exits. Since exits are driven only by supply, the reduction in total participation is a lower bound on the supply effects of HAART.

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 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

⁴² Our regression also finds an effect on exit rates in the long-run. However, that result is difficult to explain with HAART. Many who exited long after HAART was introduced likely also entered after HAART was introduced. Therefore, HAART could not have affected their exit decision.

25 in 1987 to 125 in 1997 (Neumann & Sandberg 1998). In 2001, the number had fallen to 98 (PhRMA 2002), and stayed constant at 100 until the 2010 (PhRMA 2010).

b. Quality versus price response

Even if HAART largely operated through supply effects, to what extent did HAART operate through the quality versus price of conventional treatment? Prior literature and the graphical evidence in Figure 3 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 previous analysis argued that the quality improvement would reduce the supply of trial participants, while a price increase would increase supply. This suggests that the previously estimated 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 address 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. The specific model we estimate is

$$s_{ijt} = \beta_1 HAART_t + \beta_2 NOINS_{ijt} + \beta_3 (HAART_t \times NOINS_{ijt}) + [controls] + e_{it} \quad (14)$$

The $HAART_t$ variable represents the pre- and post-HAART indicator or trend, depending on whether we are modifying regression (11) or (12). $NOINS_{ijt}$ 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.

The results are presented in Table 7. We find that that HAART reduced trial participation among insured populations. Relative to this group, uninsured populations experienced an increase in participation levels and trends, though this effect was insignificant in many specifications. This implies that the quality improvement due to HAART significantly reduced the supply of participants. The price increase after HAART, however, increased participation, though perhaps not significantly.

c. Unrelated trends in trial design

The decision to participate in a trial depends on the difference in the value of a trial and of conventional care. The value of a trial in turn depends on what the control group in a trial receives. In a placebo-controlled trial, subjects receive the equivalent of a sugar pill. In an active-controlled trial, however, the controls receive some other primary AIDS drug, often the best conventional treatment available outside the trial. Since active control trials offer more value to subjects, researchers may have responded to competition from HAART by switching from placebo to active-controls in new trials. This switch may have increased the expected value of trials and thereby offset some of the reduction in trial participation after HAART.

Our prior estimates of the effect of HAART are not conditioned on trial design. In order to provide estimates of the effect of HAART conditional on design, we estimate equation (14) separately on a subsample that focuses on active-controlled trials by excluding placebo-controlled ones (first two columns of Table 8) and a subsample that focuses on placebo-controlled trials by excluding active controlled ones (last two columns of Table 8). For active-controlled trials, we find a negative effect of HAART in short-run levels and medium- and long-run trends, all largely driven by quality changes. For placebo-controlled trials, we find significant effects only in trends estimated via logit regression. .

There is a second manner in which the design of HIV trials may have changed after 1996. The introduction of real time monitoring of viral loads by physicians outside of trials may have led new trials started after 1996 to give enrolled subjects real time information on their viral loads.⁴³ Because this design change would make it easier for subjects in trials to learn about whether the treatment they were assigned was effective, it would facilitate attrition from trials. For this reason, it is possible many new trials resisted the design change. However, if the design change were adopted in a significant fraction of new trials, it would have two observable effects. One effect is to increase long term rates of exit from trials. This is a long-term effect because trials do not change protocol midstream, meaning the change did not affect trials started before 1996. The other effect is to increase the long-term rate of entry into new trials. Subjects would be more willing to enroll in trials if it were possible to drop out when treatment in the trial seemed ineffective.

We test these predictions using the same entry and exit rate regressions we employ to distinguish whether the effect of HAART was driven by changes in supply of subjects or demand by trials. We find moderate support for the proposition that trials adopted the design change. Figure 7 suggests

⁴³ Investigators in trials had information on viral loads both before and after 1996 because viral loads were used as a surrogate endpoint in trials throughout the sample. Moreover, pre-existing trials did not change their protocols lest it modify the treatment effect estimated by data prior to 1996.

that exit rates increased only temporarily after 1996, through our regression analysis in Table 5 suggests the effect may have lasted longer. According to the same figure and Table 6, long term entry rates fell by as much as 10% after HAART. Although this is consistent with a limited adoption of the design change, we do not draw this inference because the decline in long-term entry rates could also be due to a reduction in supply of subjects or demand from trials due to improvements in conventional care.

IV. Concluding remarks and implications for future research

In this paper we argued that health care reforms, by affecting the price and quality of conventional care, have non-standard effects on development mediated by the market for research subjects, and consequently on future cost growth induced by medical innovation. We illustrated such subject market 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 market 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 conducted 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 greater 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 til tomorrow in order to learn about product quality amongst a small number of consumers today. Subject market 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. Subject Compensation

Bioethical rules impose a binding cap on wages for human research subjects. In our framework, this constraint amounts to a restriction that the price charged for experimental therapy in a trial cannot be too negative. There are two mechanisms through which this cap may impact welfare.

First, the constraint 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).

Second, the wage constraint may affect the pace of innovation due to subject market effects. To begin with, relaxing a binding wage constraint must increase innovator profits, *NPV*, for the simple reason that unconstrained profit maximization cannot lower profits. This will increase both the probability that a new drug is developed and the welfare of patients and innovators if the drug is developed. Moreover, the wage constraint may induce investigators to engage in non-price competition for subjects by altering a trial's design features. For example, investigators often lower the probability of receiving the control or employ active rather than placebo controls to raise trial enrollment. Holding constant sample size, these design changes might reduce the quality of information produced by trials. Finally, even without regard to probability of new drug development, the wage constraint may reduce static social welfare. Compensation enables subjects to internalize the positive externality that their consumption provides to future patients. Thus, when a manufacturer funds subject compensation from its future sales, it partly completes the incomplete market between current and future consumers.

D. First Mover Advantage in Medical R&D

Subject market 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 (1993)). It also has important implications for the optimal speed of creative destruction for medical products.

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.

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). 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. In particular, reforms to the quality and price of conventional care will 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.

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. 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.
3. 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.
4. 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.
5. Chandra, Amitabh and Douglas Staiger. 2007. Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. *Journal of Political Economy*. Feb: 103-140.
6. Chassang, Sylvain, Gerard Padró i Miquel and Erik Snowberg. 2010. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." Working Paper.
7. Claxton, K. and J. Posnett. 1996. "An Economic Approach to Clinical Trial Design and Research Priority-Setting." *Health economics*, 5(6), 513-24.
8. 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.
9. Congressional Budget Office. 2006. *Research and Development in the Pharmaceutical Industry*. Pub. No. 2589.
10. Cutler, David. 2004. "Your Money or Your Life: Strong Medicine for America's Health Care System." New York, NY: Oxford University Press.
11. 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.
12. 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.
13. 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.

14. 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.
15. European Federation of Pharmaceutical Industries and Associations. 2011. "The Pharmaceutical Industry in Figures." Available at: <http://www.efpia.org/content/default.asp?PageID=608>.
16. 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).
17. 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/compreg_nov2000_en.pdf.
18. 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.
19. 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.
20. 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.
21. 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.
22. Lakdawalla, D.; N. Sood and D. Goldman. 2006. "HIV Breakthroughs and Risky Sexual Behavior." *Quarterly Journal of Economics*, 121(3), 1063-102.
23. 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.
24. 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.

25. Malani, Anup. 2008. Patient enrollment in medical trials: Selection bias in a randomized experiment. *Journal of Econometrics*. 144: 341-351.
26. 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.
27. Murphy, Kevin, and Robert Topel. 2006. The Value of Health and Longevity. *Journal of Political Economy*, 114(5): 871-904.
28. Neumann, P. J. and E. A. Sandberg. 1998. "Trends in Health Care R&D and Technology Innovation." *Health Aff (Millwood)*, 17(6), 111-9.
29. New York Times. AZT's Inhuman Cost. August 28, 1989.
30. Pharmaceutical Research and Manufacturers of America. 2002. 2002 Industry Profile. PhRMA, Washington, DC.
31. Pharmaceutical Research and Manufacturers of America. 2010. Pharmaceutical Industry Profile 2010. PhRMA, Washington, DC.
32. Philipson, T. 1997. "The Evaluation of New Health Care Technology: The Labor Economics of Statistics." *Journal of Econometrics*, 76(1-2), 375-95.
33. 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.
34. Steinbrook, R. 2001. "Providing Antiretroviral Therapy for Hiv Infection." *N Engl J Med*, 344(11), 844-6.
35. Thiers, F. A.; A. J. Sinsky and E. R. Berndt. 2008. "Trends in the Globalization of Clinical Trials." *Nature Reviews Drug Discovery*, 7(1), 13-14.
36. 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>.
37. 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.
38. 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.
39. 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.

40. Heidi Williams. 2007. "Dynamic effects of medical technologies: Evidence from the prostate cancer market." Working paper.

TABLES

Table 1. Summary statistics.

VARIABLES	Units	Pre-HAART (1990-1995)			Post-HAART (1996-2005)		
		N	mean	sd	N	mean	sd
AIDS drug use							
Primary treatment	0/1	3,848	0.619	0.316	4,498	0.799	0.233
Secondary treatment	0/1	3,848	0.381	0.316	4,498	0.201	0.233
Trial participation							
Any trial	0/1	3,848	0.0774	0.215	4,498	0.0464	0.176
Placebo-controlled	0/1	3,848	0.0513	0.182	4,498	0.00931	0.0733
Employment status							
Full time	0/1	3,763	0.627	0.466	4,410	0.579	0.484
Part time	0/1	3,763	0.0893	0.265	4,410	0.0967	0.279
Income category							
<\$10,000	0/1	3,435	0.138	0.328	4,486	0.106	0.294
\$10,000-\$19,999	0/1	3,435	0.137	0.320	4,486	0.118	0.304
\$20,000-\$29,999	0/1	3,435	0.158	0.339	4,486	0.114	0.296
\$30,000-\$39,999	0/1	3,435	0.156	0.340	4,486	0.125	0.309
\$40,000-\$49,999	0/1	3,435	0.118	0.300	4,486	0.118	0.3
\$50,000-\$59,999	0/1	3,435	0.220	0.396	4,486	0.253	0.42
\$60,000-\$69,999	0/1	3,435	0.00638	0.0751	4,486	0.121	0.319
>\$70,000	0/1	3,435	0.0124	0.102	4,486	0.000111	0.00747
Refused to answer	0/1	3,435	0.0545	0.204	4,486	0.0458	0.194
Insurance status							
Government	0/1	3,848	0.141	0.334	4,498	0.229	0.406
Medicare	0/1	3,435	0.0503	0.209	4,470	0.163	0.358
Medicaid	0/1	3,436	0.0974	0.286	4,470	0.105	0.294
Vet. Admin.	0/1	3,436	0.0394	0.185	4,470	0.0272	0.155
Private	0/1	3,848	0.473	0.476	4,498	0.529	0.48
Individual plan	0/1	3,068	0.131	0.319	4,470	0.084	0.254
Group plan	0/1	3,069	0.504	0.481	4,470	0.457	0.474
Uninsured	0/1	3,848	0.418	0.468	4,498	0.292	0.431
Other							
Age	yrs	3,783	40.82	7.164	4,401	47.4	7.052
CD4 count	cells/mm3	3,819	290.6	240.2	4,470	494	276.6
Viral load	copies/ml	1,771	140,305	242,912	4,389	22,465	117,123

Table 2. Effect of HAART on trial participation rate: identification off time-trend.

	(1)	(2)	(3)	(4)
Panel A				
Pre (1995)	0.166*** (0.009)	0.184*** (0.025)	0.164*** (0.026)	4.016*** (0.473)
Post (1996)	0.130*** (0.010)	0.118*** (0.020)	0.104*** (0.021)	2.852*** (0.342)
Wald test (p-value)	0.005	0.019	0.033	0.021
Panel B				
Pre trend (1992-94)	0.009** (0.004)	0.017* (0.010)	0.014 (0.011)	1.124** (0.058)
Post trend (1997-99)	-0.050*** (0.003)	-0.051*** (0.007)	-0.046*** (0.008)	0.612*** (0.031)
Wald test (p-value)	0.000	0.000	0.000	0.000
Panel C				
Pre trend (1990-95)	0.016*** (0.002)	0.027*** (0.004)	0.027*** (0.004)	1.231*** (0.029)
Post-trend (1996-2005)	-0.031*** (0.002)	-0.030*** (0.005)	-0.033*** (0.006)	0.613*** (0.027)
Wald test (p-value)	0.000	0.000	0.000	0.000
Obs.	7,961	7,961	7,443	3,094
Identification	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	NO	Yes	Yes	Yes
Add'l covariates	NO	No	Yes	Yes
Distribution	LPM	LPM	LPM	Logit
Sample	Primary drug users	Primary drug users	Primary drug users	Primary drug users

Note: The dependent variable is whether a subject participated in a trial. Panel A reports results of regressions of the dependent variable on an indicator variable for the year just before (1995) and after (1996) HAART. Panel B replaces the one year pre and post year indicators with trends in longer pre (1992-1994) and post (1997-1999) windows. Panel C examines trends in even longer pre (1990-1995) and post (1996-2005) windows. Specifications (3) and (4) include controls for income and CD4 count. Specification (4) reports coefficients as odds ratios. Observations are at the subject x year level. They are weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Standard errors are reported under coefficient estimates. Standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between pre and post coefficients in each panel.

Table 3. Effect of HAART on trial participation rate: identification off secondary drug controls.

	(1)	(2)	(3)	(4)
Panel A				
Pre (1995)	0.141***	0.156***	0.173***	5.163***
x primary drug user	(0.012)	(0.037)	(0.032)	(0.797)
Post (1996)	0.177***	0.160***	0.156***	6.421***
x primary drug user	(0.013)	(0.029)	(0.029)	(1.344)
Wald test (p-value)	0.044	0.943	0.672	0.398
Panel B				
Pre trend (1992-94)	0.011**	0.013	0.013	1.122
x primary drug user	(0.005)	(0.012)	(0.012)	(0.091)
Post trend (1997-99)	-0.037***	-0.038***	-0.041***	0.677***
x primary drug user	(0.004)	(0.009)	(0.009)	(0.050)
Wald test (p-value)	0.000	0.000	0.000	0.000
Panel C				
Pre trend (1990-95)	0.016***	0.020***	0.022***	1.224***
x primary drug user	(0.002)	(0.005)	(0.005)	(0.039)
Post-trend (1996-2005)	-0.003**	-0.003	-0.002	1.059*
x primary drug user	(0.001)	(0.002)	(0.003)	(0.035)
Wald test (p-value)	0.000	0.000	0.000	0.000
Obs.	12,982	12,982	12,146	6,057
Identification	DD	DD	DD	DD
Individual FE	NO	Yes	Yes	Yes
Add'l covariates	NO	No	Yes	Yes
Distribution	LPM	LPM	LPM	Logit
Sample	Primary drug users	Primary drug users	Primary and secondary drug users	Primary and secondary drug users

Note: See Table 2 for a description of the dependent variable and the treatment variable definitions across Panels A – C and of control variables and coefficient reporting in Specifications (3) and (4). Observations are at the subject x drug type x year level and weighted to so each individual x drug type paid has the same weight. The sample includes subjects on primary or secondary AIDS drugs. The coefficients identify changes in the dependent variable among primary drug users relative to secondary drug users, so identification uses a difference in difference (DD) design. Standard errors are reported under coefficient estimates. In all columns, standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between pre and post coefficients in each panel.

Table 4. Effect of HAART on trial participation rate: analysis of subsamples that are more robust to attrition.

	(3)	(4)	(3)	(4)	(3)	(4)
Panel A						
Pre (1995)	0.190*** (0.036)	6.526*** (1.344)	0.197*** (0.026)	6.160*** (1.084)	0.202*** (0.034)	5.823*** (1.300)
Post (1996)	0.106*** (0.027)	3.497*** (0.644)	0.120*** (0.022)	3.451*** (0.509)	0.101*** (0.025)	3.120*** (0.616)
Wald test (p-value)	0.010	0.009	0.005	0.004	0.004	0.016
Panel B						
Pre trend (1992-94)	-0.013 (0.014)	0.893 (0.096)	-0.005 (0.010)	0.948 (0.090)	-0.019 (0.013)	0.805* (0.096)
Post trend (1997-99)	-0.040*** (0.010)	0.611*** (0.047)	-0.043*** (0.009)	0.611*** (0.039)	-0.033*** (0.010)	0.678*** (0.056)
Wald test (p-value)	0.094	0.006	0.005	0.000	0.373	0.263
Panel C						
Pre trend (1990-95)	0.022*** (0.005)	1.223*** (0.047)	0.027*** (0.005)	1.259*** (0.043)	0.025*** (0.006)	1.212*** (0.052)
Post-trend (1996-2005)	-0.029*** (0.008)	0.585*** (0.040)	-0.036*** (0.006)	0.573*** (0.032)	-0.034*** (0.007)	0.592*** (0.041)
Wald test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	3,773	1,451	5,441	2,427	3,805	1,805
Identification	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Add'l covariates	Yes	Yes	Yes	Yes	Yes	Yes
Distribution	LPM	Logit	LPM	Logit	LPM	Logit
Sample	CD4 always > 200	CD4 always > 200	In sample as of 2000	In sample as of 2000	In sample as of 2005	In sample as of 2005

Note: See Table 2 for a description of the dependent variable and the treatment variable across Panels A – C and of control variables and coefficient reporting in Specifications (3) and (4). In the first two columns, subjects are included in the sample only if their CD4 count never falls below 200. In the middle two columns, observations are kept only if the subject is still in the sample in 2000. In the last two columns, observations are kept only if the subject is still in the sample in 2005. All observations are at the subject x year level and weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year in the first four columns. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Each regression contains a linear trend, so identification is off a linear trend. Standard errors are reported under coefficient estimates. In all columns, standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1. We report the p-value for a Wald test of the difference between pre and post coefficients in each panel.

Table 5. Effect of HAART on supply: analysis of exit rates.

	(1)	(2)	(3)	(4)
Years 1992-94	-0.089** (0.035)	0.051 (0.089)	-0.215 (0.190)	0.253* (0.199)
Year 1995	-0.054 (0.051)	0.086 (0.129)	-0.258 (0.176)	0.248** (0.170)
Year 1996	0.088* (0.048)	0.324*** (0.108)	-0.051 (0.148)	0.653 (0.393)
Year 1997-99	0.023 (0.040)	0.481*** (0.101)	0.045 (0.117)	1.344 (0.504)
Constant	0.585*** (0.027)	0.315*** (0.071)	-1.261 (1.031)	
Wald tests				
1996 v. 1995	0.015	0.029	0.062	0.011
1996 v. 1992-94	0.000	0.002	0.156	0.030
1997-99 v. 1995	0.002	0.000	0.045	0.004
Obs.	760	760	753	466
Identification	exit	exit	exit	exit
Individual FE	NO	Yes	Yes	Yes
Add'l covariates	NO	No	Yes	Yes
Distribution	LPM	LPM	LPM	Logit
Sample	Primary drug users	Primary drug users	Primary drug users	Primary drug users

Note: The dependent variable is whether the subject exited a trial. See Table 2 for a description of the treatment variable across Panels A – C and of control variables and coefficient reporting in Specifications (1) - (4). Observations are at the subject x year level. They are weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year and were enrolled in a clinical trial the previous period. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Standard errors are reported under coefficient estimates. Standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between coefficients on the indicated variables.

Table 6. Effect of HAART on supply and demand: analysis of entry rates.

	(1)	(2)	(3)	(4)
1992-94	0.018*** (0.005)	-0.019 (0.019)	0.029* (0.016)	0.959 (0.493)
1995	0.032*** (0.008)	(dropped)	0.058** (0.025)	1.971 (0.880)
1996	0.047*** (0.008)	0.019 (0.022)	0.078*** (0.025)	2.944*** (1.150)
1997-99	-0.001 (0.005)	-0.043** (0.017)	0.019 (0.022)	0.817 (0.226)
Constant	0.034*** (0.004)	0.076*** (0.016)	0.160** (0.080)	
Wald tests				
1996 v. 1995	0.145	0.383	0.369	0.118
1996 v. 1992-94	0.000	0.042	0.015	0.000
1997-99 v. 1995	0.000	0.031	0.526	0.643
Obs.	6,586	6,586	6,477	1,424
Identification	enter	enter	enter	enter
Individual FE	NO	Yes	Yes	Yes
Add'l covariates	NO	No	Yes	Yes
Distribution	LPM	LPM	LPM	Logit
Sample	Primary drug users	Primary drug users	Primary drug users	Primary drug users

Note: The dependent variable is whether the subject entered a trial. See Table 2 for a description of the treatment variable across Panels A – C and of control variables and coefficient reporting in Specifications (1) - (4). Observations are at the subject x year level. They are weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year and were not enrolled in a clinical trial the previous period. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Standard errors are reported under coefficient estimates. Standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between coefficients on the indicated variables.

Table 7. Effect of HAART: separating price and quality effects.

	(1)	(2)	(3)	(4)
Panel A				
Post (1996)	-0.059*** (0.016)	(dropped)	-0.109*** (0.035)	0.488*** (0.089)
Post (1996) x uninsured	0.039** (0.020)	0.028 (0.043)	0.034 (0.043)	1.252 (0.313)
Constant	0.295*** (0.011)	0.250*** (0.024)	0.461* (0.270)	
Panel B				
Post trend (1997-99)	-0.043*** (0.004)	-0.045*** (0.009)	-0.041*** (0.010)	0.680*** (0.043)
Post trend (1997-99) x uninsured	-0.017*** (0.006)	-0.005 (0.013)	-0.006 (0.013)	0.832* (0.079)
Constant	0.135*** (0.008)	0.176*** (0.019)	0.227 (0.244)	
Panel C				
Post-trend (1996-2005)	-0.031*** (0.002)	-0.030*** (0.005)	-0.034*** (0.006)	0.600*** (0.027)
Post-trend (1996-2005) x uninsured	-0.002 (0.002)	0.001 (0.004)	0.001 (0.004)	1.069** (0.036)
Constant	0.062*** (0.006)	0.074*** (0.018)	0.174 (0.271)	
Obs.	7,961	7,961	7,443	3,094
Identification	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	NO	Yes	Yes	Yes
Add'l covariates	NO	No	Yes	Yes
Distribution	LPM	LPM	LPM	Logit
Sample	Primary drug users	Primary drug users	Primary drug users	Primary drug users

Note: See Table 2 for a description of the dependent variable and treatment variable across Panels A – C and of control variables and coefficient reporting in Specifications (1) - (4). Observations are at the subject x year level. They are weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Standard errors are reported under coefficient estimates. Standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between pre and post coefficients in each panel.

Table 8. Effect of HAART: conditioning on trial design.

	(3)	(4)	(3)	(4)
Panel A				
Post (1996)	-0.080** (0.033)	0.567*** (0.108)	-0.013 (0.020)	0.928 (0.266)
Post (1996) x uninsured	0.032 (0.042)	1.207 (0.319)	0.000 (0.027)	1.202 (0.495)
Constant	0.430** (0.201)		-0.075 (0.161)	
Panel B				
Post trend (1997-99)	-0.039*** (0.009)	0.657*** (0.047)	-0.002 (0.007)	0.697*** (0.077)
Post trend (1997-99) x uninsured	-0.005 (0.013)	0.921 (0.102)	-0.008 (0.008)	0.658** (0.129)
Constant	0.236 (0.155)		-0.126 (0.159)	
Panel C				
Post-trend (1996-2005)	-0.048*** (0.006)	0.506*** (0.029)	0.005 (0.004)	0.653*** (0.061)
Post-trend (1996-2005) x uninsured	0.004 (0.004)	1.112*** (0.041)	-0.001 (0.002)	1.040 (0.096)
Constant	0.084 (0.196)		-0.122 (0.167)	
Obs.	6,521	2,414	7,443	1,441
Identification	Linear trend	Linear trend	Linear trend	Linear trend
Individual FE	Yes	Yes	Yes	Yes
Add'l covariates	Yes	Yes	Yes	Yes
Distribution	LPM	Logit	LPM	Logit
Sample	Primary drug users, excluding subjects in placebo-controlled trials	Primary drug users, excluding subjects in placebo-controlled trials	Primary drug users, excluding subjects in active-controlled trials	Primary drug users, excluding subjects in active-controlled trials

Note: See Table 2 for a description of the dependent variable and treatment variable across Panels A – C and of control variables and coefficient reporting in Specifications (3) and (4). Observations are at the subject x year level. They are weighted so each subject has the same weight. The sample includes only subjects who were on primary AIDS drugs each year. The coefficients identify changes in the dependent variable among primary drug users relative to a linear trend. Standard errors are reported under coefficient estimates. Standard errors are clustered at the subject-level. ***/**/* indicate $p < 0.01/0.05/0.1$. We report the p-value for a Wald test of the difference between pre and post coefficients in each panel.

FIGURES

Participation rate

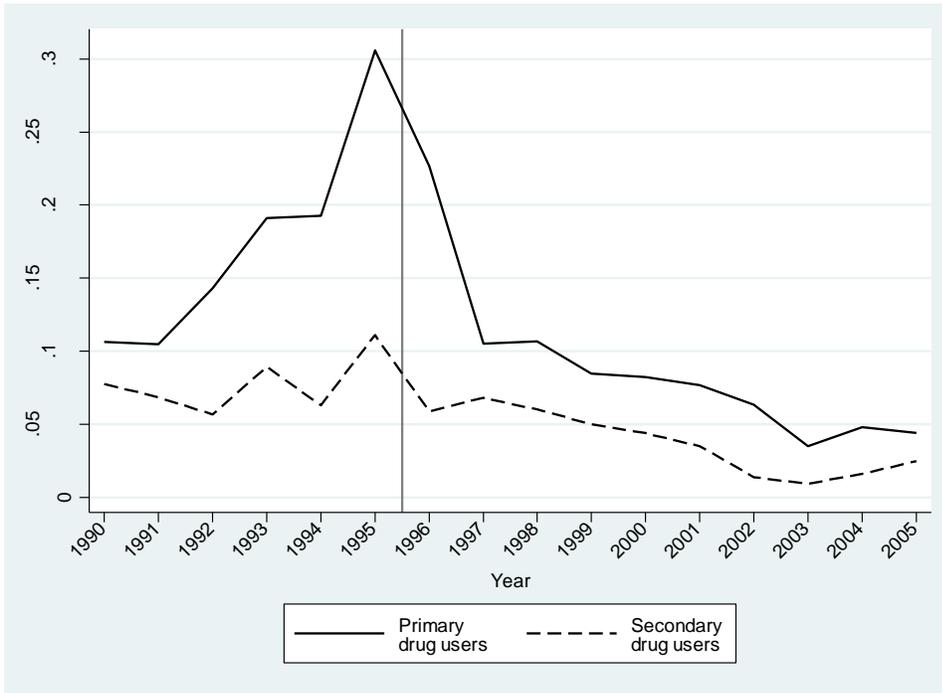


Figure 1. Effect of HAART on trial participation.

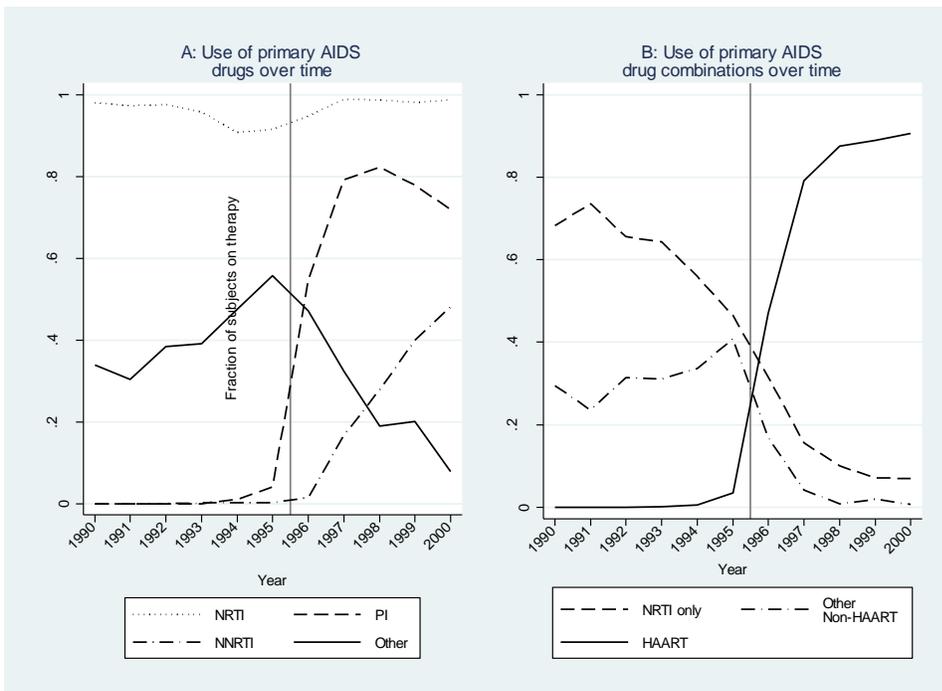


Figure 2. Effect of HAART on choice of AIDS therapy.

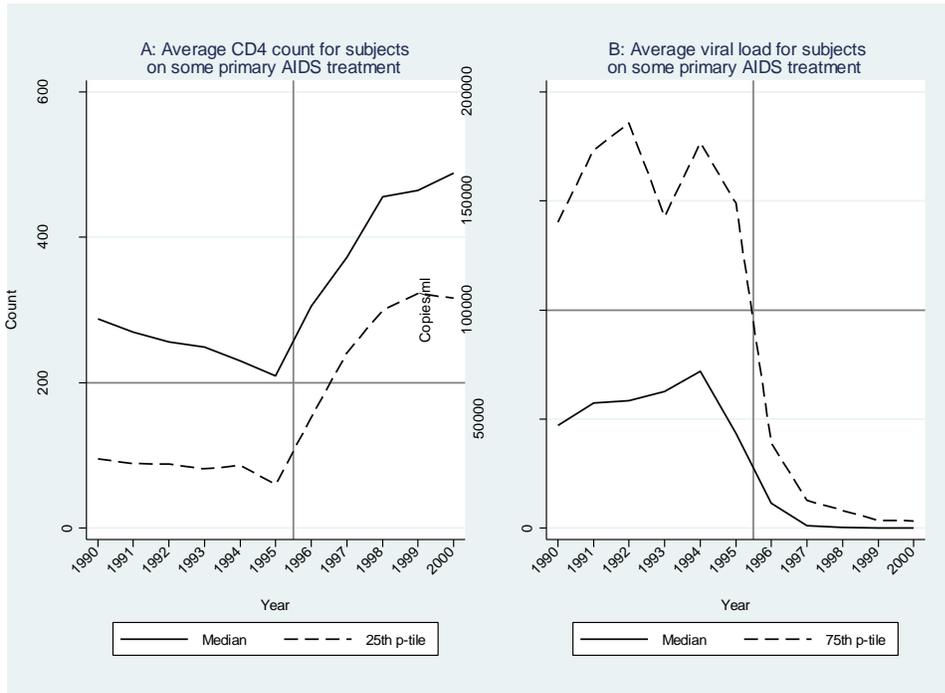


Figure 3. Effect of HAART on CD4 counts.

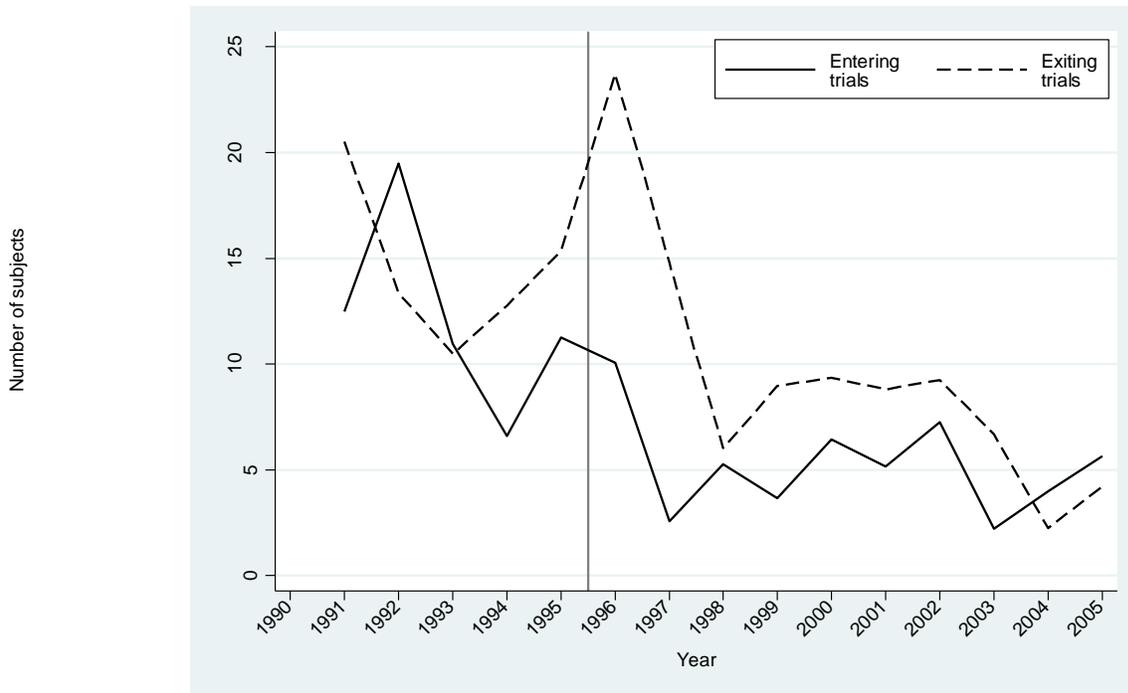


Figure 4. Effect of HAART on trial entry and exit.

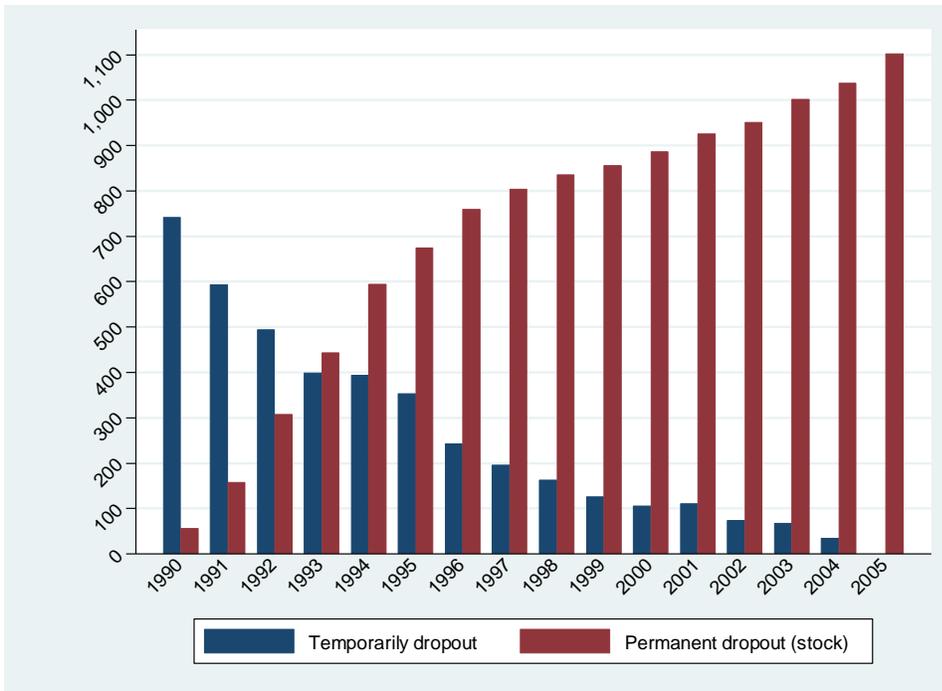


Figure 5. Temporary and permanent attrition over time.

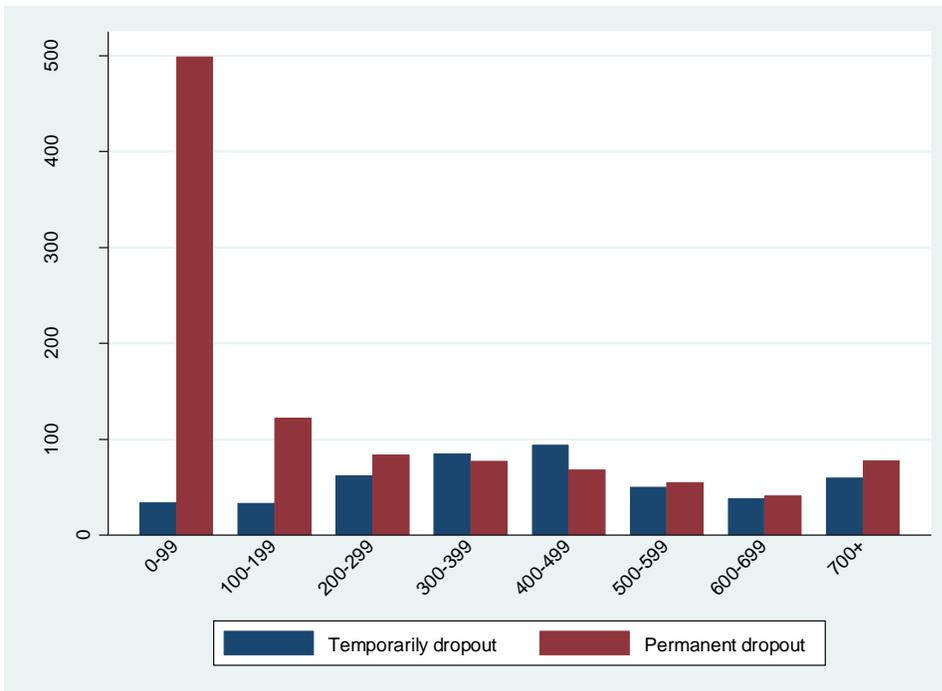


Figure 6. Temporary and permanent attrition, by CD4 count.

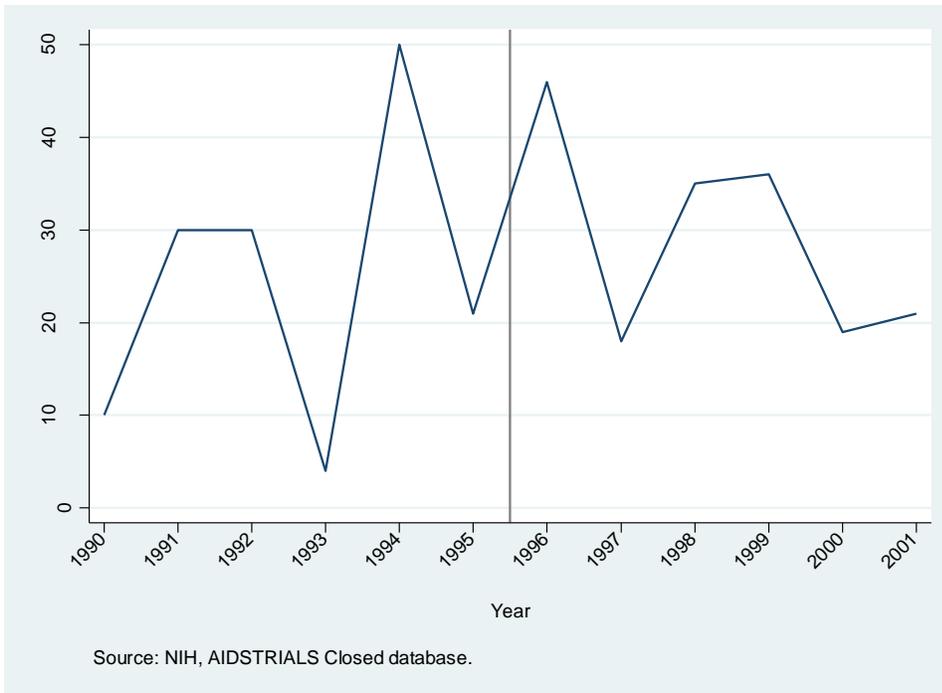


Figure 7. Termination of ACTG trials over time.

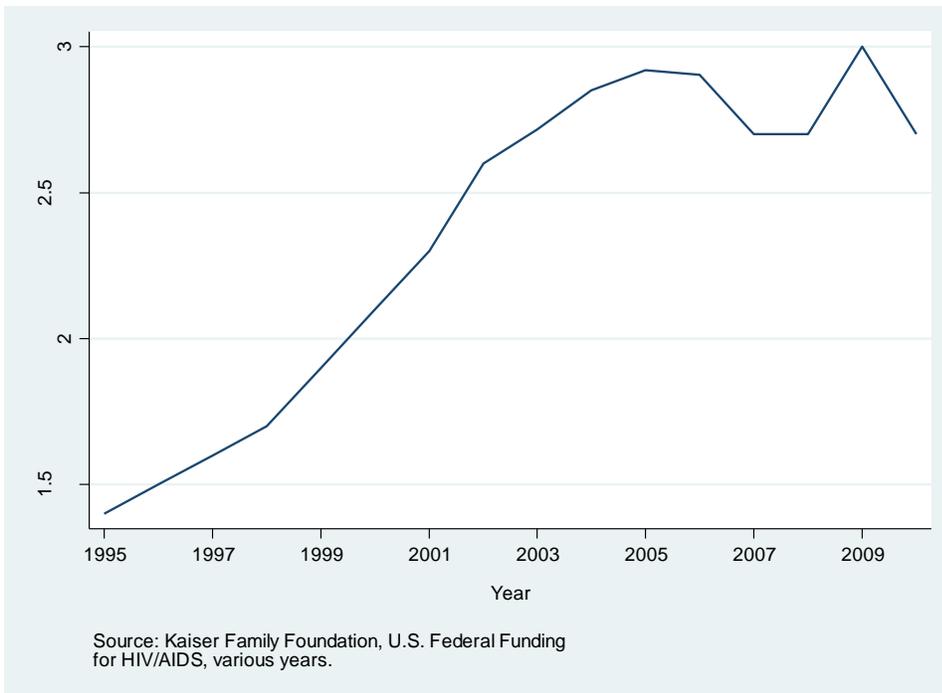


Figure 8. Federal government (NIH) spending on HIV/AIDS research.