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, Eric Helland, William Hubbard, Darius Lakdawalla, Richard Miller, Kevin Murphy, Seth Seabury, Tony Tse; workshop and conference participants at the University of Chicago, Harvard Law School, 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 and Philipson acknowledges support from the George Stigler Center for the Study of the Economy and the State and the Biotechnology Industry Organization (BIO). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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
JEL No. I1,I11

ABSTRACT

Improvements in health have been a major contributor to gains in overall economic welfare. In this paper, we argue that previous economic research on R&D has overlooked an important difference between medical R&D and R&D in other sectors. The health care sector exhibits a unique linkage between product development and output markets. Participants in clinical trials for new medical products are also potential consumers of existing approved medical products. This overlap between input supply and output demand has non-standard effects on innovative returns over time and across geography. First, medical R&D has a self-limiting effect. Contemporary innovation discourages trial participation and slows down development necessary for future innovation. Thus, medical R&D suffers increasing costs over time, driven by improvements in the standard of care. Second, policies that affect output markets, such as universal coverage and price controls, affect the returns to innovation, not only by altering the firm's variable profits, but also by increasing the length and cost of development. Third, the amount of medical R&D in a location is driven, not only by the local relative R&D talent, but also by consumer demographics and output market policies in that location. We provide evidence of the input-output linkage for the break-through HIV therapies introduced in 1996. We document the substantial drop in trial recruitment induced by these new innovations and argue that this has slowed down development and lowered returns to subsequent HIV-related innovations.

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

I. Introduction

Improvements in health have been a major contributor to gains in overall economic welfare during the last century (Murphy & Topel 2006) as well to the reduction in global inequality (Becker et al 2005). In the U.S. alone, these gains have been valued at half of GDP. A part of these gains is due to medical research and development (R&D), including improvements in medical information, procedures, drugs, biologics, and devices (e.g., Cutler & Kadiyala 2003, Lichtenberg 2003). Because of the substantial value of health improvements, the pace of medical innovation is an important concern.

Unfortunately, there is mounting evidence that the returns to medical R&D are beginning to diminish. Although the amount spent on drug R&D has doubled since 1993, the number of drugs for which the pharmaceutical industry has sought or obtained marketing approval has remained flat (GAO 2006).[2] A central component of the costs of medical R&D is clinical trials, which are required for regulatory approval to market new medical products. These trials are taking longer and becoming more costly. Recent evidence suggests that such clinical testing now accounts for over one-half of development costs ($467 of $802 million) and nearly 80% of the average 6 yearsit takes to obtain marketing approval (DiMasi et al. 2003, Adams & Brantner 2006, McKinsey and Co. 2008).[3]

Despite the importance of health to overall well-being and the slow-down of medical progress, economists have paid little attention to whether standard economic analysis of R&D applies to medical products and whether the same policy measures that have been used to stimulate other R&D activities can also promote medical progress in an efficient manner.

In this paper, we argue that there is an important way in which the process of medical R&D differs from R&D in other sectors and, as a result, the traditional positive and normative analysis of R&D in other sectors may not necessarily apply. Under the traditional economic view, the benefits of R&D are driven by future variable profits in the output market. The costs of R&D are driven by input supply, for example, the quantity and location of research talent. In the health care sector, however, both the benefits and costs of R&D are driven by output markets. The reason is that medical R&D requires

[2] The regulatory approval process varies across countries but is generally similar to that of the US, where a company must first find a chemical entity and show medical value in *in vitro* and animal testing. Thereafter, it can file an Investigational New Drug (IND) application with the Food & Drug Administration (FDA) and begin the requirement of three phases of clinical testing in humans. Phase I includes small studies of the toxicity of the drug. Phase II includes medium-sized studies of dosage and drug efficacy. Phase III, which is, by far, the most costly, includes large-scale randomized clinical trials of efficacy and safety. Upon completion, a New Drug Application (NDA) may be filed with FDA. Finally, the FDA reviews the data in the NDA to determine whether the drug is safe and effective to determine marketing approval. After approval, the FDA may require further clinical testing through, so called, "phase IV" studies.

[3] Thetime it takes to conduct Phase II and III clinical trials is much longer than the average year and a half it takes for the Food and Drug Administration (FDA) to review the evidence of the trials before deciding on marketing approval (GAO 2006).

clinical trials on human subjects.[4] Yet the same individuals who can serve as human subjects are also potential consumers of medical products as patients in the output market.[5] In otherwords, input supply for medical development overlaps with output demand for medical products. As a result, improvements in output markets, e.g., an increase in quality or reduction in price of conventional medical care, makes patients more reluctant to enroll as subjects in clinical trials. This, in turn, prolongs development timeslines for medical products.

We analyze the implications of this link between development and output markets for the rate of medical progress over time, the location of medical R&D, and the value of conventional policies aimed at stimulating R&D progress. First, we consider the return to R&D over time. As the conventional treatment become more effective and less expensive, it becomes more difficult to recruit patients into clinical trials. This slows down development and, therefore, the arrival of future profits for innovators. Because medical products are patented before they are tested, this slowdown in development also eats up the remaining patent life on the market-approvedproduct. In short, better conventional care implies both a delayed start and a shorter duration of on-patent profits for medical products. The consequence is that medical R&D has a "self-limiting" effect. This may partly explain the growth in development times and slowdown in medical R&D productivity in recent years.

Second, the link between development and output markets affects the location of medical R&D across the globe. In contrast to other products, there is an important relationship between the location of output markets and the location of R&D spending for medical products. Standard economic arguments predict no reason why the R&D, performed for world markets, should be tied to the output market or policies of the region it is conducted in. Yet, for medical R&D, the same forces that determine profitability in output markets in a country may also drive development costs in that country. For example, the high prevalence of obesity in the US enhances both profits and the speed of development for obesity-related products. Moreover, medical R&D may not take place in countries with universal insurance or price controls because local development has to compete with local availability of cheap conventional care. This offers a potential explanation for the large shift in R&D from Europe to the USin the last two decades. It may also explain why manufacturers have begun to outsource clinical trials to poorer countries, where the price of conventional care is prohibitively expensive and experimental treatment is often the best feasible care.

Third, the development-output link has non-standard implications for R&D policy. Our analysis implies that demand subsidies, such as Medicaid and Medicare in the US, may have different effects on

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

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

innovation than conjectured. Conventional wisdom suggests that subsidies affect the benefits of innovation by altering markups or quantity in output markets. However, we stress that they may also increase the costs of development because they make trial recruitment more difficult they lower the prices (co-pay) patients pay for conventional care. In addition, our analysis suggests an alternative method for stimulating medical R&D: allowing compensation for research subjects. The market for clinical observations is implicitly a labor market in which there are maximum wage policies imposed by regulations enforced by ethical bodies such as institutional review boards (IRBs) in the US. These maximum wage policies are harmful for two reasons. First, because subjects confer positive external effects by generating public knowledge about new medications, they should be given Pigouvian subsidies rather than face wage caps. Second, without a flexible price mechanism, the demand for human subjects is rationed by queuing rather than price, which increases development times and delays innovation.

To test the empirical relevance of output markets to development, we examine the pace of trial recruitment before and after the 1996 introduction of a breakthrough innovation for treating HIV/AIDS: highly active antiretroviral therapy (HAART). Using longitudinal data on over 1000 men in 4 U.S. cities, we find that the improvement in the quality of conventional care after this breakthrough was approved dramatically reduced participation in trials for new experimental HIV/AIDS treatments. To control for unobserved trends in trial participation we use a time trend or a control group of drug users that did not experience a change in conventional treatment in 1996. We find that the trial participation rate fell 5 – 10% in absolute terms (25-75% in relative terms) after HAART was approved. 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. This inference is supported by the presence of wage caps on research subjects. These caps imply excess demand for research subjects so that quantity changes are determined by changes in the lower level of supply. These results suggest that improvement in the quality of conventional HIV/AIDS care due to HAART substantially increased the development times for future HIV/AIDS-related therapies.

The paper may briefly be outlined as follows. Section II presents a simple model of the impact of output markets on the timing and cost of developing new medical products. Section III discusses how these output factors affect the returns to innovation in a non-standard manner. Section IV discusses the equilibrium impact of output markets on development when development times, rather than price, is used to ration demand for subjects. Section V presents evidence suggesting that the new AIDS treatments introduced in the mid-1990s significantly reduced trial participations and slowed down development. Section VI concludes with a discussion of the normative implications of our analysis.

II. Economic determinants of development times

We first examine 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.

A. ***The supply and demand of subjects to clinical trials***

The recruitment of trial participants comes from a stock (prevalence) of a disease, denoted z_t, in a given period. 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:

$$z_{t+1} = b + (1 - m)z_t$$

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

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. Assume that conventional treatment (or the standard of care) offers per-period utility, $U(q, p)$, that increases in the quality of care, q (synonomous with the treatment effect, health outcome, or "effectiveness" of care), and decreases in the price, p (i.e., uninsured price or co-pay for insured patients). If a patient enrolls in a clinical trial of experimental therapy, he obtains an uncertain utility, $U(q_E, p_E)$. Here, the quality of experimental care, q_E , is a random variable unknown at the time of entering the trial and p_E is the price of experimental care, potentially zero if treatment is subsidized by the trial. (Later, we will discuss the impact of regulations banning compensation of subjects, i.e., a negative experimental price.) The uncertain quality of experimental care may be the product of experimental design, e.g., the random assignment of treatments, and/or uncertainty about the effects of the experimental therapy being studied.

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

$$E[U(q_E, p_E)] \geq U(q, p)$$

where expectation is over the distribution,$F(q_E)$, describing a patient's beliefs about the quality of the experimental therapy. We assume there is heterogeneity among patients, e.g., different beliefs about the quality of experimental care or degrees of risk-aversion, that gives rise to a differentiable supply function, $s(p, q)$, that is increasing in the price of conventional care and decreasing in its quality: $s_p \geq 0, s_q \leq 0$. This function depicts the fraction of the stock of patients willing to participate in the trial, given existing output market conditions.[6]

For a simple example that illustratessuch output market effects, suppose beliefs about experimental quality, $F(q_E; \theta)$, differ across patients according to the parameter θ, which is distributed according to the cdf, $G(\theta)$. Suppose further that beliefs are increasing in a first-order stochastic

[6]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.

dominance sense in this parameter and denote, by $\theta(q,p)$, the level of the parameter of the patient who is indifferent between participating or not. Then, the supply into the trial is given by those who, relative to their peers, have more optimistic beliefs about the experimental care: $s(q,p) = 1 - G(\theta(q,p))$. An alternative source of heterogeneity among patients may be their level of risk-aversion, where analogous arguments would imply that those who are least averse to the risks of experimental care are the ones who participate in trials.

B. ***Development times and output markets***

Given a stock of patients with a disease, we assume clinical trial participants are split evenly across N existing trials so that the flow of subjects, f, into a particular trial is given by

$$f = \frac{zs}{N}$$

Thus, a smaller number of competing trials, larger prevalence and higher participation rates would each increase recruitment per period into a trial.[7]

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 a stock of n patients when a flow of patients, f , patients enroll each period:

$$Tf = n \rightarrow T = \frac{nN}{zs} = \left(\frac{m}{b}\right)\left(\frac{nN}{s(q,p)}\right)$$

This relationship implies a mechanical relationship between various factors and development times. Naturally, increases in sample sizes or the number of competing recruiting trials increase development times, as more time is needed to attain the desired number of patients. 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 rise with the mortality rate and fall with the incidence of the disease. 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, so called "me-too" innovations, for which experimental care is a close substitute to conventional care, may be interpreted as experimental treatment with quality similar to conventional care. Therefore, these me-too innovations face recruitment difficulties and longer development times because they do not offer substantial gains over the conventional treatment outside of the trial.[8]

[7]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.

[8]Breakthrough therapies involve shorter development not only because they experience easier recruitment per period, but also from the common practice of stopping successful trials due to ethical concerns over withholding beneficial therapy from control groups.

1. Direct effects of output markets on development times

In addition to these mechanical effects, there are economic incentives that affect development times as well. An important driver of the clinical trial participation rate is the quality of conventional care. Improving the treatment effect from the standard of care in output markets lengthens the time required to recruit subjects into trials for new innovations:

$$T_q = T_s s_q > 0$$

This linkage between the quality of care in output markets and the length of development times implies a self-limiting effect on R&D. Prior innovations lengthen development times for new innovations and diminish the returns to further R&D.

The output-input linkage also operates through the price in output markets. Lower prices of conventional care decrease patient incentives to participate in clinical trials and, therefore, increase development times:

$$T_p = T_s s_p < 0$$

There are three important contexts in which price mediates development times: competition in output markets, health insurance in output markets, and purchasing power across countries.

First, sequential innovation will not only raise quality but also reduce prices, thereby raising development times. In this manner, price competition in output markets is an important channel through which R&D has self-limiting effects on development. Second, because the out of pocket cost of conventional care may differ according to insurance coverage, there may be different rates of trials participation between those with better or worse insurance. Although both the insured and uninsured may face the same absolute price for experimental care (likely zero), the *relative price*of experimental care compared to conventional care is lower for the uninsured, holding other factors constant. Therefore, the higher price of conventional care for the uninsured may drive them into trials at higher rates. Lastly, price may also induce international differences in trial participation and recruitment. If conventional care is prohibitively expensive in poor nations, development will be faster. This may explain the out-sourcing of trials to countries without universal coverage or price controls.

If products are partially personally financed, then the total effect of new innovations will come from both the quality- and price effects. If prices rise with quality through a function, $p(q)$, then a new innovation affects development according to

$$\frac{dT}{dq} = T_s[s_q + s_p p_q]$$

Better conventional care has a direct negative effect by lowering participation and raising development times, but an offsetting secondary positive effect on participation because higher quality conventional care has higher prices.

2. Indirect effects of output markets on development

In addition to the direct effects of quality and price on development times through patients' willingness to participate in trials, they also have indirect effects on development times through disease prevalence and trial design.

a. The indirect link through prevalence

We previously considered the negative mechanical relationship between the prevalence of disease and development times. The steady-state prevalence of a disease increases in disease incidence,b, and decreases with the exit rate, m. For some diseases, however, the recovery or mortality 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 affects the exit rate.[9]

If conventional treatment can cure a disease, it will lower the prevalence of that disease. Conversely, if it lowers mortality of a disease without curing it, the prevalence of that disease would increase.[10] Generally speaking, there are two effects of the quality of care in output markets on development times:

$$\frac{dT}{dq} = T_s s_q + T_m m_q$$

The first term captures the participation margin, holding prevalence constant (discussed in the previous subsection). The second term captures the eligibility margin, which considers how prevalent the disease is and, thus, how many eligible patients there are.

For curative treatments, the eligibility effect of better conventional treatment supplements the participation effect, implying even larger development costs if conventional care improves. For treatments that only reduce mortality rates, eligibility effects offset participation effects. For a given prevalence rate, higher quality of conventional care makes recruitment harder, but increased quality of convention care raises prevalence by reducing mortality. In particular, when development time is $T = (m/b)(n/as)$, it follows that the relative magnitude of the elasticities of mortality with respect to quality of conventional care versus the elasticity of participation determines whether development time increases or decreases:

---

[9]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.

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

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

$$\frac{dT}{dq} > 0 \Leftrightarrow \left|\frac{s_q q}{s}\right| > \left|\frac{m_q q}{m}\right|$$

The price of conventional care also has two analogous effects on development times due to its effect on prevalence

$$\frac{dT}{dp} = T_s s_p + T_m m_p$$

Here, an increase in price of a non-curative conventional treatment has offsetting effects on development times as well; 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.

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 mortality (flow out of the disease stock) (e.g., Lakdawalla, Goldman, Sood 2006). This is particularly relevant for understanding how diagnostic technologies affect trial participation and development times. In general, two factors determine measured incidence. The first is the actual incidence or exposure to disease and the second is the diagnostic technology that identifies new patients with the disease. While improved diagnostic capabilities do not increase the actual stock of disease, they are central to recruitment and development times because only diagnosed patients can enter trials. Thus, improved diagnostic technology may speed up development if it increases the diagnosed incidence holding constant the true underlying incidence.

b. The indirect link through trial designs

Another indirect channel through which output markets may affect trial participation and development times is by forcing those running trials to change their design. For example, in order to compete with conventional care, researchers modify the design of trials to include active controls or an increased share of patients receiving experimental care. In essence, this is a form of non-price competition by those conducting experiments in response to improved output market conditions.[11]

Consider the design parameters, represented by the vector w, that affect the value of trial participation. Implicitly, this is a non-pecuniary "wage" for trial participants. One of the design parameters may be the probability of being assigned to the treatment group, w_1. A second parameter,w_2, may be the quality of the control treatment, say placebo or active control.[12] For these two design features, the expected utility of trial participation is given by

[11] In the conclusion, we discuss the impact of allowing for direct compensation of subjects, behavior that is currently banned in many countries.

[12]Active controls are only of value for those who respond well to conventional care. For non-responders, active controls serve no purpose and simply raises trial costs.

$$w_1 EU(p_E, q_E) + (1 - w_1)U(p_E, w_2)$$

It is straightforward to extend our illustrative example to this context. If $\theta(w, p, q)$denotes the reservation level of beliefs over experimental quality, where $w = (w_1, w_2)$, then the participation rate is determined, as suggested before, by

$$s(w, p, q) = 1 - G(\theta(w, p, q))$$

This implies that compensation through more beneficial designs increases participation because s is increasing in w.

When a new innovation improves the quality of conventional care, this induces changes not only in output markets but, potentially, also in experimental designs. If $w(q, p)$ generally denotes the design under a given set of output market conditions, the total impact of an increase in quality on participation is

$$\frac{ds}{dq} = s_q + s_w w_q$$

An improvement in the quality of conventional care now has both a direct negative impact on participation and an offsetting indirect positive effect from inducing a more favorable trial design.

Given the response in trial designs to output markets, one might reason that medical innovation is not self-limiting because trials can always use the current standard of care as the control treatment. However, this argument is fallacious. 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 ($w_2 = q$). In this case, receiving the control treatment is equivalent to not participating so that subjects, by going into the trial, risk substituting conventional care for experimental care, when not in the control group. Therefore, subjects only participate when the treatment arm is preferred to conventional care

$$EU(p, q_E; \theta) > U(p, q)$$

An increase in the quality of conventional care increases the expected utility outside the trial U_q*differentially more* than it raises the expected utility inside the trial, $(1 - w_1)U_q$. Therefore, an increase in the quality of care lowers trial participation even when conventional care is used as an active control.

III. Returns to innovation when development and output markets are linked

The economic determinants of development times affect the incentives to innovate in a number of ways. It is well known that longer development times reduce innovative returns by raising the costs of R&D. What is less recognized, however, is that longer development times may also reduce innovative returns by lowering the benefits of R&D, i.e. the present value of future variable profits. Markets for

medical products are peculiar because government regulation prohibits sales until development – specifically clinical testing – is completed. In such markets, longer development times have two important effects on the private benefits from R&D. First, they delay sales of the product and, thus, reduce the present value of the stream of revenue from product innovation. Second, when the product is protected by a patent, longer development time reduces the effective patent life, since patents are granted prior to development and have a fixed duration. Accounting for each of these mechanisms through which development times affect innovative returns, output markets have uniquely large effects on innovation in the medical products market.

*A. **Innovative returns and the role of output markets***

Let $c(f)$ denote the increasing cost-function of a firm in recruiting a flow of f patients each period. Let $\pi(k|q,p)$ denote the variable profits $k$years after market entry, given the quality and price of the conventional care that the innovation has to compete with upon market entry. We assume the profits from innovation are decreasing in the quality of the competing conventional care but increasing in the price of that care. The innovation has finite patent-life of length, T_P , after which the market becomes competitive and profits fall to zero. The net present value of the innovative return is given by

$$NPV \equiv B - C \equiv P\left[\int_{t=T}^{T_P} \beta^t \pi(t-T|q,p)dt\right] - \int_{t=0}^{T} \beta^t c(f)dt$$

where P is the probability of approval.[13]The first term is the expected benefit of R&D, which is the expected present value of variable profits during the period between the date of initial marketing and the expiration of the patent. The second term is the present value of R&D costs, which are incurred throughout the duration of development. In the simple case of constant annual R&D profits and costs, this net present value reduces to

$$NPV = \beta^T PA(T_P - T)\pi - A(T)c(f)$$

where $A(x)$ is the present value of a claim paying one dollar for x periods.

Output markets, through the quality and price of conventional care, affect both the benefits and costs of R&D, but we will focus on the arrival of future profits.[14] Longer development has an

[13] Given the regulatory criterion for approval, P is equal to the probability that experimental treatment is higher than some cut off rate, e.g., zero: $P = \Pr\{q_E > 0\}$.

[14]The impact on the costs of R&D is given by

$$\frac{dC}{dq} = \left(\frac{dA}{dT}\right)\left(\frac{dT}{dq}\right)c + A\left(\frac{dc}{df}\right)\left(\frac{df}{ds}\right)\left(\frac{ds}{dq}\right)$$

$$\frac{dC}{dp} = \left(\frac{dA}{dT}\right)\left(\frac{dT}{dp}\right)c + A\left(\frac{dc}{df}\right)\left(\frac{df}{ds}\right)\left(\frac{ds}{dp}\right)$$

unambiguously negative effect on the present value of these future profits.The reason is that longer development times both delay profits and reduce the duration of the profit stream.[15]

$$\frac{dB}{dT} = P\pi\left[\left(\frac{d\beta^T}{dT}\right)A(T_P - T) - \beta^T\left(\frac{dA}{dT}\right)\right] < 0$$

Longer development reduces the present value of a dollar of variable profits, $\beta^T A(T_P - T)$, by reducing the value of those profits when they start,β^T, and by shortening effective patent life during which the firm makes profits, $A(T_P - T)$.

Thus, in markets where development eats up patent life, like those for medical products, the value of faster development lies not only in its effect on R&D costs, but in the hastening and extension of the patent profit window. Indeed, for medical products, the output market effects on the benefits side of innovation may be much larger than any incremental changes in R&D costs.

Using the relationship between development time and output-market variables, $T(p, q)$, the total effect of output markets on innovative returns can be derived. First, consider the impact of an increase in the price of conventional care

$$NPV_p = P[V\pi_p + V_T T_p \pi] - C_p$$

The first term is positive as the profits from a new product increase in the price of conventional care (assuming that price does not respond, e.g., due to competitive pricing). The second term is also positive as higher prices hasten recruitment and reduce development times. These two effects reinforce each other, as both experimental care and the final product do better competing against a more expensive substitute. For example, consider European markets, which have more generous insurance subsidies and, thus, lower prices. Choosing such a site would both delay the start of lower profits and the duration of those profits. Development will be most attractive in markets where price of conventional care is high, such as the US, where private markets allow for higher pricing, or in developing countries, where conventional care outside of trials is often prohibitively expensive or infeasible.

The impact of the quality of conventional care on the full innovative returns is given by

$$NPV_q = P[V\pi_q + V_T T_q \pi] - C_q$$

The first term is negative, due to lower profits when the quality of competing conventional care improves. The second term is also negative as development times go up when an experimental therapy

---

Thus, an increase in the quality raises R&D costs by lengthening development but has the offsetting effect of lowering the cost per period of recruitment. Analogous offsetting effects occur for the effects of the price of conventional care.

[15] See Philipson and Sun (2010), which provides estimates of the magnitudes of these delay costs for several drug classes.

competes with higher quality conventional care. These two negative quality effects reinforce each other. An important implication of this negative quality impact from conventional care is a self-limiting effect of innovation. New innovation not only lowers the future profits from subsequent innovations but also slows down future development.

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.

*B. **Implications for the location of R&D***

The link between output markets and development for medical products has some non-standard implications for the location of medical R&D. In standard markets, where the knowledge generated by R&D is exported cheaply, it is efficient for the location of R&D spending to be independent of output market conditions across regions. Traditional R&D may be performed anywhere to serve the world output market; it is not restricted to the domestic market where it is performed. 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.

Despite this logic, the locus of medical R&D has shifted dramatically in the last two decades. In 1990, two-thirds of global medical R&D took place in Europe, but by 2010 two-thirds of R&D took place in the U.S. A common explanation is that European price controls are responsible for this large shift towards the US in R&D. This argument is dismissed by economists who use the logic above to argue that R&D should not track output markets. However, our analysis suggests that the non-standard links between output markets and medical development imply that the naive policy argument is the correct one.

Consider recruiting from two locations A and B with for an overall sample size of n. Overall development time is determined by

$$n = T\left(\frac{z_A s_A}{N_A} + \frac{z_B s_B}{N_B}\right)$$

This implies that the relative size of the recruited sample in the two regions is given by relative magnitudes of prevalence, subject participation rates, and the number of competing trials in each location. This in turn implies that profits and development may be positively linked across regions, contrary to standard arguments.

One way in which the two are related is that both profits and recruitment may be smaller in price-controlled markets, such as those in Europe. Price controls in output markets not only lower future profitability but also make recruitment in the region more difficult. Another way profits and development are related is that disease prevalence may vary across regions. A larger prevalence of a

disease increases demand in output markets and thus raise profits. But it also speeds up recruitment through two channels: a direct effect from greater prevalence and an indirect effect from higher prices of conventional care. Because of the prevalence effect, it is easier to recruit subjects for obesity trials in the US, which is also where the sales of any successful product will take place.

Indeed, looking at the overall pattern of R&D spending across regions, few industries besides medical products have as much development spending in the poorer continents of Africa and South America. The reason is that development talent is generally more concentrated in rich countries. For medical products, however, this factor must be balanced against the much quicker recruitment in very poor countries where, due to the lack of conventional care, the only means of getting any high quality care may be through experimental therapies.

*C. **Cross-disease effects***

The previous discussion concerned a single disease for which experimental and conventional care were gross substitutes (so s is increasing in p). When there are multiple diseases that have separate trials recruiting subjects, the price and quality of conventional care for a disease not only affect innovative returns for that disease, but also for other diseases through cross-disease effects.

Consider two diseases for which conventional care has the qualities $q = (q_1, q_2)$ and prices $p = (p_1, p_2)$. There are two channels through which conventional care for one disease affects the development times and innovative returns of another. The first is by affecting the stock of the other disease, either by affecting entry into or exit out of that disease state. This mechanism may make the stock of one disease dependent on the prices and qualities of care for the other disease: $z_1(p, q)$ and $z_2(p, q)$. The second channel by which cross-disease effects may operate is through subjects' incentives to participate in trials: $s_1(p, q)$ and $s_2(p, q)$.

The fact that output markets affect prevalence and trial participation across diseases implies that conventional care has cross-disease effects on development times and innovative returns. The impact of quality and price of conventional care of one disease on the development times of the other is given by

$$\frac{dT_2}{dq_1} = \frac{\partial T}{\partial s_2}\left[\frac{ds_2}{dq_1}\right] + \frac{\partial T_2}{\partial z_2}\left[\frac{dz_2}{dq_1}\right]$$

$$\frac{dT_2}{dp_1} = \frac{\partial T}{\partial s_2}\left[\frac{ds_2}{dp_1}\right] + \frac{\partial T_2}{\partial z_2}\left[\frac{dz_2}{dp_1}\right]$$

In a manner analogous to the impact of quality and price on the disease itself, quality and price of conventional care for one disease affects the development time for another disease by raising or lowering the prevalence of that disease and by affecting trial participation rates among those eligible for investigational research. For example, if quality or prices improve for treating heart disease, this leads to a larger population facing co-morbid risks, such as Alzheimer's disease. This in turn affects development times for treatments of such competing risks. Whereas innovation may be self-limiting within a disease, it may increase innovative returns for competing diseases.

The cross-disease impact of conventional care on innovative returns may be illustrated by an example where the quality of care of a given disease increases:

$$\frac{dNPV_2}{dq_1} = V_2\left(\frac{d\pi_2}{dq_1}\right) + \frac{dV_2}{dT_2}\left(\frac{dT_2}{dq_1}\right)\pi_2 - \left(\frac{dC_2}{dq_1}\right)$$

The first term captures the direct effect of conventional care for the first disease on the profitability of treatments for the second disease. For example, the first treatment may affect prevalence of the second disease or it may be a substitute for or complement to care of the second disease. The second term captures the effect that care for the first disease has on the development times of the second disease as discussed above. This affects both the arrival and duration of profits of marketable products. The last term concerns the direct effect on development costs.

IV. Rationing Supply through Development Times rather than Price

An important feature that distinguishes the labor market for human research subjects 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 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, Institutional Review Boards do not even allow trials to fully compensate for non-time expenditures (Ripley et al. 2010).

Since innovators cannot use wages to equilibrate the demand and supply of subjects, other factors determine the profitability of trials. One such factor is the speed at which ideas are generated and then developed into new products. In general, this is determined by the number and talent of researchers. In the health care sector, however, 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. An increase in the number of trials implies an increase in development times and a reduction in the profitability of trials. In the presence of binding wage controls, these delays are how subjects are rationed. In this section we examine how the price and quality of conventional care affect the development time under rationing by delay.

As discussed before, trials take longer the more there are competing for a given supply of subjects, $T_N > 0$. In addition, the present value of profits falls in development time, $NPV_T < 0$. This implies that the present value of development falls in the number of trials. The equilibrium number of trials, implicitly defined by (N, q, p), dissipates the profits from entering

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

Defined in this manner, the equilibrium number of trials will be affected by output markets. As previously discussed, the development time, $T(N, q, p)$, increases in the number of trials and in the quality of conventional care and falls in its price. Applying the implicit function theorem implies that the number of trials in equilibrium is affected by output markets:

$$\frac{dN}{dq} = \frac{1}{T_N}\left[\frac{NPV_q}{-NPV_T} - T_q\right] \leq 0$$

$$\frac{dN}{dp} = \frac{1}{T_N}\left[\frac{NPV_p}{-NPV_T} - T_p\right] \geq 0$$

This implies that 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 leading to fewer trials in equilibrium. Analogous arguments imply that increasing the price of conventional care raises the number of trials in equilibrium. Given this effect of quality and price on the number of trials, these output market parameters affect development times in equilibrium according to

$$\frac{dT}{dq} = T_q + T_N\left(\frac{dN}{dq}\right)$$

$$\frac{dT}{dp} = T_p + T_N\left(\frac{dN}{dp}\right)$$

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

It is important to note that when development times ration trial participant supply in this manner, the flow of subjects recruited *per period* into trials, s, is independent of demand, N, because price controls do not allow wages to increase when demand rises. However, the number of periods or length of development is not independent of demand; it rises with increases in N, $T_N > 0$, to ration supply. 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 supply at the constrained price, not the higher demanded amount.

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

The link between R&D and output markets for medical products 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 flow of patients into clinical trials. In this section we examine empirical evidence for this link in the case

of innovation in AIDS drugs, focusing on the effects of output quality and price on participation: $s_q < 0, s_p > 0$.

First, we present reduced-form estimate 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 25-75%within a few years. This result stands up even when we employ a difference-in-difference identification strategy the compares the effect on trial participation for AIDS drugs that do not directly compete against HAART versus those that do.

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 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/mm^3, the patient becomes acutely vulnerable to non-HIV infections, so-called secondary infections, and is classified as having AIDS (CDC 1993).[16]

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 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

[16] 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.

own were unable to affect progression from HIV to AIDS. In December 1995, the FDA approved the first protease inhibitor and, in January 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). 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 (Chaisson& Moore 1999).

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.[17] We first isolate a subsample of 1,861 unique subjects observed in the period 1992-2005.[18] 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.[19]

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 1,214 individuals in our subsample who actually took primary or secondary drugs at some point and the timing of their pharmaceutical regimens. Of these, 174 took only primary drugs, 56 took only secondary drugs, and 984 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 three quarters of subjects have some medical insurance before and after HAART, though that masks a substantial change in the sources of insurance. The most common source of

[17] More information on the MACS study is available at http://www.statepi.jhsph.edu/macs/macs.html.

[18] We begin our analysis in 1992 rather than 1984 because MACS changed the questionnaire that asked about trial participation in that year. We are not certain that pre-1992 data on participation are comparable to post-1992 data.

[19] 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 is private coverage which falls nearly 10 percent from 61% pre-HAART to 53% after. Government insurance picks up most of the slack, covering 17% of the sample pre-HAART but 23% after. The most important program for government insurance 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 - $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 1 documents the changing pattern of primary AIDS drug use over time. Panel A gives a breakdown of primary drug use by class of antiviral drugs. While use of NRTIs such as AZT are constant at nearly 100% throughout the 1990s, there is a sharp rise in use of protease inhibitors (PI) and non-NRTIs (NNRTI) after they are introduced in December 1995 and the end of 1996, respectively. At the same time we see a sharp drop in use of other classes of antiviral drugs.[20] These results are consistent with the summary statistics in Table 1, which show that the number subjects using primary drugs increased from almost 60% to 80% and the number of subjects using secondary drugs fell from 41% 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.[21]

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 275 to 494/mm^3. (Recall that a higher CD4 count implies better health: an HIV patient with a CD4 count below 200 is considered to have AIDS.) Figure 2, which plots CD4 counts for subjects on primary drugs, provides even more detail on the trend. 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 bottom 25% of subjects. Until 1995, all these subjects had AIDS (CD4 count < 200/mm^3), but after HAART, their CD4 counts improved to the point where, by 1997, the 25th percentile subject did not have AIDS.

Finally, we turn to trial participation. According to Table 1, 8.3% 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. Figure 3 goes further and compares the changes in trial participation for primary drugs to trial participation for secondary drugs. The important difference between the two

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

[21] 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.

types of drugs is that HAART increased the quality of conventional primary drugs but did not affect the quality of conventional secondary drugs. Thus, drug users saw an increase in the quality of conventional care for primary drugs after 1995, but not for secondary drugs. Our model implies that patients should have dropped out of trials for primary drugs after 1995, but not out of trials for secondary drugs.[22]

This is the basic pattern in the raw data. The probability that a primary drug user obtained his medications in a trial increased during 1992-1995 to a peak of 30% and then fell dramatically to 10% in 1997. After that it fell more gradually, eventually reaching 5% by 2005. In contrast, the probability that a secondary drug user obtained his drug in a trial had more modest changes. It rises to 10% in 1995 and then drops to 7% in 1997 and around 3% in 2005.[23] 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 within 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 by the total number of subjects not in trials and the exit rates 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 5 indicates, exit rates rose from 30% to over 40% in 1996 and 1997. Moreover, the average exit rate rose from around 20% prior to HAART to over 30% after HAART.

E. Regression analysis

1. Main results

Our primary strategy for estimating the effect of HAART on trial participation involves a regression of the form:

$$s_{it} = \beta_1 HAART_t + \gamma_i + \lambda X_{it} + e_{it} \qquad (1)$$

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 coefficient on an indicator $HAART_t$ that proxies for the

[22] Later, we will use trial participation for secondary drug use as a control to net out unrelated trends in trial participation. In that context, the difference between the trial participation for primary drug use and trial participation for secondary drug use – a difference-in-difference estimator – is used to estimate the effect of HAART on trial participation among primary drug users.

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

introduction of HAART. We employ three proxies for HAART's introduction. The first is simply an indicator that turns on in the period 1996-2005. The second proxy examines three-year pre and post windows. The pre window spans 1992-1994 and the post window spans 1997-1999. As is common with such estimation strategies, we drop 1995 and 1996 as transition years. The third proxy examines one-year pre-post windows. The pre window covers 1995 and the post covers 1996. While the pre-post windows focus on the short-term treatment effects of HAART, the post-HAART indicator provides an estimate of the long-term effect. In some specifications we included individual fixed effects (γ_i), so that treatment effects are identified from within-subject changes in participation. Some specifications also included time-varying controls for income and CD4 count.

In most cases we treat the regression equation as a linear probability model and estimate it using OLS. However, we verify the results 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. The first three columns in each panel report the results of linear regression specifications. Specification 2 adds subject fixed effects and specification 3 further adds time-varying covariates for income and CD4 count. The last specification keeps all the controls in specification 3 but estimates a logit regression. HAART appears to lower trial participation rates by 4-12 percentage points in linear specifications. We can estimate the proportional effect in these specifications by comparing the treatment effect to the constant, which captures the pre-HAART participation rate. This comparison suggests that HAART lowered trial participation roughly 17% in the short run (panel C) to nearly 48% in the long run (panel A).[24] The logit regression gives larger estimates for all proxies. All treatment effects are highly 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.

[24] Since the participation rate is rising prior to HAART, the baseline is lower in Panel A, where the pre-period is 1992-1995, than Panel B, where the pre-window is 1993-1995.

One way we address is this to include a parametric (say linear or quadratic) time trend in our basic regressions. We also take an alternative approach by employing 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.[25]

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 is not fatal to identification, 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, there is no reason why 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 regression:

$$s_{ijt} = \beta_1 HAART_t + \beta_2 PRIM_{ijt} + \beta_3\left(HAART_t \times PRIM_{ijt}\right) + \gamma_i + \lambda X_{it} + e_{it} \qquad (2)$$

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. We also 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 (1).

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.[26] 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

[25] 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.

[26] 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.

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 the last two columns of 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 3. In the medium- and long-run, however, HAART reduce participation by more than 50%.

Because time trends are a crude control for unrelated trends in trial participation and the estimates from the difference-in-difference specifications are roughly the same as from the before-after comparison in the medium- and long-term, we used just the latter in subsequent analyses.

b. Survival bias

A limitation of the results above is that they may be driven in part by the changing composition of subjects in the data over time. Subjects enter and exit our sample after the initial enrollment. A subject may show up for visits in one year but not another. Moreover, subjects may drop out permanently due to death.[27]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. If drop out or death is driven by sickness, survival bias is somewhat diminished by our inclusion of CD4 counts in specifications 3 and 4 of our basic regression model (1). That control attempts to hold constant a key measure of HIV infection severity so that the $HAART_t$ variable compares people with similar levels of viral load. However, the power of that control depends heavily on the adequacy of a linear specification.

To better test how substantial this bias may be for permanent drop outs, we estimated equation (1) on a subset of subjects that we confirmed were alive as of 2000 or as of 2005. 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 presented in Table 4 suggest that for either the alive-in-2000 or the alive-in-2005 sample, the results are similar to that from analysis of the more inclusive specification from before (Table 2). This implies that there is no substantial survival bias in our estimates.[28]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

[27] Because of temporary drop out, our subsample starts in 1992 with 1,649 subjects rather than all 1,861 unique subjects. Because the rate of drop out exceeds the rate of entry, the sample size falls monotonically to 738 subjects by 2005. On average, each subject is observed for six years.

[28] 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.

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 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, did it operate through improvements in conventional care's quality or 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 IV documented that prices are capped in the market for human subjects. As a result, there is likely to be excess demand for research 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 length of development, T, is affected by demand. Therefore, our estimates of the effect of HAART on per period trial participation rate is 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.[29] In other words, our identifying assumption for the effect of HAART on supply is that the equilibrium wage is positive.

Second, we also 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.[30] In standard labor markets "quit" versus "fired" is hard to identify, but this is less of an issue for clinical trial recruitment. In practice, we can rule out "fired" 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).[31] The rules do not permit cancellation because some other drug outside the trial worked. Empirical support

[29] 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.

[30] 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.

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

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 4, 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 5: there was a sharp increase in the exit rate in 1996 with the level of exit rates remaining elevated throughout 1996-2005. The regression analyses in Table 5 also provide additional evidence of HAART's impact. We employed the same specifications as in the previous subsection, 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. The results suggest that exit rates increased sharply after HAART for all time horizons we examine. In the short-run, they increased by 73% or more (depending on the specification in Panel C). In the long-run, they increased by 62%or more (Panel A).

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 supply and demand influences 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 40%,which is consistent with the graphical data in Figure 5. We find no medium-term effects on entry and small and mostly insignificant effects in the long run. From these results we infer that the reduction in total trial participation is entirely explained by an increase in exits. Since exits are driven only by the subjects, 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 6 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 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 2 suggest HAART clearly improved the quality of conventional treatment (as measured by health outcomes). But HAART also increased the price of such treatment. 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

$$s_{ijt} = \beta_1 HAART_t + \beta_2 NO\_INS_{ijt} + \beta_3\left(HAART_t \times NO\_INS_{ijt}\right) + \gamma_i + \lambda X_{it} + e_{it} \quad (3)$$

whereNO_INS_{it} is an indicator for not having insurance. The common effect across the insured and uninsured groups (β_1) captures the effect of a change in quality after HAART. The differential impact on the uninsured (β_3) captures the effect of a change in price after HAART.

Since different types of insurance may have different co-pays for drug therapies, we provide estimates for four groups with different insurance: those with any insurance; those with private insurance; those on any form of government insurance; and those specifically with Medicaid, for whom we can confirm generous drug coverage.[32]For each we restrict the sample to the insured group of interest and those subjects without any form of insurance. Moreover, for each comparison we examine two outcomes: overall participation and exit rates.

The results are presented in Table 7. We find that that HAART reduced trial participation (increased trial exit) among insured populations. Relative to this group, uninsured populations experienced an increase in participation (and reduced exit), though this effect was insignificant in many specifications. This implies that the quality improvement due to HAART significantly reduced the supply

[32] Medicaid beneficiaries faced the same co-pays for conventional therapeutic drugs throughout the sample, whether it was AZT prior to 1996 or HAART after. Moreover, these programs covered all approved drugs and did not have annual or lifetime limits on HIV therapies during the sample period.

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-controlled 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 (3) separately for active- and placebo-controlled trials. The results are presented in Table 8. They confirm our prior findings in the long run. There are also consistent effects in the short run for active controlled trials and in the medium run for placebo controlled trials.

VI. Concluding remarks and implications for medical R&D policy

In this paper we argued that there is a link between output markets and development for medical products that distinguishes medical R&D from other forms of development. When the returns to development are linked to the quality and price in output markets, medical R&D has a self-limiting effect: past innovation lowers future innovative returns by making development more difficult. In addition, output market policies, such as universal coverage and price controls, impact R&D costs beyond their standard effects on R&D benefits. We examined evidence of this link between output markets and development in the case of break-through AIDS therapies introduced to the US market around 1996 and found a substantial reduction in the supply of subjects induced by these new innovations.

The linkage between output markets and development not only helps explain how medical innovation differs from other R&D, but also affects normative analysis. Optimal methods to stimulate medical R&D differ from the standard methods proposed in other markets. For example, implicit “pull” measures to stimulate innovation, e.g., demand subsidies such as Medicare in the US, may not have the standard effects when applied to medical innovation, especially when those subsidies lower the price of the standard of care and, hence, slow down trial recruitment. Likewise, “push” policies, such as NIH funding in the US, may affect research but not development, when the latter is driven by output markets.

A more central normative concern for medical R&D policy is the appropriate balance within the inter-temporal consumption problem that implicitly defines medical innovation; the costs of those consuming treatments in experiments must be balanced against the benefits for future consumers. Specifically, trial participants confer a positive external effect – information about treatment effects – on future patients in output markets. Thus, Pigouvian subsidies from consumers in output markets to subjects in experimental care would ideally compensate them for the public service of generating the knowledge about the quality for a future standard of care.

However, public regulation of clinical trials often bans compensation of subjects. That is, they require a positive experimental price, p_E, in our framework. These controls are usually enforced through public or private review of trial protocols, usually by Institutional Review Boards (IRBs) in the US. This oversight requires that trials first obtain the informed consent of patients before enrollment and not offer patients "excessive" financial compensation for enrollment. The informed consent requirement seems fully justified given the history of highly offensive studies conducted on humans during World War II and on US prisoners and African-Americans after the war. The benefits of wage-controls, however, may be dominated by the social costs discussed in this paper. So long as consenting subjects understand the risks of research, there is nothing special about compensating for health risks in trials. The same is done for other types of public service with positive external effects, e.g., service in the military.

Wage controls have two harmful effects as discussed in our analysis. First, they lead to inefficiently slow trials because development time, rather than price, is used to ration for the supply of subjects. Without wage controls, more trials might not necessarily mean longer development times. Second, these wage controls may induce investigators to engage in non-price competition for subjects, by altering a trial's design features. For example, investigators might lower the probability of receiving the control or employ active- rather than placebo controls to raise trial enrollment. These design changes might reduce the quality of information produced by trials.

Rather than constraining wages at zero, the efficient level of compensation should take into account the hastened arrival of the total social surplus from the innovation. In our framework, the present value of surplus would be $\beta^{T(p_E)}S + np_E$ where S is the immediate value of the surplus and p_E represents the price of experimental care. Optimal trial compensation, represented by a negative experimental price, would balance the value of shorter development times with the spending on compensation.[33] Although manufacturers would partially internalize the inter-temporal tradeoff between future consumers and trial subjects, they would under-pay subjects relative to the efficient Pigouvian subsidy if they fail to fully appropriate the entire social surplus.

More generally, we hope we have provided support for the claim that the unique relationship between output markets and development in medical innovation suggests that standard positive and normative analysis of R&D may not necessarily apply. Given the enormous economic benefits of

[33] This optimal balance may also have implications for the optimal designs of trials, since optimal sample sizes would be guided by economic-, rather than statistical-, efficiency.

previous medical advances and the potentially large gains that medical progress may deliver in the future, a better understanding of the unique dynamics of medical R&D seems warranted.

References

1. Christopher P. Adams and Van V. Brantner. Estimating The Cost Of New Drug Development: Is It Really $802 Million? Health Affairs, DOI 10.1377/hlthaff.25.2.420, (March/April 2006).

2. T. Chan, B. Hamilton. Learning, Private Information and the Economic Evaluation of Randomized Experiments. Journal of Political Economy (forthcoming).

3. Amitabh Chandra and Douglas Staiger. 2007. Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks. Journal of Political Economy. Feb: 103-140.

4. Sylvain Chassang, Gerard Padró i Miquel and Erik Snowberg. 2010. Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. Working Paper.

5. Joseph A. DiMasi, Ronald W. Hansen, and Henry G. Grabowski. The price of innovation: new estimates of drug development costs. Journal of Health Economics, 22: 151–185 (2003).

6. Richard D. Mooreand Richard E. Chaisson. 1999. Natural History of HIV Infection in the Era of Combination Antiretroviral Therapy. AIDS, 13(14), 1933-42.

7. GAO. New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts. GAO-07-49 (November 2006).

8. Pierre-Yves Geoffard and Tomas Philipson. 2002. Pricing and R&D when consumption affects longevity. RAND Journal of Economics. 33: 85-95.

9. Christine Grady, Neal Dickert, Tom Jawetz, Gary Gensler, Ezekiel Emanuel. 2005. An analysis of U.S. practices of paying research participants. Contemporary Clinical Trials. 26: 365-375.

10. Gulick, Roy M.; John W. Mellors; Diane Havlir; Joseph J. Eron; Charles Gonzalez; Deborah McMahon; Douglas D. Richman; Fred T. Valentine; Leslie Jonas; Anne Meibohm, et al. 1997. Treatment with Indinavir, Zidovudine, and Lamivudine in Adults with Human Immunodeficiency Virus Infection and Prior Antiretroviral Therapy." New England Journal of Medicine, 337(11), 734-39.

11. Hammer, Scott M.; Kathleen E. Squires; Michael D. Hughes; Janet M. Grimes; Lisa M. Demeter; Judith S. Currier; Joseph J. Eron; Judith E. Feinberg; Henry H. Balfour; Lawrence 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. New England Journal of Medicine, 337(11), 725-33.

12. Darius Lakdawalla and Neeraj Sood. Health Insurance as a Two-Part Pricing Contract. NBER Working Paper 12681 (2006).

13. Darius Lakdawalla, Neeraj Sood, and Dana Goldman. 2006. HIV Breakthroughs and Risky Sexual Behavior. Quarterly Journal of Economics, August 2006, 1063-1102.

14. Anup Malani. 2006. Identifying Placebo Effects with Data from Clinical Trials. Journal of Political Economy, 114: 236-256.

15. Anup Malani. 2008. Patient enrollment in medical trials: Selection bias in a randomized experiment. Journal of Econometrics. 144: 341-351.

16. Kevin Murphy and Robert Topel. 2006. The Value of Health and Longevity. Journal of Political Economy, 114(5): 871-904.

17. New York Times. AZT's Inhuman Cost. August 28, 1989.

18. Tomas Philipson and Jeffrey DeSimone. 1997. Experiments and Subject Sampling. Biometrika, 84(3): 618-632.

19. Tomas Philipson and Larry Hedges. 1998. Subject Evaluation in Social Experiments. Econometrica 66(2): 381-409.

20. Tomas Philipson and Anup Malani. 1999. Measurement Errors: A Principal Investigator-Agent Approach. Journal of Econometrics, 91:273.

21. Tomas Philipson. 1997. The Evaluation of New Health Care Technology: The Labor Economics of Statistics. Journal of Econometrics, 76 (1-2): 375-396.

22. Ripley, 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.

23. Robert Steinbrook. 2001. Providing Antiretroviral Therapy for Hiv Infection. New England Journal of Medicine. 344(11), 844-46.

24. U.S. Centers for Disease Control & Prevention. 1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults (1993), available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm.

TABLES

Table 1.Summary statistics.

		Pre-HAART (1992-1995)			Post-HAART (1996-2005)		
VARIABLES	Units	N	mean	sd	N	mean	sd
AIDS drug use							
AIDS treatment	0/1	2,413	0.584	0.322	4,498	0.799	0.233
AIDS secondary	0/1	2,413	0.416	0.322	4,498	0.201	0.233
Trial participation							
Any trial	0/1	2,413	0.083	0.216	4,498	0.0462	0.176
Placebo-controlled	0/1	2,413	0.0414	0.162	4,498	0.00931	0.0733
Employment status							
Full time	0/1	2,368	0.592	0.475	4,410	0.579	0.484
Part time	0/1	2,368	0.0921	0.268	4,410	0.0967	0.279
Income category							
<$10,000	0/1	2,411	0.148	0.337	4,486	0.106	0.294
$10,000-$19,999	0/1	2,411	0.145	0.327	4,486	0.118	0.304
$20,000-$29,999	0/1	2,411	0.14	0.319	4,486	0.114	0.296
$30,000-$39,999	0/1	2,411	0.152	0.334	4,486	0.125	0.309
$40,000-$49,999	0/1	2,411	0.116	0.295	4,486	0.118	0.3
$50,000-$59,999	0/1	2,411	0.244	0.413	4,486	0.253	0.42
$60,000-$69,999	0/1	2,411	0	0	4,486	0.121	0.319
>$70,000	0/1	2,411	0	0	4,486	0.000111	0.00747
Refused to answer	0/1	2,411	0.0542	0.2	4,486	0.0458	0.194
Insurance status							
Government	0/1	2,413	0.173	0.364	4,498	0.229	0.406
Medicare	0/1	2,410	0.0622	0.229	4,470	0.163	0.358
Medicaid	0/1	2,411	0.108	0.298	4,470	0.105	0.294
Veterans Admin.	0/1	2,411	0.035	0.173	4,470	0.0272	0.155
Private	0/1	2,413	0.61	0.471	4,498	0.529	0.48
Individual plan	0/1	2,410	0.131	0.318	4,470	0.084	0.254
Group plan	0/1	2,411	0.494	0.479	4,470	0.457	0.474
Other							
Age	yrs	2,381	41.53	7.081	4,401	47.4	7.052
CD4 count	cells/mm3	2,392	275.1	226.8	4,470	494	276.6

Table 2.Effect of HAART on trial participation rate: individual level, before-after comparisons.

	(1)	(2)	(3)	(4)
Panel A				
HAART	-0.081***	-0.117***	-0.097***	0.379***
	(0.006)	(0.015)	(0.016)	(0.035)
Constant	0.170***	0.194***	0.266***	
	(0.004)	(0.007)	(0.033)	
Panel B				
Post (3 year)	-0.048***	-0.046***	-0.045***	0.585***
	(0.007)	(0.018)	(0.018)	(0.068)
Constant	0.155***	0.152***	0.240***	
	(0.004)	(0.009)	(0.034)	
Panel C				
Post (1 year)	-0.039***	-0.061**	-0.057**	0.745**
	(0.013)	(0.028)	(0.028)	(0.106)
Constant	0.274***	0.298***	0.323***	
	(0.009)	(0.024)	(0.038)	
Specifications				
Obs.	6,579	6,579	6,529	2,695
Model	LPM	LPM	LPM	Logit
Subject FE		Y	Y	Y
Other covariates			Y	Y

Note: Panel A reports results of regressions of whether a subject participated in a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. Specifications (3) and (4) include controls for income and CD4 count. The table only reports the treatment effect and constant to calculate elasticities. 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 primaryAIDS drugs each year. Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

Table 3.Effect of HAART on trial participation rate: identification off of time trends or secondarydrug controls.

	(3)	(4)	(3)	(4)	(3)	(4)
Panel A						
HAART	-0.050**	0.786*	-0.116***	0.421***	-0.068***	0.711***
	(0.020)	(0.102)	(0.028)	(0.061)	(0.016)	(0.079)
Constant	0.275***		0.220***		0.130***	
	(0.032)		(0.033)		(0.020)	
Panel B						
Post (3 year)	0.041**	1.657***	0.015	0.763	-0.084***	0.425***
	(0.021)	(0.228)	(0.031)	(0.135)	(0.019)	(0.067)
Constant	0.240***		0.227***		0.115***	
	(0.033)		(0.032)		(0.021)	
Panel C						
Post (1 year)	-0.056**	0.753**	-0.057**	0.730**	-0.020	1.357
	(0.028)	(0.108)	(0.028)	(0.104)	(0.039)	(0.339)
Constant	0.355***		0.345***		0.134***	
	(0.037)		(0.038)		(0.026)	
Specifications						
Obs.	6,529	2,690	6,529	2,690	10,653	5,264
Model	LPM	Logit	LPM	Logit	LPM	Logit
Time trend	Linear	Linear	Quadratic	Quadratic	None	None
Secondary drug userDD	N	N	N	N	Y	Y

Note: Panel A reports results of regressions of whether a subject participated in a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. Specifications (3) and (4) include individual fixed effects and controls for income and CD4 count. The table only reports the treatment effect and constant to calculate elasticities. Specification (4) reports coefficients as odds ratios. In the first four columns, 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 primaryAIDS drugs each year in the first four columns. In the last two columns, 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. In all columns, standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

Table 4.Effect of HAART on trial participation rate: subsample of subjects alive late in the sample.

	(1)	(2)	(3)	(4)
Panel A				
HAART	-0.099***	0.361***	-0.105***	0.353***
	(0.017)	(0.039)	(0.023)	(0.049)
Constant	0.277***		0.329***	
	(0.037)		(0.047)	
Panel B				
Post (3 year)	-0.050***	0.604***	-0.045*	0.684**
	(0.019)	(0.083)	(0.025)	(0.120)
Constant	0.249***		0.301***	
	(0.038)		(0.049)	
Panel C				
Post (1 year)	-0.078***	0.570***	-0.099***	0.559**
	(0.028)	(0.110)	(0.035)	(0.139)
Constant	0.354***		0.420***	
	(0.042)		(0.055)	
Specifications				
Obs.	5,202	2,270	3,650	1,690
Model	LPM	Logit	LPM	Logit
Subjects in sample at least until	2000	2000	2005	2005

Note: Panel A reports results of regressions of whether a subject participated in a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. Specifications (3) and (4) include individual fixed effects and controls for income and CD4 count. The table only reports the treatment effect and constant to calculate elasticities. 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 AIDS primary drugs each year and were either in the sample as of 2000 (first two columns of results) or 2005 (last two columns). Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

Table 5.Effect of HAART on supply: before-after comparisons of exit rates.

	(1)	(2)	(3)	(4)
Panel A				
HAART	0.209***	0.259***	0.320***	6.242***
	(0.042)	(0.090)	(0.093)	(2.416)
Constant	0.518***	0.230***	0.074	
	(0.017)	(0.044)	(0.165)	
Panel B				
Post (3 year)	0.135***	0.164	0.217*	3.295***
	(0.044)	(0.119)	(0.125)	(1.291)
Constant	0.505***	0.226***	0.121	
	(0.018)	(0.046)	(0.194)	
Panel C				
Post (1 year)	0.186***	0.223**	0.253**	3.318***
	(0.054)	(0.114)	(0.112)	(1.136)
Constant	0.478***	0.154	0.030	
	(0.041)	(0.103)	(0.169)	
Specifications				
Obs.	673	673	670	400
Model	LPM	LPM	LPM	Logit
Subject FE		Y	Y	Y
Other covariates			Y	Y

Note: Panel A reports results of regressions of whether a subject exited a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. Specifications (3) and (4) include controls for income and CD4 count. The table only reports the treatment effect and constant to calculate elasticities. 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 include only subjects that started in trials for AIDS primary drugs each year. Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

Table 6.Effect of HAART on supply: before-after comparisons of entry rates.

	(1)	(2)	(3)	(4)
Panel A				
HAART	-0.084***	-0.071***	-0.021	0.925
	(0.008)	(0.019)	(0.020)	(0.054)
Constant	0.636***	0.622***	0.774***	
	(0.005)	(0.009)	(0.044)	
Panel B				
Post (3 year)	0.003	0.023	0.015	1.070
	(0.011)	(0.026)	(0.026)	(0.074)
Constant	0.630***	0.608***	0.785***	
	(0.005)	(0.011)	(0.045)	
Panel C				
Post (1 year)	0.239***	0.268***	0.280***	5.204***
	(0.020)	(0.036)	(0.037)	(0.688)
Constant	0.684***	0.656***	0.721***	
	(0.014)	(0.030)	(0.049)	
Specifications				
Obs.	6,475	6,475	6,425	5,738
Model	LPM	LPM	LPM	Logit
Subject FE		Y	Y	Y
Other covariates			Y	Y

Note: Panel A reports results of regressions of whether a subject entered a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. Specifications (3) and (4) include controls for income and CD4 count. The table only reports the treatment effect and constant to calculate elasticities. 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 include only subjects that started in trials for AIDS therapeutic drugs each year. Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

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

Panel A								
HAART	-0.106***	0.558***	-0.100***	0.542***	-0.173***	0.480***	-0.169***	0.372*
	(0.019)	(0.067)	(0.021)	(0.071)	(0.042)	(0.153)	(0.056)	(0.204)
HAART x	0.037	-0.349**	0.037	-0.391***	0.083*	-0.255	0.087	-0.025
uninsured	(0.029)	(0.136)	(0.032)	(0.139)	(0.045)	(0.193)	(0.060)	(0.239)
Constant	0.273***	-0.011	0.245***	-0.109	0.352***	0.332	0.345***	0.467*
	(0.034)	(0.164)	(0.050)	(0.178)	(0.044)	(0.228)	(0.059)	(0.270)
Panel B								
Post (3 year)	-0.042**	0.481***	-0.037	0.493***	-0.083*	0.228	-0.061	0.020
	(0.020)	(0.107)	(0.023)	(0.108)	(0.043)	(0.269)	(0.055)	(0.282)
Postx	-0.007	-0.177	-0.008	-0.231	-0.001	0.061	-0.014	0.282
uninsured	(0.033)	(0.181)	(0.037)	(0.187)	(0.050)	(0.295)	(0.062)	(0.301)
Constant	0.235***	0.073	0.202***	-0.098	0.290***	0.514**	0.274***	0.585**
	(0.035)	(0.211)	(0.051)	(0.191)	(0.045)	(0.246)	(0.060)	(0.248)
Panel C								
Post (1 year)	-0.103***	0.384***	-0.111***	0.350**	-0.125*	0.432*	-0.130	0.422
	(0.035)	(0.133)	(0.039)	(0.140)	(0.076)	(0.226)	(0.099)	(0.294)
Post x	0.166***	-0.543**	0.179***	-0.446*	0.210**	-0.520	0.214*	-0.319
uninsured	(0.062)	(0.261)	(0.065)	(0.258)	(0.090)	(0.331)	(0.111)	(0.388)
Constant	0.357***	0.191	0.336***	0.070	0.401***	0.549	0.405***	0.767**
	(0.042)	(0.183)	(0.057)	(0.197)	(0.070)	(0.334)	(0.080)	(0.349)
Specifications								
Obs.	6,529	670	5,635	597	3,506	324	2,863	250
Dep. variable	Partici-pation	Exit	Partici-pation	Exit	Partici-pation	Exit	Partici-pation	Exit
Treatment group	Un-insured	Un-insured	Un-insured	Un-insured	Un-insured	Un-insured	Un-insured	Un-insured
Control group	Insured	Insured	Private insured	Private insured	Govt insured	Govt insured	Medicaid	Medicaid

Note: Panel A reports results of regressions of whether a subject participated in a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. All regressions include individual fixed effects and controls for income and CD4 count; they are comparable to specification (3) in prior tables. Estimation is by OLS. The table only reports the treatment effect and constant to calculate elasticities. 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 AIDS primary drugs each year and who fall either in the treatment group or the control group. Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

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

Panel A				
HAART	-0.059***	0.377***	-0.060***	0.593***
	(0.016)	(0.094)	(0.013)	(0.153)
HAART x	0.041*	-0.263*	-0.007	-0.202
uninsured	(0.024)	(0.155)	(0.020)	(0.418)
Constant	0.229***	0.634***	0.084***	0.276
	(0.030)	(0.208)	(0.021)	(0.271)
Panel B				
Post (3 year)	0.003	0.332***	-0.053***	0.602***
	(0.018)	(0.122)	(0.013)	(0.141)
Postx	0.017	-0.075	-0.019	0.295
uninsured	(0.027)	(0.195)	(0.024)	(0.302)
Constant	0.187***	0.663***	0.088***	0.406
	(0.031)	(0.240)	(0.021)	(0.248)
Panel C				
Post (1 year)	-0.090***	0.135	-0.006	0.360
	(0.033)	(0.157)	(0.020)	(0.233)
Postx	0.097*	-0.196	0.057	-1.028**
uninsured	(0.054)	(0.277)	(0.038)	(0.494)
Constant	0.307***	0.864***	0.084***	0.489**
	(0.037)	(0.259)	(0.025)	(0.244)
Specifications				
Obs.	6,477	523	6,529	210
Type of trial	Active-controlled	Active-controlled	Placebo-controlled	Placebo-controlled
Dep variable	Participation	Exit	Participation	Exit
Model	LPM	LPM	LPM	LPM
Treatment group	Uninsured	Uninsured	Uninsured	Uninsured
Control group	Insured	Insured	Private insured	Private insured

Note: Panel A reports results of regressions of whether a subject participated in a trial on an indicator variable for the HAART era (1996-2005). Panel B replaces the 1996-2005 HAART dummy with pre (1992-1994) and post (1997-1999) windows. Panel C employs even shorter 1 year pre (1995) and post (1996) windows. The pre window indicator is omitted and thus captured by the constant. All regressions include individual fixed effects and controls for income and CD4 count; they are comparable to specification (3) in prior tables. The table only reports the treatment effect and constant to calculate elasticities. 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 AIDS primary drugs each year. Standard errors are clustered at the subject-level. ***/**/* indicate p<0.01/0.05/0.1.

FIGURES

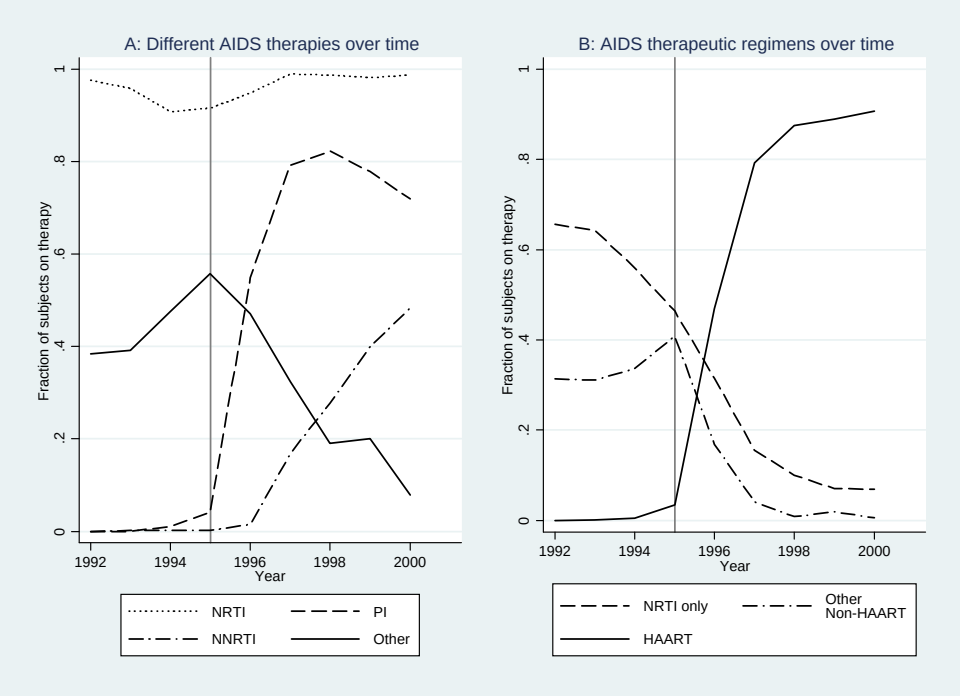

Figure 1.Effect of HAART on choice of AIDS therapy.

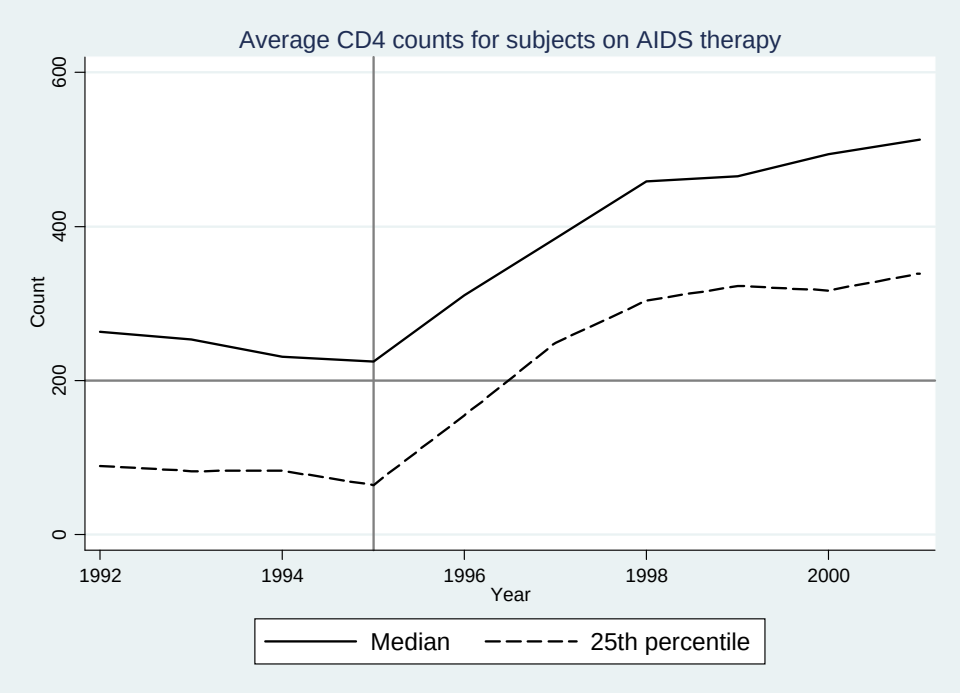

Figure 2.Effect of HAART on CD4 counts.

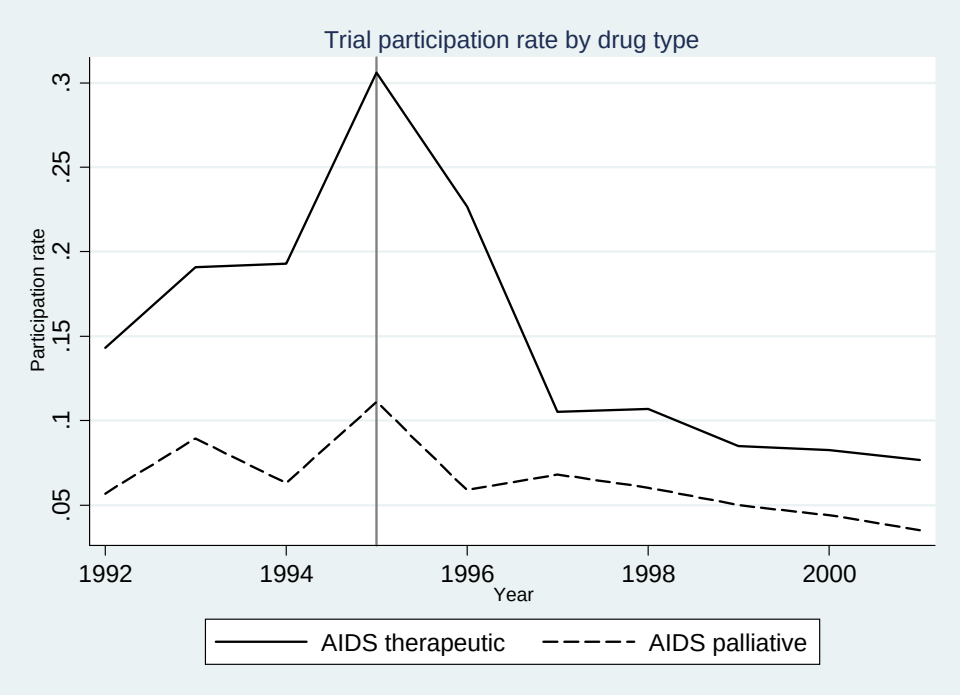

Figure 3.Effect of HAART on trial participation.

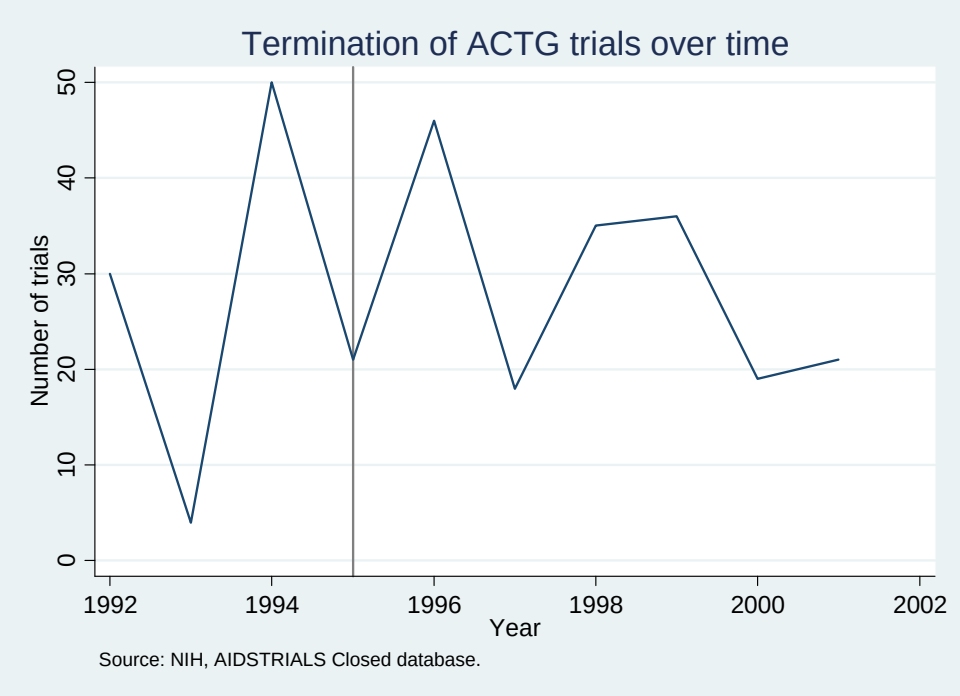

Figure 4.Termination of ACTG trials over time.

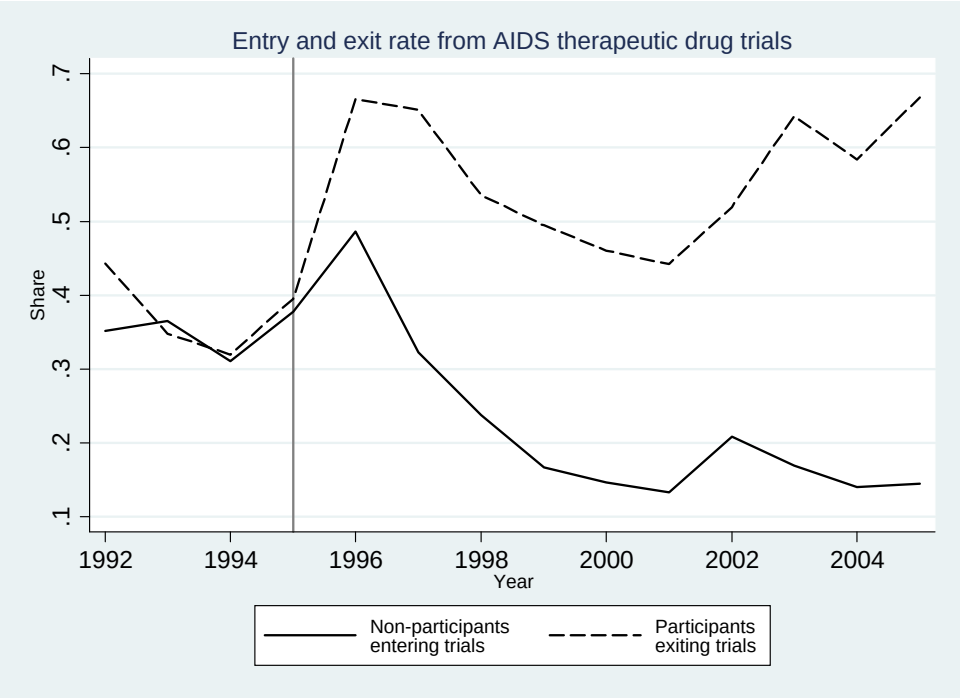

Figure 5.Effect of HAART on trial entry and exit.

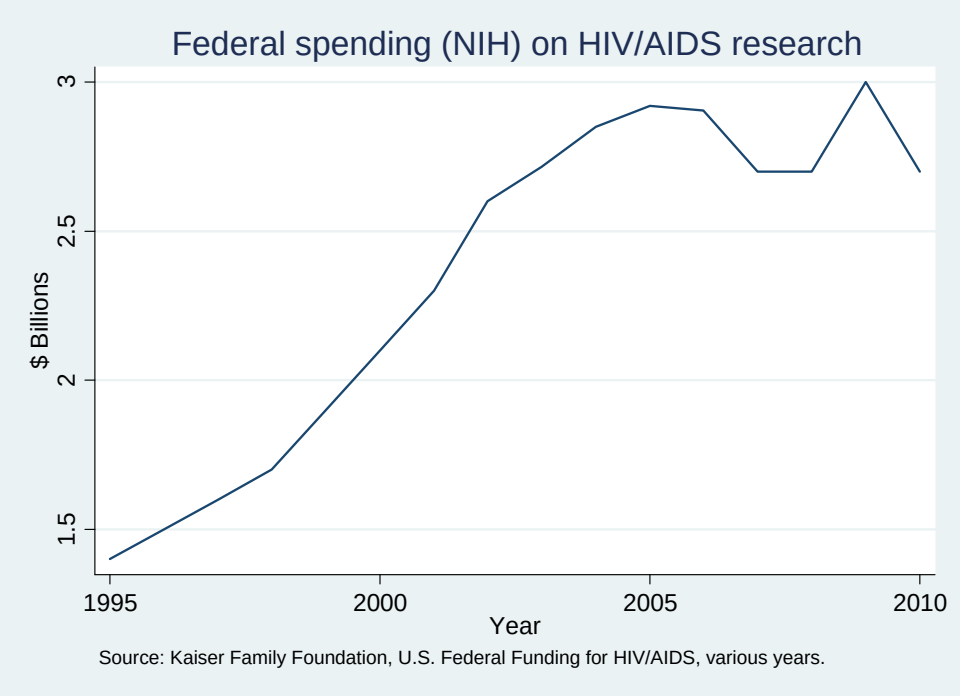

Figure 6.Federal government spending on HIV/AIDS research.