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

Anup Malani
University of Chicago

Tomas J. Philipson
University of Chicago

Medical R&D differs from other R&D because of a unique linkage between output and input markets: potential consumers of existing medical products are also potential subjects in clinical trials required to develop new products. Therefore, a quality increase or price reduction for an existing treatment reduces patients' incentive to participate in trials of new treatments. We label this linkage the "subject supply effect" and provide evidence of it from trials of hepatitis C and AIDS treatments. The subject supply effect has important positive implications for how policies affect the rate of medical R&D and normative implications for whether subjects ought to be compensated for enrolling in clinical trials.

Medical research and development (R&D) plays a central role in health care economics. On the one hand, it is an important driver of the benefits of health care. Improvements in health have been a major contributor to gains in overall economic welfare during the past century (Murphy and Topel 2006) as well to the reduction in global inequality (Becker, Philipson, and Soares 2005). A significant part of these gains has been attributed to medical R&D, including improvements in medical knowledge, procedures, drugs, biologics, and devices (Lichtenberg 1998; Cutler et al. 2003; Cutler 2007). On the other hand, innovation is an important driver of health care costs. The medical consumer price index (CPI) has recently

We thank Leila Agha, Gary Becker, Tatyana Deryugina, Einer Elhauge, Jose Fernandez, Dana Goldman, Eric Helland, William Hubbard, Darius Lakdawalla, Richard Miller, Kevin Murphy, Julian Reif, Seth Seabury, Neeraj Sood, Tony Tse, Heidi Williams, and workshop and conference participants at the University of Southern California, the University of Chicago, Harvard Law School, the National Bureau of Economic Research Summer Institute, the Association Lecture at the Southern Economic Association, HEC (*Hautes études commerciales*) Montreal, and the Area Health Education Centers 2012 meeting for helpful comments; and Ilya Beylin, Mete Karakaya, Rafeh Qureshi, Nate Reid, and Christopher Whaley for research assistance. Malani acknowledges financial support from the Samuel J. Kersten Faculty Fund and the Microsoft Fund at the University of Chicago Law School, and Philipson acknowledges support from the Biotechnology Industry Organization and the George Stigler Center for the Study of the Economy and the State at the University of Chicago.

[*Journal of Human Capital*, 2019, vol. 13, no. 2]

been 1.5 percent higher than the consumer CPI, and health expenditures increased from 8 percent of GDP in 1990 to 16 percent in 2008. Newhouse (1992) and others have attributed much of this growth in costs to technology.

Given the importance of medical innovation in the health care system, there is considerable attention given to the high and growing costs of developing medical R&D as well as the slow pace of development. DiMasi, Hansen, and Grabowski (2003) estimate that the cost of bringing a new drug to market is over $800 million. More recent estimates suggest that these development costs may be as high as $1.7 billion per drug (Adams and Brantner 2006; Brichet and Cohen 2007). In addition, because of regulation-mandated clinical testing, it takes 7.5 years, on average, from start of testing to marketing of drugs (DiMasi et al. 2003). A key hurdle to completing testing in a timely manner is the difficulty in recruiting human subjects into trials (Abraham, Young, and Solomon 2006). One study of 2,685 trials at 14 cancer research centers found that more than 50 percent of trials failed to recruit even one patient (Durivage and Bridges 2009). An important gap in the literature on clinical trials, however, is the absence of positive analysis of what drives the costs and pace of medical R&D.

Under the traditional economic view, the private benefits of R&D are driven by future variable profits in the output market, and the costs of R&D are driven by input supply, for example, the quantity and location of R&D talent. We argue in Section I, however, that an important difference between medical and nonmedical R&D is a unique linkage between medical output markets and medical R&D inputs. This linkage is driven by the twin facts that medical R&D requires clinical trials on human subjects[1] and that individuals who can serve as human subjects are also potential consumers of medical products as patients in the output market.[2] In other words, input supply for medical R&D overlaps with output demand for medical products. The central implication of this overlap, which we term the "subject supply effect," is that improvements in output markets, for example, an increase in quality or a reduction in price of conventional medical care, make patients more reluctant to enroll as subjects in clinical trials.

A recent example of the subject supply effect can be seen in clinical trials for drugs to treat hepatitis C (HCV), a viral infection of the liver. Be-

[1] 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 that their 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.

[2] 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.

fore 2013, the primary treatment for HCV was interferon alfa. The cure rate was less than 50 percent, and there were serious side effects from interferon therapy (e.g., fatigue and headaches). In 2013, the US Food and Drug Administration (FDA) approved the antiviral sofosbuvir (sold as Sovaldi), a drug produced by Gilead, for the treatment of genotype 1 HCV in patients, the most common type of HCV. After Sovaldi, the cure rate jumped to 75–90 percent, depending on a patient's genotype. However, Sovaldi had to be used with interferons to achieve maximal efficacy, and interferons had serious side effects. Some relief came in 2014, when the FDA approved the combination of antivirals sofosbuvir and ledipasvir (sold together as Harvoni by Gilead). Harvoni is as effective as Sovaldi for patients of genotype 1, but Harvoni has far fewer side effects because it does not require the concomitant use of interferons.[3]

Our theory suggests that enrollment in trials of Gilead's Harvoni and trials of competitors' drugs that did not use interferon should have declined after approval of that drug in 2014. As demonstrated in figure 1, that is exactly what we see.[4] After Gilead filed for a drug approval application (a new drug application), new enrollment in Harvoni trials fell dramatically. Moreover, after approval of Harvoni, new enrollment in trials of Harvoni competitors also fell dramatically. The former effect is almost surely driven by demand: Gilead likely stopped Harvoni trials after they filed for drug approval, as additional trial results had no further value for its FDA application. The latter was plausibly driven by subject supply effects, however, as patients dropped out of trials to access Harvoni directly.[5] Of course, these data are not definitive, as demand for research subjects among competitors may also have fallen. The figure shows an equilibrium quantity, which reflects both supply and demand.

Therefore, in Section II, we provide somewhat more rigorous evidence of the supply effect in the context of HIV/AIDS, specifically, the introduction of highly active antiretroviral therapy (HAART) in 1996. HIV is a virus that, when it reaches a critical level of concentration in the bloodstream, triggers AIDS, a condition that cripples the immune system, rendering a person defenseless against secondary infections by bacteria and

[3] Sovaldi was also approved for HCV of genotypes 2, 3, and 4, though those are less common.

[4] The data on enrollment in HCV trials were scraped from https://clinicaltrials.gov. We searched for clinical trials of the components of Harvoni or competitors Daklinza, Zepatier, and Viekira Pak. We scraped the total number of subjects enrolled, the start date for recruitment, and the end date for recruitment. The flow of subjects into a trial is calculated under the assumption that an equal number of subjects were enrolled each month. The sum of subjects for a drug is the sum of calculated enrollment flow across trials for that drug in a given month. A trial is called a Harvoni (Harvoni competitor) trial if the best treatment available in the trial is Harvoni (a Harvoni competitor).

[5] Some additional evidence of the subject supply effect may be the ethical qualms doctors have about enrolling their patients in Sovaldi trials (Rao and Daugherty 2015). These moral concerns discourage doctors from recommending patients to trials. If the doctors are acting as agents for patients, this is simply a mechanism for restricting the supply of human subjects to trials.

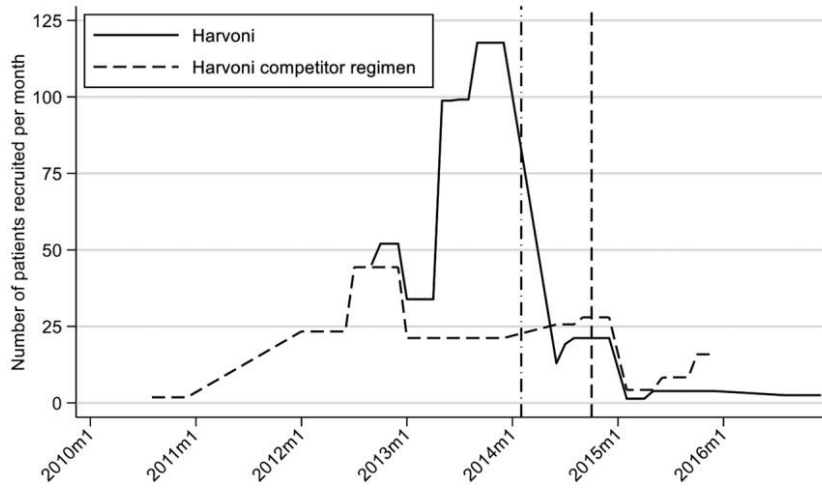

Figure 1.—Enrollment in trials of Harvoni and Harvoni competitors (for genotype 1 HCV).

fungi. HAART is a cocktail of so-called antiviral medications that effectively slows the reproduction of the HIV virus and prevents AIDS. Using longitudinal data from the Multicenter AIDS Cohort Study (MACS) on over 1,000 gay and bisexual men in four US cities, we show that HAART dramatically reduced participation in trials for other, potentially innovative antiviral medications. Our basic result is summarized by the cross tabulation in table 1, which reports the fraction of sample HIV patients participating in clinical trials before and after the introduction of HAART in 1996. In panel A, we find that, among all HIV-positive subjects, participation in trials of antiviral drugs (hereafter "primary" drugs) fell from 11.5 to 7.3 percent after introduction of HAART. Among the subset of HIV-positive patients actually using such primary drugs, the reduction was more dramatic: from 20.1 to 9.1 percent.[6] To control for unrelated changes in trial participation during the sample period, we use time trends and a control group of HIV patients who consume non-antiviral medications (hereafter "secondary" drugs).[7]

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

[7] Whereas antiviral (i.e., primary) drugs target replication of the HIV virus, aiming to prevent AIDS, non-antiviral (i.e., secondary) drugs target the secondary infections that take advantage of AIDS. In short, primary drugs address the cause of AIDS while secondary drugs tackle the symptoms of AIDS. HAART was a dramatically more effective primary drug. HIV patients often consume both primary and secondary drugs, though the trials for the two drugs are often different. While HAART reduced the number of HIV patients with AIDS

TABLE 1
Trial Participation Rate before and after HAART

	Participation in Primary-Drug Trials		Participation in Secondary-Drug Trials	
	Observations	Rate	Observations	Rate
	A. Subjects with HIV			
Pre-HAART (1990–95)	6,669	.1149	6,669	.0393
Post-HAART (1996–2005)	5,464	.0728	5,464	.0194
Difference		−.042		−.020
p-value		(.000)		(.000)
Difference-in-differences				−.022
p-value				(.001)
	B. Subjects Taking Relevant Drug			
Pre-HAART (1990–95)	3,645	.2093	2,651	.0977
Post-HAART (1996–2005)	4,366	.0912	2,495	.0425
Difference		−.118		−.055
p-value		(.000)		(.000)
Difference-in-differences				−.063
p-value				(.000)

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

An important question is whether the observed decline in trial participation was driven by a contraction in the supply of subjects, a dynamic that this paper highlights, or by the fact that HAART reduced demand for further R&D and thus for subjects. We demonstrate in three ways that supply was the more important factor. First, we document that the decline in trial participation after HAART was largely driven by exit from ongoing trials. Because ethics rules prohibit the closure of ongoing trials by drug companies, an increase in exits signals a reduction in the supply of subjects rather than one in the demand for subjects. Second, we show that funding for

and thus the overall level of demand for secondary drugs, it did not change the relative quality of secondary drugs in the output market versus those in trials and thus should not affect the rate at which patients who wanted to take secondary drugs obtained them in secondary-drug trials rather than on the market. Panel B of table 1 shows that, among HIV-positive patients, the participation rate in secondary-drug trials fell only 2 percentage points and that, among secondary-drug users, the participation rate fell 5.5 percentage points after HAART. This yields a difference-in-differences estimates of a reduction in trial participation of 2.2 percentage points among HIV-positive subjects and 6.3 percentage points among primary-drug users.

Table 1 uses a sample of HIV-positive individuals who were consuming primary (secondary) drugs in either the output market or trials. If we expand the sample (not reported) to all HIV-positive individuals, whether they consumed primary (secondary) drugs or not, trial participation in primary (secondary) drug trials fell 4.1 (2) percentage points, also highly significant. The difference in difference estimate is 2.1 percentage points ($p = .0177$).

R&D in HIV/AIDS did not diminish after the introduction of HAART. Third, we observe that, because ethics rules also cap payment of trial subjects, equilibrium quantity is dictated by subject supply rather than demand.

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

The subject supply effect also implies that health care policies that operate on output markets have dual effects on the net returns to medical innovation. Conventional analysis suggests that health care reforms that expand insurance coverage increase the demand for innovation and thus innovative returns, while policies that impose price controls reduce demand for innovation and thus innovative returns. The subject supply effect suggests that both policies also improve patients' prospects in the output market and reduce their incentive to participate in clinical trials. The resulting delay in development increases the cost of and reduces the return to innovation. Thus, the subject supply effect mitigates the positive innovation effects from insurance expansion, while it compounds the negative innovation effects from price controls.

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

The subject supply effect also has important welfare implications. As alluded to above, an important feature that distinguishes the labor mar-

[8] Finkelstein (2004) finds that vaccine promotion policies increased vaccine approvals 7 or more years later. This is not inconsistent with the development delay effect that we high-

ket for human research subjects from other labor markets is that wages are capped by ethics rules. Specifically, bioethicists—and thus institutional review boards (IRBs) that regulate trials—frown upon research compensation because it may encourage subjects to enroll in trials for nonmedical reasons.[9] Because of these ethical wage caps, we argue that observed delays in development due to the subject supply effect are likely to be inefficient. According to conventional economic analysis, improvements in output markets, such as an improvement in conventional treatment, may reduce innovation, but this is likely to be efficient because those improvements reduce the demand for innovation. Likewise, one could argue that improvements in output markets that affect innovation through the subject supply effect are also efficient because they reflect an increase in the opportunity cost to patients of enrolling in trials. However, as a result of ethical wage caps, subjects willing to participate in trials are rationed among trials by queuing rather than by price. Thus, a component of the development delay induced by the subject supply effect is the misallocation of subjects.

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

This paper builds on a prior literature on subject enrollment in clinical trials. The empirical component of this literature focuses on patients' reported incentives to enroll in trials. Most of that literature uses willingness-to-pay surveys rather than behavior, but there is some correlation between survey response and enrollment (Halpern et al. 2001; Abraham et al. 2006). These willingness-to-pay surveys show that patients weigh costs and benefits of enrollment (see, e.g., Carroll et al. 2012). There is heterogeneity in preferences by demographic features, consistent with our derivation of a supply function of trial participants (see, e.g., reviews by Ross et al. 1999 and Abraham et al. 2006). There is evidence that altruism is an important motivator for participation (Truong et al. 2011), though that is not incon-

light, because she does not examine whether trials took longer. Consistent with the demand-side analysis, Acemoglu and Linn (2004) find that larger markets, as measured by disease prevalence, are associated with greater innovation—but that conflates a positive demand for innovation and a larger supply of subjects, both of which increase innovation.

[9] As a result, IRBs 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 $2,000 (Grady et al. 2005). Indeed, surveys suggest that, in practice, IRBs do not even allow trials to fully compensate subjects for nontime expenditures (Ripley et al. 2010).

sistent with our theory, which allows things other than price and quality of treatment to affect utility. The literature also finds that research design, that is, randomization and placebo rather than active controls, tends to discourage participation (Agoritsas, Deom, and Perneger 2011). Consistent with our assumption that the wage paid in trials matters, there is a moderate literature showing that payments encourage trial participation (see, e.g., Grady 2001). Finally, some literature suggests that better information about the design, costs, and benefits of trials promotes enrollment (Van Epps, Volpp, and Halpern 2016). The feature that distinguishes our analysis from that of these papers is that we theoretically model patient incentives to enroll in trials, test those predictions to validate our model, and then draw welfare implications from our positive model.

Our theoretical modeling is a subset of the more general literature on data markets (Philipson and Malani 1999; Philipson 2001) that examines the incentives people have to supply data for research. Within this literature, our paper employs a model of patient enrollment in trials similar to that of Malani (2008), who examines how selection into medical trials affects the external validity of treatment effects estimated with data from trials, and of Philipson (1997*b*) and Chan and Hamilton (2006), who examine how attrition out of trials reveals information about the quality of treatments provided in trials. A more general version of the subject of self-selection models in these papers may be found in Chassang, Padró i Miquel, and Snowberg (2012); conversely, a less formal version is presented by Dunn and Gordon (2005). The feature that distinguishes those papers from our own is that they are concerned with implications of patient selection into trials for statistical inferences from trials. By contrast, our interest is the implications for product development and welfare. From this perspective, the papers that are closest in objective to our paper are Philipson (1997*a*) and Philipson (1997*b*), which examine optimal wages in the market for supply of data by subjects. Nevertheless, both are focused on the relationship between wages and the information generated from trials, not product innovation—the final output—or welfare.

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

ucts, including the implications for location of innovation and a reexamination of Pareto optimal sample sizes (Geoffard and Philipson 1998).

I. Development and Changes in Output Markets

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

A. *The Supply of Subjects to Clinical Trials*

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

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

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

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

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

are binding ethics regulations that prohibit monetary compensation of subjects and thus a negative price for experimental care.

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

$$\mathrm{E}_F[U(q^e, p^e)] \geq U(q, p), \tag{2}$$

where the expectation is taken over the cumulative distribution function, $F(q^e)$, describing a patient's beliefs about the quality of the experimental therapy. We assume that heterogeneity among patients—for example, different beliefs about the quality of experimental care or degrees of risk aversion—gives rise to a differentiable supply function, $s(q, p)$, that is decreasing in the quality of conventional care and increasing in its price: $s_q \leq 0$, $s_p \geq 0$.[11] This function depicts the fraction of the stock of patients willing to participate in the trial, given existing output market conditions.[12]

B. Development Times

Given a stock of patients with a disease, N, and a steady-state fraction, s, who participate in trials, we can derive the supply of subjects Ns. We assume that, because of binding ethical wage caps, clinical-trial participants are allocated evenly across existing trials.[13] For simplicity, we also assume that trials are of identical sample size and design. Given M trials and a flow of subjects, f, into each particular trial, demand for subjects is Mf, and the equilibrium condition in the subject market is that demand equal supply:

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

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

Suppose a trial seeks to recruit a sample size of n patients, assumed to be determined by considerations of statistical power. The development

[11] 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 under 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 show below. Higher quality reduces not only direct participation but also the ability to recruit from nonresponders to conventional care.

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

[13] A number of services have emerged to help match supply (patients) and demand (trials). Many of these are web based, such as the National Institutes of Health's (NIH's) https://clinicaltrials.gov database.

time or duration, T, it takes to complete this trial is the number of periods required to recruit n subjects, given a flow of patients, f, enrolling in each period:

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

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

C. *Changes in the Supply of and Demand for Subjects*

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

$$\begin{aligned} T_q &= T_s s_q > 0, \\ T_p &= T_s s_p < 0. \end{aligned} \tag{5}$$

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

Changes in conventional care can also shift the demand for subjects. While this is not the first paper to highlight this effect, our simple model can capture it and relate it to the novel subject supply effect. Because innovation competes with conventional care, demand for research is decreasing in the quality ($M_q < 0$) and increasing in the price ($M_p > 0$) of that care. The full effects of changes in conventional care on develop-

[14] A number of predictions flow from this relationship. For example, some oncology trials may take long to complete because patients die quickly and are available only for observation during a short time window. Moreover, the relationship suggests that so-called me-too innovations, for which experimental care is a close substitute for conventional care, may face recruitment difficulties and longer development times because they do not offer substantial gains over the conventional treatment outside of the trial.

ment times are ambiguous because supply and demand effects may flow in opposite directions:

$$\begin{aligned} T_q &= T_s s_q + T_M M_q, \\ T_p &= T_s s_p + T_M M_p. \end{aligned} \tag{6}$$

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

Even if improvements in conventional care do not induce changes in the demand for R&D, they might induce investigators, who are barred by ethics rules from raising subject wages, to modify the design of experiments to make them more attractive to patients. Because these are second-order effects, we explore them only in the appendix (available online). Likewise, even if the demand for R&D does not change, changes in the output market may also yield changes in regulatory requirements, such as the requirement for greater statistical confidence in efficacy.[15] This, in turn, could increase the sample size n required in future trials, thus increasing development times.

II. Evidence of the Subject Supply Effect for AIDS Innovations

In this section, we examine empirical evidence for the subject supply effect in the case of AIDS treatment. We employ the introduction of HAART—a breakthrough AIDS treatment—in 1996 as a change in conventional treatment. Our aim is to identify, or at least sign, the effect of the implied change in output quality and price on the total supply of subjects in HIV/AIDS clinical trials, that is, to validate the predictions $\partial(Ns)/\partial q < 0$ and $\partial(Ns)/\partial p > 0$ and the participation rates, that is, validate the predictions $s_q < 0$ and $s_p > 0$. We do this in three steps.

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

[15] For example, a drug regulator may require that future treatments be proven effective relative not to no treatment but to the conventional treatment (O'Connor 2010). If conventional treatment has more variable outcomes relative to no treatment, then even the sample size required for a given level of confidence will rise.

the price effect of HAART offsets the reduction in participation due to the quality effect of HAART, we likely underestimate the purely quality-driven effects of HAART on subject supply.[16]

A. Background on HIV/AIDS and HAART

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

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

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

HAART is the label for a regimen of antiviral medications that, together, substantially slow the progression of HIV into AIDS. This regimen typically includes two nucleoside reverse transcriptase inhibitors (NRTIs) and either one protease inhibitor (PI) or one non-nucleoside reverse transcriptase inhibitor (NNRTI). HAART was first introduced to the market in 1996. Although the first NRTI—zidovudine (AZT)—was approved by the FDA in 1987, NRTIs on their own were unable to affect progression from HIV to AIDS. In December 1995, the FDA approved the first PI, and in July 1996, scientists demonstrated that the combination of NRTIs and a PI was highly effective at controlling AIDS progression (Gulick et al. 1997; Hammer et al. 1997; Bartlett 2006). Later in 1996, the FDA approved a

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

[17] The 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.

new class of antivirals, the NNRTIs. These also were shown to be highly effective at controlling the HIV virus when used with NRTIs. Together, these combinations of primary AIDS drugs, which quickly acquired the name HAART, reduced AIDS deaths by a half before the end of the decade (Palella et al. 1998; Moore and Chaisson 1999). In 1997, HAART was pronounced the standard of care by the US panel of the International AIDS Society, the guideline-setting body for HIV treatment (Carpenter et al. 1997).

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

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

B. Data

Our empirical analysis employs data from the MACS. Given the relatively high prevalence of HIV in the gay community, this study tracked 6,972

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

homosexual and bisexual men in four cities (Baltimore, Chicago, Pittsburgh, and Los Angeles) longitudinally during the period 1984–2005.[19] We first isolate a subsample of 1,725 unique subjects with HIV observed in the period 1990–2005.[20] This is a subsample that was observed both before and after the introduction of HAART in 1996 and for which we have reliable data on HIV-positive status and trial participation.

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

We conduct our empirical analysis on either the entire sample of HIV-positive patients or a subset of patients who used HIV drugs. In the HIV-positive sample, we include only individuals in the years after they were diagnosed with HIV. In the HIV drug sample, we include only individuals in the years they report using a primary drug or a secondary drug, as appropriate. Among HIV-positive patients, 1,435 individuals took primary drugs and 1,267 took secondary drugs at some point during the sample period, whether the source of drug supply was inside or outside a trial. Of all HIV-positive individuals, 216 took only primary drugs, 48 took only secondary drugs, and 1,219 took both a primary drug and a secondary drug at some point.

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

C. Descriptive Statistics

Table 2 provides descriptive statistics for all subjects with HIV. Statistics are calculated at the subject-year level and weighted so that each subject has an identical weight. The data are separately described for the years before HAART (1990–95) and the years after HAART (1997–2005) to provide an unconditional assessment of the effect of HAART. Roughly the same fraction—roughly 75 percent—of subjects has some medical

[19] More information on the MACS study is available at http://www.statepi.jhsph.edu/macs/macs.html. Not all 6,972 subjects were observed each year. Instead, subjects were enrolled in three large waves: in 1984–85, 1987–91, and 2001–03. Only subjects enrolled in the first two waves were observed before HAART. Moreover, half of the confirmed HIV-negative subjects were administratively censored in 1993. We omitted these subjects because they were not observed after HAART.

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

TABLE 2
SUMMARY STATISTICS

Category, Variable	Unit	Pre-HAART (1990–95)			Post-HAART (1997–2005)		
		Observations	Mean	Standard Deviation	Observations	Mean	Standard Deviation
AIDS drug use:							
Primary treatment	0/1	9,483	.442	.497	10,269	.644	.479
Secondary treatment	0/1	9,483	.326	.469	10,269	.313	.464
Trial participation:							
Primary-drug trial	0/1	9,483	.092	.288	10,269	.051	.22
Placebo controlled	0/1	9,483	.041	.199	10,269	.009	.096
Secondary-drug trial	0/1	9,483	.032	.175	10,269	.011	.102
Employment:							
Full time	0/1	7,817	.748	.434	8,369	.574	.494
Part time	0/1	7,817	.117	.321	8,369	.135	.342
Income:							
<$10,000	0/1	7,703	.142	.349	9,272	.206	.404
$10,000–$19,999	0/1	7,703	.173	.378	9,272	.16	.367
$20,000–$29,999	0/1	7,703	.213	.41	9,272	.14	.347
$30,000–$39,999	0/1	7,703	.213	.41	9,272	.127	.333
$40,000–$49,999	0/1	7,703	.145	.352	9,272	.126	.331
$50,000–$59,999	0/1	7,703	.232	.422	9,272	.207	.405

$60,000–$69,999	0/1	7,703	.007	.086	9,272	.133	.34
>$70,000	0/1	7,703	.015	.122	9,272	.004	.062
Refused to answer	0/1	7,703	.105	.306	9,272	.079	.269
Insurance:							
Any	0/1	7,907	.758	.429	9,233	.737	.44
Government	0/1	7,801	.136	.343	9,438	.325	.469
Medicare	0/1	7,795	.042	.202	9,209	.17	.375
Medicaid	0/1	7,798	.081	.273	9,208	.171	.377
VA coverage	0/1	7,795	.038	.192	9,208	.027	.162
Private	0/1	7,807	.665	.472	9,219	.487	.5
Individual	0/1	7,084	.146	.353	9,215	.092	.288
Group	0/1	7,092	.531	.499	9,219	.431	.495
None	0/1	7,907	.358	.479	9,233	.36	.48
Health:							
Age	years	7,856	40	7	8,336	45	8
CD4 count	cells/mm^3	7,718	466	357	8,368	563	315
Viral load	copies/mL	3,294	114,547	313,340	7,479	29,121	146,625

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

insurance before and after HAART. However, there is a substantial change in the sources of insurance after HAART. While private insurance coverage falls from 67 percent to 49 percent after HAART, government coverage rises from 14 percent to 33 percent after HAART. An important program for government insurance after HAART is, surprisingly, Medicare. This is because Medicare covers not just the elderly but also the long-term disabled population. Our sample has relatively low income. Only 57 percent are employed after HAART, and the median income is between $20,000 and $29,000.

D. Evidence on the Unconditional Impact of HAART

Before we present our regression results, we offer some basic graphical evidence of the effect of HAART on the quality of conventional treatment and on trial participation. Figure 2 documents the changing pattern of primary AIDS drug use over time. Figure 2*A* gives a breakdown of primary-drug use by class of antiviral drugs. While use of NRTIs such as AZT was constant, at nearly 100 percent, throughout the 1990s, there was a sharp rise in use of PIs and NNRTIs after they were introduced in December 1995 and the end of 1996, respectively. At the same time, we see a sharp drop in use of other classes of antiviral drugs.[21] For context, table 2 notes that the fraction of HIV-positive subjects using primary drugs increased from 44 percent to 64 percent after HAART, while the fraction of these subjects using secondary drugs was constant, at roughly 32 percent, over time. Figure 2*B* gives a breakdown of how the usage of primary-drug regimens or combinations changed. Note that, although an individual may be taking multiple primary drugs in figure 2*A*, he can be treated with only one of the three primary-drug regimens (NRTI only, HAART, and other non-HAART regimen) in figure 2*B*. The main takeaway is that HAART use skyrocketed after 1996, while the use of other regimens fell.[22]

The dramatic rise in usage of HAART after 1996 led to a dramatic improvement in subjects' health. According to table 2, after HAART the average CD4 count in our sample of HIV positive subjects increased from 466 to 563/mm^3. A higher CD4 count implies better health: an HIV patient with a CD4 count below 200 is considered to have AIDS. Moreover, average viral loads fell from roughly 114,547 to 29,121 copies/mL, and a patient is considered to have AIDS with a load above 100,000 copies/mL. Figure 3, which plots CD4 counts and viral loads for subjects taking primary drugs, provides even more detail on the trend. According to figure 3*A*, CD4 counts for the median subject taking primary drugs declined steadily

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

[22] Although the use of NRTI-only regimens had been falling since 1991, much of the decline before 1996 was offset by growing use of non-HAART cocktails on the market. These regimens typically combined an NRTI with one of the "other" antiviral drugs depicted in fig. 2*A*.

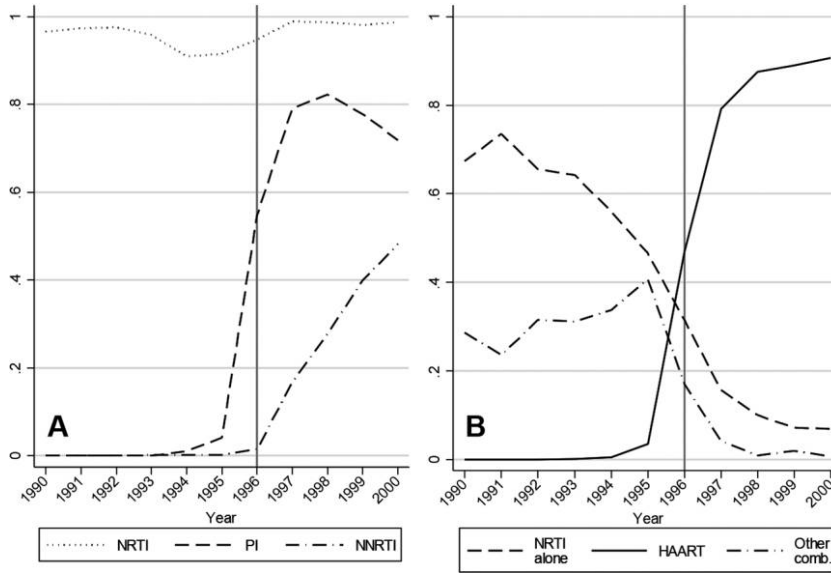

Figure 2.—Effect of HAART on choice of AIDS therapy.

until 1995, right before the introduction of HAART, after which they improved dramatically. We illustrate the effects for AIDS progression by plotting the CD4 counts for the 25th-percentile subject. Until 1995, this subject had AIDS (CD4 count < 200/mm^3), but after HAART, their CD4 counts improved to the point where, by 1997, the 25th-percentile subject

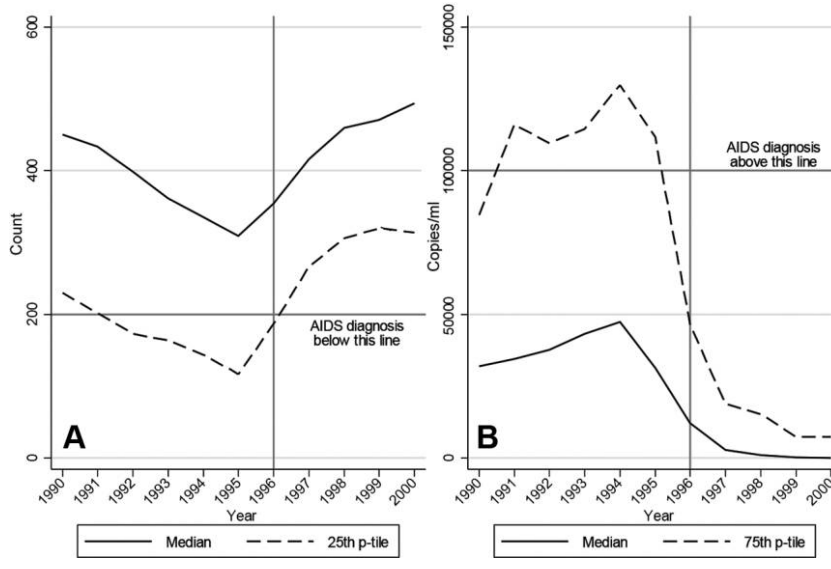

Figure 3.—Effect of HAART on CD4 counts; p-tile: percentile.

did not have AIDS. Likewise, figure 3*B* shows that viral loads for the median patient fell dramatically, starting in 1996. The effect was even more pronounced for the 75th-percentile subject in our sample. Until 1995, this individual had loads above the AIDS threshold. Within a year after HAART, his load fell so that he was no longer at risk.

Finally, we turn to trial participation. According to figure 4*A* (right *Y*-axis), the total number of subjects in trials hovered around 120 until 1995. After HAART, the total fell dramatically to around 70 in 1997 and

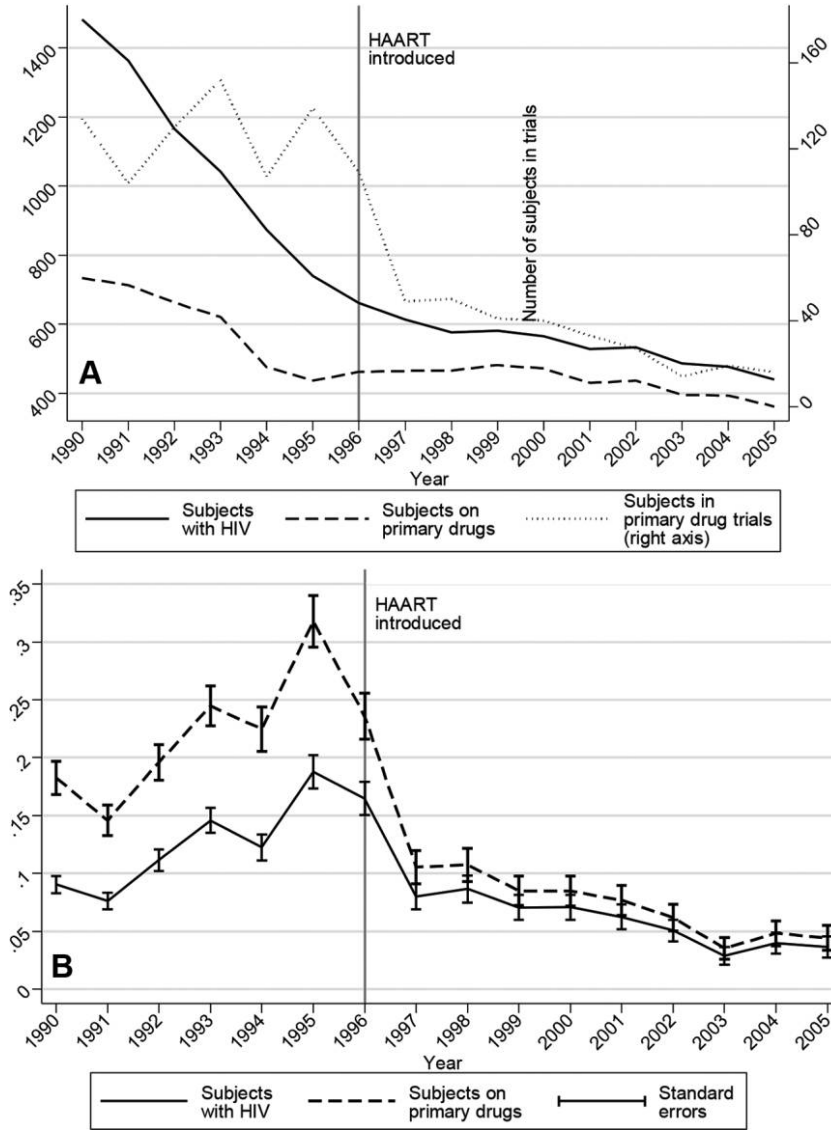

Figure 4.—Effect of HAART on trial participation.

trickled down to 50 by 2005. These totals correspond to the equilibrium number of subjects in trials, $Mf = Ns$. We can also look at equilibrium participation rates, that is, $Mf/N = s$. According to table 2, 9.2 percent of HIV-positive subjects were in clinical trials for primary AIDS drugs before HAART. However, after HAART was introduced, only 5.1 percent were participating in trials.[23] Figure 4*B* illustrates that, behind these changes, there is an even more dramatic change in the trend in trial participation over time. Trial participation increased to nearly 20 percent among HIV-positive subjects in 1995, and then trial participation plummeted and eventually settled at 5 percent by 2005, roughly half the level in 1990.[24] The change was even more dramatic among primary-drug users, where the participation rose to over 30 percent by 1995 and fell to a quarter of the 1990 rate by 2005.

We can use figure 4 to unpack the causes of the observed changes in trial participation rates. From figure 4*A* we can rule out an increase in demand for research subjects, *Mf*, as the cause of the dramatic rise in trial participation rates before 1996. An increase in demand should increase total subjects in trials, but that was steady before 1996. As between the alternative explanations, an increase in the supply rate *s* or a decrease in the number of subjects with HIV, figure 4*B* favors the latter: the number of subjects with HIV fell from over 1,400 to just 800 by 1996. The drop is largely due to high AIDS mortality rates before 1996. While primary-drug utilization rates—a precursor to participation in trials—also fell, the fall was less dramatic, from 700 to 500, largely as a result of evidence emerging that AZT, the foremost primary drug before 1996, was ineffective at controlling HIV.

The fall in the trial participation rate after HAART could be due to either a reduction in aggregate demand for subjects, *Mf*, or a reduction in the supply rate, *s*, of subjects. It cannot be attributed to an increase in the number of subjects with HIV or using primary drugs, *N*. To rule out changes in aggregate demand, figure 5 examines the rates of entry and exit into trials. Entry (exit) rates are calculated by taking the number of subjects who enter (exit) trials in a given period and dividing by the number of subjects not in (in) trials in the previous period. A reduction in demand can take only the form of a decrease in the number of new trials started, because ethics rules prevent researchers from closing down ongoing trials and kicking out or "firing" enrolled subjects. Thus, changes in demand will be manifested only as decreases in entry. By contrast, a decrease in the supply rate can cause subjects either to exit trials or to enter trials at a lower rate. Figure 5 shows a sharp (short-term) spike in exits

[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 percent rate of trial participation for patients with diseases other than HIV (e.g., Lara et al. 2001).

[24] The pre-HAART average participation rate in table 2 is somewhat lower than the pre-HAART average in fig. 4 because the two use different weighting schemes. The former gives each subject equal weight.

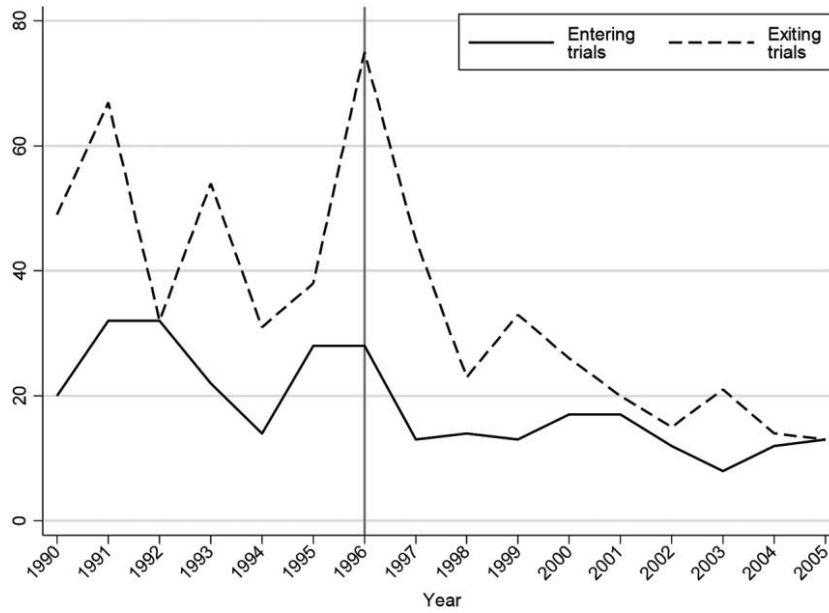

Figure 5.—Effect of HAART on trial entry and exit.

in 1996 but only a mild reduction in entrances after HAART. Thus, it is more likely that the decline in the equilibrium trial participation rate immediately after HAART is due to a reduction in the supply rate of subjects, consistent with the prediction of our theory.

E. Regression Analysis

1. Empirical Strategy

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

$$P_t = \beta_1 + \beta_2 \text{POST\_HAART}_t + \beta_3 \text{TREND}_t + u_t. \quad (7)$$

The dependent variable P_t is calculated by summing the total number of subjects reporting that they participated in a primary-drug trial in year t. The treatment effect is captured by the coefficient on POST_HAART_t, an indicator for the period 1996–2005 or, in some specifications, 1997–2005. In the latter case, we drop the year 1996 from the sample. From figure 4*A*, we see that the total number of subjects in trials is more variable before 1996 than after; therefore, we calculate robust standard errors when conducting the regression above. Standard errors are unbiased only if the regression errors are not serially correlated; we report the p-value from Durbin's alternative test to check this condition.

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

$$p_{it} = \beta_1 \text{PRE\_HAART}_t + \beta_2 \text{POST\_HAART}_t + \beta_3 \text{TREND}_t + [\text{controls}] + e_{it}. \quad (8)$$

The dependent variable is an indicator variable p_{it} for whether subject i in year t participated in a clinical trial. The treatment effect is captured by the difference between the coefficient on an indicator for some *pre-window* before the introduction of HAART and the coefficient on an indicator for some *post-window* after the introduction of HAART. We employ three specifications of the pre- and post-windows to capture short- and long-term changes in participation rates. One specification, which we label the "short run," uses a pre-window of just 1995 and a post-window of just 1997. A second, which we label the "medium run," uses a pre-window of 1993–95 and a post-window of 1997–99. The last, which we label the "long run," uses a pre-window of 1990–95 and a post-window of 1997–2005. In all but one regression, we omit 1996 from the pre- and post-windows, though we do not drop that year from the data.[25] The one exception is our analysis of trial exits. There we focus on 1996, because that is the year in which we observe a spike in exits in figure 5.[26] We include a linear trend in the regression above to capture unobserved trends in participation. Thus, the treatment effect is, to some extent, identified off a linear trend.[27]

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

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

[27] Although the regression analysis reported in this paper focuses on the level of trial participation, much of the effect of HAART was manifested through changes in the trend of participation. The level of trial participation depends on subjects' knowledge of the quality of experimental treatments and newly approved treatments. However, it may take time for subjects to learn this in the case of HAART. Although trials of PIs and NNRTIs began in the late 1980s and early 1990s, it took a few years before patients learned about the value of those drugs. Likewise, even though the first PI was approved in December 1995, it took a number of years to gather decisive evidence of the benefit of drug cocktails. For example, the first findings demonstrating HAART's effect on CD4 counts, a biomarker, were published in 1997 (Gulick et al. 1997; Hammer et al. 1997). Evidence of the reduction in death rates, the key end point, did not emerge until the end of the decade (Moore and Chaisson 1999). This may explain the sharp reversal in trial participation trends after 1996 in fig. 2*B*. In a previous version of this paper, we included regressions in which the pre- and post-windows were interacted with the trend term:

$$p_{it} = \beta_1 \text{PRE\_HAART\_TREND}_t + \beta_2 \text{POST\_HAART\_TREND}_t + [\text{controls}] + e_{it}.$$

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

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

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

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

2. Effect of HAART on Equilibrium Participation

Our initial results concerning the equilibrium quantity of subjects in primary-drug trials are presented in panel A of table 3. (We discuss panels B–D, which examine specifically placebo- or active-controlled trials, along with fig. 8, in the appendix.) We find that the number of subjects in trials fell by roughly 90 from a base of 128, that is, a drop of roughly three-quarters. We obtain a slightly larger estimate when we drop 1996, our transition year, from the sample. When we add linear drift to the regression, the baseline rises to 143 participants and the HAART-induced reduction falls to 67 participants, that is, a reduction of roughly one-half. In all specifications, we cannot reject that the errors are not serially correlated, suggesting that our OLS estimates are adequate.

Table 4 presents our basic results on the equilibrium rate of individual participation in primary-drug trials. Panel A reports results for regression

TABLE 3
EFFECT OF HAART ON NUMBER OF INDIVIDUALS ENROLLED IN PRIMARY-DRUG TRIALS

	1996 Included		1996 Not Included	
	(1)	(2)	(3)	(4)
	A. Total Subjects Enrolled in Primary-Drug Trials			
HAART	−87.9***	−39.1	−95.6***	−66.6***
	(−11.6)	(−25.7)	(−8.8)	(−13.9)
Trend		−6.1**		−3.4**
		(−2.4)		(−1.2)
Constant	127.7***	155.1***	127.7***	143.0***
	(−7.5)	(−14.1)	(−7.5)	(−9.6)
Observations	16	16	15	15
Durbin's alternative (*p*-value)	.18	.69	.84	.21
	B. Total Subjects Enrolled in Active-Controlled Primary-Drug Trials			
HAART	−54.4***	−25.2	−60.0***	−47.8**
	(−13.4)	(−24.1)	(−12.3)	(−18.8)
Trend		−3.6		−1.4
		(−2.3)		(−1.9)
Constant	89.7***	106.1***	89.7***	96.1***
	(−11.5)	(−17.7)	(−11.6)	(−16.4)
Observations	16	16	15	15
Durbin's alternative (*p*-value)	.22	.36	.73	.83
	C. Total Subjects Enrolled in Placebo-Controlled Primary-Drug Trials			
HAART	−49.9***	−14.8	−53.1***	−21.1*
	(9.6)	(10.1)	(9.1)	(10.6)
Trend		−4.4***		−3.8**
		(1.2)		(1.4)
Constant	58.7***	78.4***	58.7***	75.6***
	(8.9)	(9.6)	(9.0)	(10.8)
Observations	16	16	15	15
Durbin's alternative (*p*-value)	.26	.02	.00	.00
	D. Total Subjects Enrolled in Placebo-Controlled Primary-Drug Trials			
Lagged subjects	.7***	.6***	.6***	.6***
	(.1)	(.1)	(.1)	(.1)
HAART	−8.9	−8.4	−15.8***	−16.5***
	(7.6)	(8.4)	(4.6)	(4.6)
Trend		−.3		.3
		(.6)		(.4)
Constant	9.2	13.0*	15.9***	12.4
	(7.5)	(6.5)	(4.8)	(7.2)
Observations	15	15	14	14
Durbin's alternative (*p*-value)	.42	.46	.84	.81

Note.—Dependent variable is total subjects in primary-drug trials indicated at top of each panel. Regression in panel D includes first lag of dependent variable as a control. Regressions in cols. 3 and 4 omit the transition year 1996. Robust standard errors are in parentheses. Durbin's alternative test checks whether the null hypothesis of no serial correlation in residuals can be rejected.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

TABLE 4
EFFECT OF HAART ON INDIVIDUAL TRIAL PARTICIPATION RATE: IDENTIFICATION OF TIME TREND

	Specification (See Sec. II.E.1)					
	1	2	3	4	4	4
	A. HIV Patients					
Trend	−.003***	−.001	−.010	−.064*	−.054*	−.044
	(.000)	(−.001)	(−.009)	(−.033)	(−.033)	(−.035)
Pre-HAART	.081***	.086***	.064***	.650***	.644***	
	(−.008)	(−.016)	(−.017)	(−.105)	(−.072)	
Post-HAART	−.012	−.004	−.011	−.161	−.030	−.263**
	(−.009)	(−.011)	(−.012)	(−.158)	(−.108)	(−.128)
Treatment effect	−.093	−.089	−.075	−.811	−.674	−.263
Wald test (p)	.000	.000	.000	.000	.000	.040
Observations	12,133	12,133	10,862	4,187	4,187	4,187
	B. Primary-Drug Users					
Trend	−.010***	−.012***	−.025**	−.221***	−.216***	−.155***
	(.001)	(.002)	(.011)	(.040)	(.041)	(.042)
Pre-HAART	.135***	.147***	.119***	.895***	.857***	
	(.012)	(.026)	(.026)	(.119)	(.081)	
Post-HAART	−.037***	−.037***	−.042***	−.438**	−.288**	−.780***
	(.013)	(.014)	(.015)	(.171)	(.120)	(.145)

Treatment effect	−.171	−.184	−.160	−1.333	−1.145	−.780
Wald test (p)	.000	.000	.000	.000	.000	.000
Observations	8,011	8,011	7,421	3,377	3,377	3,377
	C. Details on Specifications					
Pre-window	1995	1995	1995	1995	1993–95	1990–95
Post-window	1997	1997	1997	1997	1997–99	1997–2005
Identification	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend
Individual fixed effects	No	Yes	Yes	Yes	Yes	Yes
Additional covariates	No	No	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	Logit	Logit	Logit

Note.—Dependent variable is whether a subject participated in a primary-drug trial. Sample includes the years 1990–95 and 1997–2005 and individuals indicated at top of each panel. Observations are at the individual-year level. Pre-window and post-window variables are dummies for the years indicated in panel C. Other controls, as well as estimation method—OLS or logit—are indicated in panel C. Additional covariates include binned income and CD4 count. Coefficients identify change in trial participation rate relative to a linear trend. Standard errors, clustered at the individual level, are reported in parentheses. Treatment effect is the coefficient on post-HAART minus the coefficient on pre-HAART in the first five columns; in the last column, it is simply the coefficient on post-HAART. The last row of panels A and B reports the p-value for a Wald test of whether the treatment effect is zero.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

samples including all HIV patients. Panel B reports results for samples including only subjects taking primary drugs. Panel C describes the specification of each regression, though the number at the top of each column is a quick reference to the nature of controls and estimation method employed. The first four columns of the table report short-run specifications with one-year treatment and control windows, the next column reports medium-run specifications with three-year windows, and the last column reports long-run specifications. At the bottom of panels A and B, we report treatment effects (usually post-window minus pre-window coefficients) and the *p*-value from a Wald test that the difference is zero. We use this template for many of the remaining tables (for a discussion of table 5, see the appendix).

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

3. Robustness of Equilibrium Participation Results

A concern with simple pre-window versus post-window 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 as a result of trial participation. If this reduced participation, we may overestimate the reduction in participation after HAART.

One way we address is this to include a linear time trend as a control in our basic regressions. An alternative approach, which we report in the appendix, is to employ subjects using secondary AIDS drugs as a control group. If there are secular changes in trial participation incentives, they should affect individuals' willingness to participate in both primary AIDS drug trials and secondary AIDS drug trials. If we assume identical secular changes in both drug categories, then we can employ changes in participation rates in secondary-drug trials as a control for secular changes in participation rates in primary-drug trials. We report tabular results using this identification strategy in note 8 of the appendix. We explain the identification strategy more fully and report regression results in the appendix.

F. Explaining the Observed Changes in the Participation Rate

The previous subsections identified the reduced-form relationship between HAART and trial participation to illustrate a robust relationship between innovation and equilibrium trial participation. We now examine what the change in equilibrium participation reveals about changes in

the supply of research subjects, the mechanism behind the subject supply effect. In the appendix, we further try to decompose the effect of HAART into a positive quality effect, which should increase supply, and a positive price effect, which should decrease supply, and we examine the effect of HAART on the design of trials, an alternative way for researchers to compete for subjects when ethics rules cap monetary wages.

Recall from equation (3) that the equilibrium participation satisfies $Mf = Ns$ or, in rates, $Mf/N = s$, where M measures demand, f reflects pro rata allocation of subjects across trials, N is the number of eligible subjects (either HIV-positive individuals or primary-drug users), and s is the supply function for individuals. For a number of reasons, we think that the relationship between HAART and the equilibrium trial participation rate is driven by changes in the supply of subjects, s, and not merely changes in demand for subjects, that is, the number of trials M.[28]

First, ethics rules cap monetary wages in the market for human research subjects. As a result, there is likely to be excess demand for subjects in each period. Because quantity is the minimum of supply and demand when prices are capped, quantity will be set by supply. In equilibrium, the supply of subjects is rationed across trials by queuing, which is reflected in f in our model. Therefore, our estimates of the effect of HAART on the per-period trial participation rate identify the supply behavior of subjects. Even if the introduction of HAART reduced demand by decreasing trials, M, we predict that it would not cause the observed decline in trial participation per period.[29] In other words, a sufficient assumption to identify the effect of HAART on supply is that the equilibrium wage is determined by a binding cap.[30]

Second, we investigate whether HAART affected exits from existing clinical trials, a response that is likely to be driven only by the supply rate. In theory, patients may have exited trials after HAART because they no longer wanted to participate—reduced supply—or because drug companies canceled trials—as a result of reduced demand for innovation.[31] In standard labor markets, "quits" versus "fires" are hard to distinguish empirically, but this is less of an issue for clinical-trial recruitment. In practice, we can rule out "fires" because it is standard ethical practice not to cancel ongoing trials unless the treatment drug in that trial clearly works or the experimental drug has a severe side effect (Psaty and Rennie 2003;

[28] We can rule out that the decline in participation is due to an increase in the number of eligible patients N because fig. 2*A* shows that the number of patients with HIV or taking primary drugs was roughly the same after 1996 as in 1995.

[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] Consistent with this view, and beyond the evidence presented in the introduction, Tam-McDevitt et al. (2007) finds that demand for subjects exceeds supply, at least in oncology.

[31] 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.

TABLE 5
EFFECT OF HAART ON TRIAL PARTICIPATION RATE: ANALYSIS OF SUBSAMPLES MORE ROBUST TO ATTRITION

	Specification (See Sec. II.E.1)								
	3	3	3	3	3	3	3	3	3
	A. HIV Patients								
Trend	−.011	−.011	−.013*	−.010	−.009	−.010	−.028***	−.026***	−.024***
	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)	(.007)	(.006)	(.007)
Pre-HAART	.063***	.023***		.081***	.041***		.164***	.096***	
	(.017)	(.008)		(.015)	(.010)		(.026)	(.018)	
Post-HAART	.017	.013	.022	−.010	.000	.002	−.036**	−.021*	−.061***
	(.015)	(.011)	(.017)	(.012)	(.009)	(.014)	(.016)	(.012)	(.021)
Treatment effect	−.045	−.009	.022	−.091	−.041	.002	−.200	−.117	−.061
Wald test (p)	.044	.488	.178	.000	.001	.910	.000	.000	.004
Observations	5,995	5,995	5,995	7,996	7,996	7,996	5,478	5,478	5,478
	B. Primary-Drug Users								
Trend	−.027***	−.026***	−.024***	−.011	−.010	−.011	−.031***	−.029***	−.027***
	(.007)	(.007)	(.007)	(.008)	(.008)	(.008)	(.006)	(.006)	(.006)
Pre-HAART	.156***	.069***		.083***	.041***		.174***	.096***	
	(.033)	(.019)		(.019)	(.011)		(.032)	(.021)	
Post-HAART	−.012	−.016	−.051**	−.015	.004	.001	−.038**	−.016	−.063**
	(.020)	(.015)	(.026)	(.014)	(.011)	(.017)	(.018)	(.014)	(.025)

Treatment effect	−.167	−.085	−.051	−.098	−.038	.001	−.213	−.112	−.063
Wald test (p)	.000	.000	.048	.000	.010	.966	.000	.000	.011
Observations	3,603	3,603	3,603	5,916	5,916	5,916	4,077	4,077	4,077
	C. Details on Specifications								
Sample	CD4 always ≥ 200	CD4 always ≥ 200	CD4 always ≥ 200	Survived until 2000	Survived until 2000	Survived until 2000	Survived until 2005	Survived until 2005	Survived until 2005
Pre-window	1995	1993–95	1990–95	1995	1993–95	1990–95	1995	1993–95	1990–95
Post-window	1997	1997–99	1997–2005	1997	1997–99	1997–2005	1997	1997–99	1997–2005
Identification	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend	Linear trend
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Note.—Dependent variable is whether a subject participated in a primary-drug trial. Sample includes years 1990–2005, excluding 1996. Sample includes individuals who have HIV or use a primary drug, as indicated at top of each panel, and who survived until the year 2000 or 2005 or had a CD4 count above 200 cells/mm^3 for the entire sample period, as indicated in panel C. Observations are at the individual-year level. Pre-window and post-window variables are dummies for the years indicated in panel C. Other controls, as well as estimation method (OLS or logit), are indicated in panel C. Additional covariates include binned income and CD4 count. Coefficients identify change in trial participation rate relative to a linear trend. Standard errors, clustered at the individual level, are reported in parentheses. Treatment effect is the coefficient on post-HAART minus the coefficient on pre-HAART in the first eight columns; in the last column, it is simply the coefficient on post-HAART. The last row of panels A and B reports the p-value for Wald test of whether the treatment effect is zero.

* $p < .10$

** $p < .05$.

*** $p < .01$.

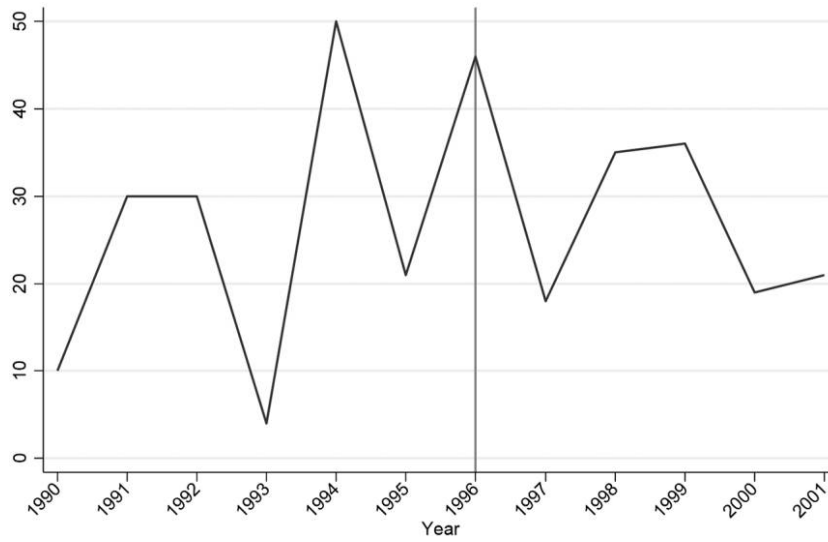

Figure 6.—Termination of ACTG trials over time.

Cannistra 2004).[32] Ethics rules do not permit cancellation because some drug outside the trial, in our case HAART, works better. Empirical support for this view comes from figure 6, which plots data on trial terminations from the AIDS Clinical Trials Group (ACTG), the main organizing body for HIV/AIDS trials during the period. According to figure 6, while the number of trials terminated in 1996 was higher than average, it was less than the number terminated in 1994, before HAART. Moreover, average annual trial terminations in the years after 1996 did not exceed those before 1996. If reduced demand for innovation after HAART was not causing trial terminations, then the exit rate from trials is a reasonable measure of the supply side of the trial participant market.

Raw data on exit rates after HAART can be seen in figure 5: there was a sharp increase in the exit rate in 1996, after which exits declined gradually, just as they had before HAART. The regression analyses reported in table 6 confirm this finding. Observations are at the year level, and the dependent variable is the number of exits (cols. 1–2) or entrances (cols. 3–4) in a year. The results suggest that exits increased by 80 percent (36 subjects) in 1996 and then actually fell by 45 percent (19 subjects) after 1996, relative to the pre-HAART period. While the decline in exits vanishes once we insert a linear-drift term, it is clear there is an isolated spike in exits in 1996, which is consistent only with a supply effect. By contrast, there is no significant change in entrances in 1996. After 1996, entrances fall by 45 percent (11 subjects), or 20 percent (6 subjects) net of trend. This last effect is consistent with either a supply or a demand effect.

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

TABLE 6
Effect of HAART on Supply: Analysis of Exits and Entrances

	Exits		Entrances	
Dependent Variable	(1)	(2)	(3)	(4)
Year 1996	36.0***	47.7***	5.3	7.5
	(−5.4)	(−5.2)	(−4)	(−5.2)
Years 1997–2005	−19.4***	8.9	−10.9**	−5.7
	(−6.4)	(−9.1)	(−4.2)	(−7.6)
Trend		−3.3***		−.6
		(−.8)		−.6
Constant	43.0***	51.3***	24.7***	26.2***
	(−5.4)	(−5.2)	(−4)	(−4.1)
Observations	16	16	16	16
Durbin's alternative (*p*-value)	.79	0	.98	.7

Note.—Dependent variable in first (last) two columns is total subjects exiting from (entering into) primary-drug trials. Sample includes years 1990–2005. Constant picks up exits or entrances in the years 1990–95. Robust standard errors are in parentheses.
** $p < .05$.
*** $p < .01$.

Recall that 1997 saw the introduction not only of the HAART drug regimen but also of real-time viral-load monitoring as the standard of care. We can use data on exits and entrances to distinguish between the effects of these two changes. First, whereas HAART would induce exits from ongoing trials, the introduction of viral-load monitoring would not affect existing trials because the FDA and good trial practice do not allow researchers to change trial protocols midstream; such changes make treatment effects before and after the protocol change less comparable. Thus, the 1996 spike in exits, which affected only ongoing trials, can only be the effect of the introduction of HAART that year.[33] Second, the introduction of viral-load monitoring in trials should increase entrances as well as exits from new trials after 1996. With real-time viral-load monitoring, enrolled subjects can exit as soon as they judge that their group assignment—whether treatment or control—is not reducing their viral load as desired. Moreover, a more effective option to exit should make subjects more willing to enroll in trials. Yet we find a reduction in exits in the long run, unless we include a trend in the exit regression. Moreover, we find a reduction rather than an increase in entrances after 1996. These two findings lead us to conclude that viral-load monitoring cannot explain the changes we see in trial participation from 1996 onward.

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 7 plots NIH spending on R&D for HIV/AIDS between 1995 and 2010. It shows that federal spending continued to rise

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

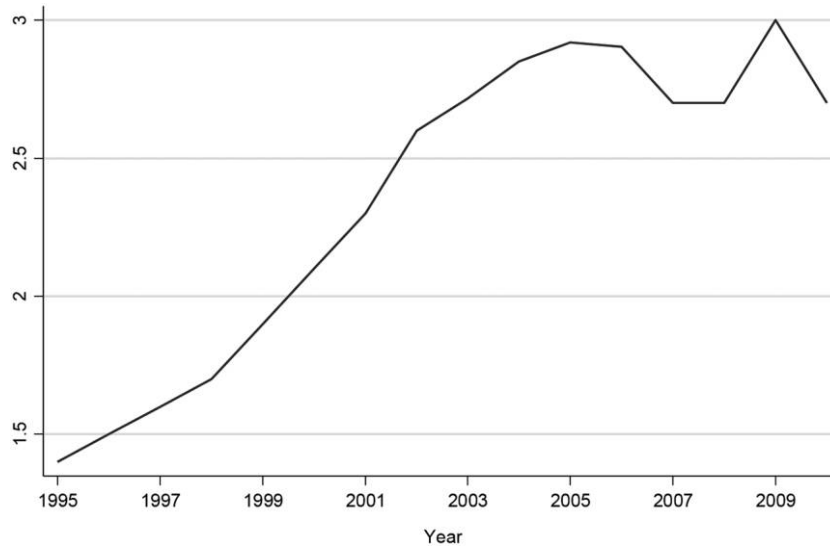

Figure 7.—Federal government (NIH) spending on HIV/AIDS research.

dramatically even after the introduction of HAART. Indeed, the data suggest 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 sponsored 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 and Sandberg 1998). By 2001, the number had fallen to 98 (PhRMA 2002), and it stayed constant at 100 until 2010 (PhRMA 2010).

III. Positive Implications of the Subject Supply Effect

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

A. Innovative Returns

We assume that the innovation has a finite period, T^{P}, during which it is under patent. After this period, the market becomes competitive and profits fall to zero. Let $A(x)$ be the value of a standard annuity paying \$1 for

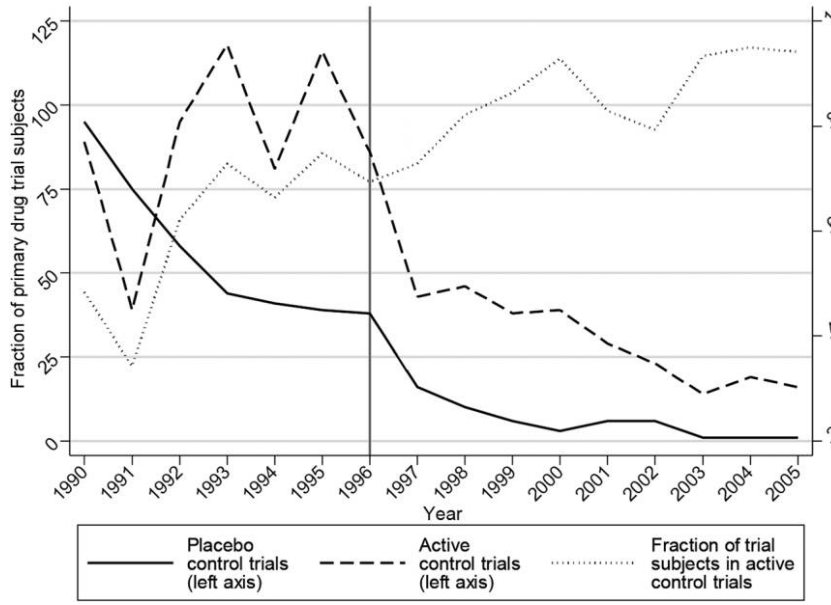

Figure 8.—Enrollment in placebo- and active-control trials.

x periods. The net present value of the overall innovative return can be written as

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

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

Delays in development have two negative effects on the present value of innovative returns. First, they delay when the firm begins earning positive variable profits. This effect operates through the term β^T. Second, delays in development eat up patent life, that is, the period during which firms have exclusive rights to sell their new product, reducing the number of periods the firm earns positive profits. This effect operates through the term $A(T^{\mathrm{P}} - T)$. Prior research suggests that the cost of development delays can be economically significant. For example, DiMasi et al. (2003) find that financial opportunity costs attributable to the duration of clinical trials account for half the total cost of clinical research.[34] Even that is likely to be an underestimate, because it does not address the effect of trial duration on remaining patent life.

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

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

$$\frac{d\text{NPV}}{dx} = V\Pi_x + V_T T_x \Pi, \tag{10}$$

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

For example, an improvement in the quality of conventional care reduces the nominal per-period patent profits a firm can expect from a future innovation. The same improvement in the quality of conventional care also reduces the propensity of patients to enroll in trials. The resulting delay reduces the present value of per-period patent profits. If we instead consider changes in the price of conventional care, standard arguments suggest that the first term is positive: the profits from a new product increase in the price of conventional care. Subject supply effects suggest that the second term is also positive because of faster development. Thus, a higher price for conventional care would increase the return to innovation.[35]

When extrapolating from this theory to predict the effect of health care reforms on innovation, it is important to be careful about the definition of price. In some cases it will be straightforward. Reforms that reduce the reimbursements that health care providers receive clearly reduce the price of conventional care, which will reduce innovative returns. Other cases may not be so simple. For example, reforms that expand insurance coverage increase the price at which innovative firms can expect to sell their products. However, it also lowers the price at which patients are able to purchase conventional care, so-called out-of-pocket payments. This reduces their willingness to enroll in clinical trials. Thus, the firm price effect controls the sign of the first term in equation (10), while the patient price effect controls the signs of the second term in equation (10). The net effect is ambiguous: insurance expansions may increase or decrease innovation, depending on which effect is larger. More generally,

[35] 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 a 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.

standard effects are driven by the price of conventional treatment faced by health care producers, while the nonstandard effects we highlight—the subject supply effects—are driven by the price faced by consumers.

B. Free Entry and the Equilibrium Duration of Development with a Wage Cap

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

Rationing of subjects by queuing suggests that, holding the supply of subjects constant, an increase in trials increases development times: $T_M > 0$. In the preceding section, we saw that an increase in development time reduces the present value of profits: $\text{NPV}_T < 0$. Together, these facts imply that the present value of innovation falls in the number of trials.

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

$$\text{NPV}(T(M, q, p), q, p) = 0.$$

The equilibrium number of trials will be affected by output market conditions, that is, $M(q, p)$. While development time, $T(M, q, p)$, increases in the number of trials, it also increases in the quality of conventional care and falls in its price.[36]

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

$$\frac{dM}{dq} = \frac{1}{T_M}\left(\frac{\text{NPV}_q}{-\text{NPV}_T} - T_q\right) \leq 0,$$

$$\frac{dM}{dp} = \frac{1}{T_M}\left(\frac{\text{NPV}_p}{-\text{NPV}_T} - T_p\right) \geq 0.$$

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

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

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

$$\frac{dT}{dq} = T_q + T_M \frac{dM}{dq},$$
$$\frac{dT}{dp} = T_p + T_N \frac{dN}{dp}.$$

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

Note that, when development times ration trials, the flow of subjects recruited per period into trials, s, is independent of demand, N, because price controls do not allow wages to rise when demand increases. However, the number of periods or length of development is not independent of demand; it rises with N, that is, $T_N > 0$, thereby rationing further demand. This is important when empirically investigating changes in the participation rates (quantities per period), which will be driven by supply. Put another way, under a maximum-wage policy, the market quantity observed is the lower amount supplied at the constrained price, not the higher amount demanded at that constrained price.

IV. Normative Implications of the Subject Supply Effect

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

A. Free Entry and Inefficiently Long Development Times with Wage Caps

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

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

$$\mathrm{NPV} + M * \mathrm{NPV}_T T_M = 0.$$

This directly implies that profits are positive, $\mathrm{NPV} > 0$. By contrast, in the presence of wage caps, there is free entry until $\mathrm{NPV} = 0$. The analysis in

the previous section implies that development takes inefficiently long relative to what maximizes industry profits. The industry can earn positive profits by restricting entry into development. This will speed up development by reducing the competition for subjects.

B. Privately and Socially Optimal Wages

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

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

$$\text{NPV}(w) = -nw + D(T(w))\Pi, \tag{11}$$

where $T(w)$ is how duration responds to higher wages and $D(T(w)) = \beta^{T(p,q)}A(T^{\text{P}} - T(p, q))$ is the present value of a dollar received throughout the patent window.[37] The necessary first-order condition equates the marginal increase in present value of expected profits as wages speed development with the marginal cost of a higher wage bill:

$$D_T T_w \Pi = n.$$

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

Of course, private profits do not capture the entirety of social welfare. The shortfall has two components:[38]

[37] Recall that a positive wage w for trial subjects implies a negative price p^{e} for experimental treatment in eq. (2); i.e., a positive wage implies $p^{\text{e}} = -w$.

[38] We acknowledge, however, that the wage cap may satisfy some ethical criteria that may matter to welfare. For example, it may serve some deontological aims, such as preventing patients from being coerced by payment into dehumanizing action; preventing patients from making medical decisions motivated by "impure" aims, i.e., something other than their own health or altruism; or it may challenge the professional ethos that elevates care of patients in medicine (e.g., Reiser 2005). Alternatively, the constraint may satisfy a consequentialist aim, such as preventing irrational individuals from making decisions that do not promote their own utility (e.g., Fry et al. 2006). Some bioethicists argue, however, that more robust informed consent and counseling can address both the deontological and consequentialist concerns with subject compensation (e.g., Grady 2001).

$$S(w) = -nw + D(T(w))\left(\Pi + \mathrm{CS}_{\mathrm{patent}}\right) + \beta^{T(p)}A(\infty)\mathrm{CS}_{\mathrm{generic}}.$$

Even when medical products are under patent, consumers earn per period surplus $\mathrm{CS}_{\mathrm{patent}}$, though less than first best from a static perspective. Moreover, consumers earn a larger surplus $\mathrm{CS}_{\mathrm{generic}}$ when products go off patent. Even without wage caps, a firm's compensation to research subjects will not capture the full value of consumer surplus under a patent. If development times are so long that the net present value of product development (eq. [11]) at the privately optimal wage becomes negative, the firm will not develop the product. Yet if the firm internalized the loss of surplus when the product went off patent, it might have reversed course.

The failure of firms to fully internalize the benefit to future consumers of clinical trials suggests the need for Pigouvian subsidies to encourage research subjects to enroll in clinical trials. Therefore, the ethical wage cap has a double cost. It not only prevents the firm from efficiently capturing future producer surplus; it also prevents it from capturing future consumer surplus. The result is yet longer development times in a Pareto sense.[39]

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

It might seem intuitive to address cap-induced delays in development by extending patent life. This would not be a Pareto efficient solution. First, additional patent length generates additional static deadweight loss without speeding up recruitment, hurting future patients in two ways. Second, wage caps would continue to prevent the firm from using the additional producer surplus to compensate subjects. A better use of the extended patent term might be to enable the firm to internalize more of the social surplus from innovation in a regime without a wage cap. The ques-

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

tion of whether direct Pigouvian subsidies or an extended patent term is a superior tool for this problem depends on the same considerations that determine whether it is better to encourage innovation with patents or rewards: whether the government or the firm has a better sense of the social surplus from their innovation.

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

V. Concluding Remarks and Implications for Future Research

In this paper, we have argued that changes in the price and quality of conventional care have nonstandard effects on development mediated by the market for research subjects, and consequently on the health benefits from and future cost growth induced by medical innovation. We illustrated such subject supply effects by exploring the introduction of HAART in 1996, a dramatic increase in the quality of conventional treatment for HIV/AIDS. We documented a substantial reduction in the supply of subjects due to this new innovation. We also provided some evidence in the introduction consistent with a subject supply effect for trials of HCV treatments. In our conclusion, we explore a number of positive and normative issues that would be productive avenues for future research.

A. Additional Evidence of the Subject Supply Effect

We have provided evidence for the subject supply effect for some notable innovations, such as HIV/AIDS and HCV. Another promising field in which to search for evidence of subject supply effects is patients with chronic myeloid leukemia, for whom tyrosine kinase inhibitors such as Gleevec have dramatically improved outcomes outside of trials.

We have also examined the responsiveness of those with and without insurance. The subject effect could likewise be tested by comparing individuals with likely higher or lower sensitivity to changes in conventional care. For example, better-educated patients are more likely to be aware of changes in conventional care than those less well educated and thus exhibit greater subject supply effects. Moreover, because individuals who are poorer are perhaps less likely to have health insurance or be otherwise more sensitive to price, they may remain in trials even when conven-

tional care improves, that is, exhibit smaller subject supply effects. Finally, patient subgroups known not to respond to a newly approved therapy are less likely to exit or avoid trials than subgroups that are responsive to new approved drugs. Each of these is a testable prediction of the subject supply effect.

B. Location of Medical R&D

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

Medical R&D does not appear to conform to this wisdom. The locus of medical R&D has shifted dramatically in the past two decades from Europe toward the United States. In 1990, Europe conducted 15 percent more R&D than the United States, but by 2010 it conducted only 72 percent of US R&D (EFPIA 2011). A common explanation in medical-policy circles for this change is that European price controls are responsible for the shift toward the United States (Gambardella, Orsenigo, and Pammolli 2000; Hassett 2004). This argument is dismissed by economists who use the logic in the preceding paragraph to argue that R&D does not track output markets. However, our analysis of subject markets implies that the naive policy argument may have more credit than it is given.

The competition among trials for research subjects suggests that the amount of research across locations depends on the relative disease prevalence, subject participation rates, and the number of competing trials in each location. This implies that there may be a connection between output markets and development across regions, contrary to standard arguments. Here are three examples. First, trials of treatments for obesity may be easier to conduct in the United States, where those treatments will disproportionately be sold. The high prevalence of obesity there speeds up recruitment through two channels: a direct effect from greater prevalence and an indirect effect from high prices of conventional care. Second, price controls in Europe may be responsible for the decline of clinical trials there. Price controls not only lower future profitability in local output markets but also dissuade patients from participating in local trials. Third, developing countries with low-quality conventional care or high relative prices for that care are good candidates for clinical trials. These conditions speed up trial recruitment. This may explain the recent growth in clinical research in those countries (Thiers, Sinskey, and Berndt 2008). It may explain why few industries other than medical products spend so much on development in the continents of Africa and South America.

C. *Optimal Sample Size Determination*

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

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

D. *First-Mover Advantage in Medical R&D*

Subject supply effects have direct implications for the differential R&D costs associated with first-in-class entry versus subsequent, or so-called me-too, innovation. Our analysis suggests that first-in-class drugs have a first-mover advantage for two reasons. First, firms have lower R&D costs when they alone are recruiting subjects in a disease class. Second, first-in-class innovations, by raising the quality of conventional care in the output market, slow subsequent development. In other words, medical R&D may have a self-limiting effect, with earlier innovations raising the cost of subsequent innovations. This first-mover advantage may have important market structure implications (see Tirole 1994). It also has important implications for the optimal speed of creative destruction for medical products.

E. *Innovations That Affect (Known) Disease Prevalence*

Subject supply effects may differ for curative treatments and new diagnostic tests. Development time falls with the steady-state prevalence of a disease, which, in turn, increases with its incidence, b, and decreases with mortality or the exit rate, m. For some diseases, the exit rate may be affected by the price and quality of conventional care, $m(q, p)$. The cheaper or more effective that conventional care is, the more patients utilize it. Therefore, its quality and price affect not only participation rates but also the stock of subjects available to participate.[40] For example, if a newly approved con-

[40] 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 $1/m$

ventional treatment can cure a disease, improving its quality or reducing its price will lower the prevalence of that disease, reducing trial participation and increasing development times, compounding the usual subject market effect. Conversely, if a newly approved conventional treatment lowers mortality of a disease without curing it, improving it will increase the prevalence of that disease, offsetting the standard subject market effect of from new treatments.[41]

There may also be indirect effects of price and quality of conventional care that operate through the incidence (flow into the disease stock), rather than through cure or mortality (flow out of the disease stock).[42] This is particularly relevant for understanding how diagnostic technologies affect development times. While improved diagnostic capabilities do not increase the actual stock of disease, they may improve recruitment and shorten development times because only diagnosed patients can enter trials.[43]

F. Comorbidities and Cross-Disease Effects

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

For example, conventional care for one disease may affect the prevalence of a second disease. For example, if conventional care for treating heart disease improves, there will be a larger population facing comorbid risks, such as Alzheimer's disease. This, in turn, lowers development times for Alzheimer's research. As a result, while innovation may be self-limiting within a disease, it may increase innovative returns for competing diseases (Geoffard and Philipson 2002).[44] This has important implications for the debate over whether and why there has been a slowdown in aggregate productivity of medical research (e.g., Cockburn 2006). The produc-

---

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.

[41] 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 receiving conventional care, then the overall exit of the stock of disease is given by $m(p, q) = s(p, q)\mathrm{E}[m_E(p_E, q_E)] + (1 - s(p, q))m_C(p, q)$.

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

[43] By contrast, Williams (2010) stressed the "pull" effect of diagnostics. By identifying new patients for treatment, diagnostics increase the returns to treatment innovations.

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

tivity of medical R&D is negatively affected by the self-limiting effects of R&D within a disease but positively affected by cross-disease effects.

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

References

Abraham, Ned S., Jane M. Young, and Michael J. Solomon. 2006. "A Systematic Review of Reasons for Nonentry of Eligible Patients into Surgical Randomized Controlled Trials." *Surgery* 139 (4): 469–83. doi:10.1016/j.surg.2005.08.014.

Acemoglu, Daron, David Cutler, Amy Finkelstein, and Joshua Linn. 2006. "Did Medicare Induce Pharmaceutical Innovation?" *A.E.R.* 96 (2): 103–7. doi:10.1257/000282806777211766.

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Q.J.E.* 119 (3): 1049–90.

Adams, Christopher P., and Van V. Brantner. 2006. "Estimating the Cost of New Drug Development: Is It Really $802 Million?" *Health Affairs* 25 (2): 420–28. doi:10.1377/hlthaff.25.2.420.

Agoritsas, Thomas, Marie Deom, and Thomas V. Perneger. 2011. "Study Design Attributes Influenced Patients' Willingness to Participate in Clinical Research: A Randomized Vignette-Based Study." *J. Clinical Epidemiology* 64 (1): 107–15. doi:10.1016/j.jclinepi.2010.02.007.

Bartlett, John G. 2006. "Ten Years of HAART: Foundation for the Future." Medscape. https://www.medscape.org/viewarticle/523119.

Becker, Gary S., Tomas J. Philipson, and Rodrigo R. Soares. 2005. "The Quantity and Quality of Life and the Evolution of World Inequality." *A.E.R.* 95 (1): 277–91. doi:10.2307/4132680.

Blume-Kohout, Margaret E., and Neeraj Sood. 2008. "The Impact of Medicare Part D on Pharmaceutical R&D." Working Paper no. 13857 (March), NBER, Cambridge, MA.

Brichet, Stephanie, and Nicole Cohen. 2007. "The Pursuit of High Performance through Research and Development Understanding Pharmaceutical Research and Development Cost Drivers." White paper, Accenture.

Cannistra, Stephen A. 2004. "The Ethics of Early Stopping Rules: Who Is Protecting Whom?" *J. Clinical Oncology* 22 (9): 1542–45. doi:10.1200/JCO.2004.02.150.

Carpenter, Charles C. J., Margaret A. Fischl, Scott M. Hammer, et al. 1997. "Antiretroviral Therapy for HIV Infection in 1997: Updated Recommendations of the International AIDS Society—USA Panel." *J. American Medical Assoc.* 277 (24): 1962–69.

Carroll, Ricki, Jules Antigua, Darren Taichman, Harold Palevsky, Paul Forfia, Steven Kawut, and Scott D. Halpern. 2012. "Motivations of Patients with Pulmonary Arterial Hypertension to Participate in Randomized Clinical Trials." *Clinical Trials* 9 (3): 348–57. doi:10.1177/1740774512438981.

CDC (US Centers for Disease Control and Prevention). 1993. "1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults." *MMWR Recommendations and Reports* 41:RR-17. https://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm.

Chan, Tat Y., and Barton H. Hamilton. 2006. "Learning, Private Information, and the Economic Evaluation of Randomized Experiments." *J.P.E.* 114 (6): 997–1040.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *A.E.R.* 102 (4): 1279–1309.

Claxton, Karl, and John Posnett. 1996. "An Economic Approach to Clinical Trial Design and Research Priority-Setting." *Health Econ.* 5 (6): 513–24.

Clemens, Jeffrey. 2012. "The Effect of US Health Insurance Expansions on Medical Innovation." Discussion Paper 11-016, Stanford Institute for Economic Policy Research, Stanford, CA.

Cockburn, Iain M. 2006. "Is the Pharmaceutical Industry in a Productivity Crisis?" In *Innovation Policy and Economics*, vol. 7, edited by Adam B. Jaffe, Josh Lerner, and Scott Stern, 1–32. Cambridge, MA: MIT Press (for NBER).

Cutler, David M. 2007. "The Lifetime Costs and Benefits of Medical Technology." *J. Health Econ.* 26 (6): 1081–1100.

Cutler, David M., Srikanth Kadiyala, Kevin Murphy, Eve Rittenberg, Sherwin Rosen, Bob Topel, and Milt Weinstein. 2003. "The Return to Biomedical Research: Treatment and Behavioral Effects." Paper presented at the conference Measuring the Gains from Medical Research, Washington, DC, December 1999.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski. 2003. "The Price of Innovation: New Estimates of Drug Development Costs." *J. Health Econ.* 22 (2): 151–85.

Dunn, Laura B., and Nora E. Gordon. 2005. "Improving Informed Consent and Enhancing Recruitment for Research by Understanding Economic Behavior." *J. American Medical Assoc.* 293 (5): 609–12.

Durivage, Henry, and Kerry D. Bridges. 2009. "Clinical Trial Metrics: Protocol Performance and Resource Utilization from 14 Cancer Centers." Paper presented at the American Society of Clinical Oncology annual meeting, Orlando, FL.

EFPIA (European Federation of Pharmaceutical Industries and Associations). 2011. "The Pharmaceutical Industry in Figures." Available from Malani.

FDA (US Food and Drug Administration). 1998. "Guidance on Statistical Principles for Clinical Trials." 63 Fed. Reg. 49583 (Sept. 16, 1998).

Finkelstein, Amy. 2004. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Q.J.E.* 119 (2): 527–64.

Fry, Sara T., and Robert M. Veatch. 2006. *Case Studies in Nursing Ethics*. 3rd ed. Sudbury, MA: Jones & Bartlett.

Gambardella, Alfonso, Luigi Orsenigo, and Fabio Pammolli. 2000. "Global Competitiveness in Pharmaceuticals: A European Perspective." MPRA Paper 15965, Munich Personal RePEc (Research Papers in Economics) Archive.

Geoffard, Pierre-Yves, and Tomas Philipson. 1998. "Data Markets and Optimal Sample Size Determination." DELTA Working Paper 98-15, Département et Laboratoire d'Économie Théorique Appliquée, École Normale, Supérieure, Paris.

———. 2002. "Pricing and R&D When Consumption Affects Longevity." *RAND J. Econ.* 33 (1): 85–95.

Grady, Christine. 2001. "Money for Research Participation: Does It Jeopardize Informed Consent?" *American J. Bioethics* 1 (2): 40–44.

Grady, Christine, Neal Dickert, Tom Jawetz, Gary Gensler, and Ezekiel Emanuel. 2005. "An Analysis of U.S. Practices of Paying Research Participants." *Contemporary Clinical Trials* 26 (3): 365–75.

Gulick, Roy M., John W. Mellors, Diane Havlir, et al. 1997. "Treatment with Indinavir, Zidovudine, and Lamivudine in Adults with Human Immunodeficiency Virus Infection and Prior Antiretroviral Therapy." *New England J. Medicine* 337 (11): 734–39.

Halpern, Scott D., David S. Metzger, Jesse A. Berlin, and Peter A. Ubel. 2001. "Who Will Enroll? Predicting Participation in a Phase II AIDS Vaccine Trial." *JAIDS: J. Acquired Immune Deficiency Syndromes* 27 (3): 281–88.

Hammer, Scott M., Kathleen E. Squires, Michael D. Hughes, 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 J. Medicine* 337 (11): 725–33.

Hassett, Kevin A. 2004. "Pharmaceutical Price Controls in OECD Countries." Testimony before the Department of Commerce, International Trade Administration, 2. http://www.aei.org/publication/pharmaceutical-price-controls-in-oecd-countries/.

Lakdawalla, Darius, Neeraj Sood, and Dana Goldman. 2006. "HIV Breakthroughs and Risky Sexual Behavior." *Q.J.E.* 121 (3): 1063–1102.

Lara, Primo N., Jr., Roger Higdon, Nelson Lim, et al. 2001. "Prospective Evaluation of Cancer Clinical Trial Accrual Patterns: Identifying Potential Barriers to Enrollment." *J. Clinical Oncology* 19 (6): 1728–33.

Lichtenberg, Frank R. 1998. "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth." Working Paper no. 6569 (May), NBER, Cambridge, MA.

Malani, Anup. 2008. "Patient Enrollment in Medical Trials: Selection Bias in a Randomized Experiment." *J. Econometrics* 144 (2): 341–51.

Moore, Richard D., and Richard E. Chaisson. 1999. "Natural History of HIV Infection in the Era of Combination Antiretroviral Therapy." *AIDS* 13 (14): 1933–42.

Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *J.P.E.* 114 (5):871–904.

Murray, Jeff. 2011. "Accelerated Approval (AA) Overview of HIV Drug Approvals." Presentation to the US Food and Drug Administration. http://www.healthworkscollective.com/wp-content/uploads/2015/07/UCM243224.pdf, 33–50.

Neumann, Peter J., and Eileen A. Sandberg. 1998. "Trends in Health Care R&D and Technology Innovation." *Health Affairs* 17 (6): 111–19.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *J. Econ. Perspectives* 6 (3): 3–21. doi:10.1257/jep.6.3.3.

O'Connor, Alec B. 2010. "Building Comparative Efficacy and Tolerability into the FDA Approval Process." *J. American Medical Assoc.* 303 (10): 979–80. doi:10.1001/jama.2010.257.

Palella, Frank J., Jr., Kathleen M. Delaney, Anne C. Moorman, et al. 1998. "Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection." *New England J. Medicine* 338 (13): 853–60.

Philipson, Tomas. 1997*a*. "Data Markets and the Production of Surveys." *Rev. Econ. Studies* 64 (1): 47–72. doi:10.2307/2971740.

———. 1997*b*. "The Evaluation of New Health Care Technology: The Labor Economics of Statistics." *J. Econometrics* 76 (1–2): 375–95. doi:10.1016/0304-4076(95)01798-4.

———. 2001. "Data Markets, Missing Data, and Incentive Pay." *Econometrica* 69 (4): 1099–1111. doi:10.1111/1468-0262.00232.

Philipson, Tomas, and Anup Malani. 1999. "Measurement Errors: A Principal Investigator-Agent Approach." *J. Econometrics* 91 (2): 273–98. doi:10.1016/S0304-4076(98)00078-5.

Philipson, Tomas J., and Eric Sun. 2010. *Cost of Caution: The Impact on Patients of Delayed Drug Approvals*. Project FDA Report 2. New York: Manhattan Institute.

PhRMA (Pharmaceutical Research and Manufacturers of America). 2002. "2002 Industry Profile." Washington, DC: PhRMA.

———. 2010. *Pharmaceutical Industry Profile 2010.* Washington, DC: PhRMA. http://www.phrma-jp.org/wordpress/wp-content/uploads/old/library/industryprofile/Profile_2010_FINAL.pdf.

Psaty, Bruce M., and Drummond Rennie. 2003. "Stopping Medical Research to Save Money: A Broken Pact with Researchers and Patients." *J. American Medical Assoc.* 289 (16): 2128–31. doi:10.1001/jama.289.16.2128.

Rao, Vijaya L., and Christopher Daugherty. 2015. "Is It Ethical to Enroll a Patient in a Hepatitis C Virus Clinical Trial When Current Standard of Care Is Highly Effective and Safe?" *Clinical Liver Disease* 6 (5): 120–21. doi:10.1002/cld.512.

Reiser, Stanley Joel. 2005. "Research Compensation and the Monetarization of Medicine." *J. American Medical Assoc.* 293 (5): 613–14.

Ripley, Elizabeth, Francis Macrina, Monika Markowitz, and Chris Gennings. 2010. "Who's Doing the Math? Are We Really Compensating Research Participants?" *J. Empirical Res. Human Res. Ethics* 5 (3): 57–65.

Ross, Sue, Adrian Grant, Carl Counsell, William Gillespie, Ian Russell, and Robin Prescott. 1999. "Barriers to Participation in Randomised Controlled Trials: A Systematic Review." *J. Clinical Epidemiology* 52 (12): 1143–56. doi:10.1016/S0895-4356(99)00141-9.

Tam-McDevitt, Jennifer T., Lodovico Balducci, Robert S. Hauser, Daniel Gura, Aram Paraghamian, Heidi Thomas, and Stuart M. Lichtman. 2007. "Has Demand for Clinical Trial Participants Outpaced Supply?" *J. Nat. Cancer Inst.* 99 (1): 86–87. doi:10.1093/jnci/djk012.

Thiers, Fabio A., Anthony J. Sinskey, and Ernst R. Berndt. 2008. "Trends in the Globalization of Clinical Trials." *Nature Rev. Drug Discovery* 7 (1): 13–14.

Tirole, Jean. 1994. *The Theory of Industrial Organization.* Cambridge MA: MIT Press.

Truong, Tony H., Jane C. Weeks, E. Francis Cook, and Steven Joffe. 2011. "Altruism among Participants in Cancer Clinical Trials." *Clinical Trials* 8 (5): 616–23. doi:10.1177/1740774511414444.

Van Epps, Eric M., Kevin G. Volpp, and Scott D. Halpern. 2016. "A Nudge toward Participation: Improving Clinical Trial Enrollment with Behavioral Economics." *Sci. Translational Medicine* 8 (348): 348fs313. doi:10.1126/scitranslmed.aaf0946.

Williams, Heidi. 2010. "Essays on Technological Change in Health Care Markets." PhD diss., Harvard Univ.