NBER WORKING PAPER SERIES

INNOVATION AND DIFFUSION OF MEDICAL TREATMENT

Barton H. Hamilton
Andrés Hincapié
Robert A. Miller
Nicholas W. Papageorge

Working Paper 24577
http://www.nber.org/papers/w24577

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018

This paper has benefited from presentations in economics departments at Arizona State, Central Florida, Houston, KU Leuven, Minnesota, NYU, Sydney, Toronto, UCLA, Washington, Wisconsin - Madison, and Yale, and in business schools at Carnegie Mellon, Harvard, Stanford, UNC and Washington University in St. Louis, as well as conference presentations at the 2016 AHEW, the Federal Reserve Board NYC, the 2017 SEA meeting, and the 2018 NBER Productivity, Innovation and Entrepreneurship spring meeting. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Innovation and Diffusion of Medical Treatment
Barton H. Hamilton, Andrés Hincapié, Robert A. Miller, and Nicholas W. Papageorge
NBER Working Paper No. 24577
May 2018
JEL No. O31

ABSTRACT

This paper develops and estimates a dynamic structural model of demand for a multi-attribute product. The demand side equilibrium supports a product spectrum, the characteristics of which evolve over time in response to supply innovations induced by the composition and extent of aggregate demand. The direction and speed of innovation is inefficient because individuals create an externality by not accounting for their influence on the discovery process. We apply the model to drugs invented to combat the HIV epidemic, during which frequent, incremental innovations in medication were punctuated by sporadic breakthroughs. In this application products differ in their efficacy and their propensity to cause side effects. Our biennial data on four American cities track a replenished panel of individuals for over twenty years, from when drugs were not only ineffective but also created debilitating side effects, to when the market matured. We find that the externalities are quantitatively important and that even a temporary subsidy would have improved average social welfare and been more equitable.

Barton H. Hamilton
Olin Business School
Washington University in St. Louis
Campus Box 1133
One Brookings Dr.
St. Louis, MO 63130
hamiltonb@wustl.edu

Andrés Hincapié
Department of Economics
University of North Carolina at Chapel Hill
CB # 3305 Gardner Hall
Chapel Hill, NC 27599
andres.hincapie@unc.edu

Robert A. Miller
Carnegie Mellon Universtiy
Graduate School of Industrial Admin.
Pittsburgh, PA 15213
ramiller@andrew.cmu.edu

Nicholas W. Papageorge
Department of Economics
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
and NBER
papageorge@jhu.edu

An Animated Appendix to "Innovation and Diffusion of Medical Treatment" is available at
https://www.dropbox.com/s/zkw64rwqv3wcjxc/treatmentevolutionNew.mp4?dl=0

1 Introduction

Economists have long recognized that innovation is not only determined by science and luck, but is also affected by profit-seeking firms. They direct their inventive activity towards developing new products that meet currently latent consumer demand (Hicks, 1932). Sometimes referred to as demand-pull innovation (Schmookler, 1966; Scherer, 1982), this factor is also the source of an externality (Jovanovic and MacDonald, 1994; Waldfogel, 2003; Finkelstein, 2004). The externality arises because the benefits an individual indirectly confers upon all (other) future individuals through his effect on innovative activity are not reflected in the competitive equilibrium price he pays in the current period.

This paper develops a multi-attribute-product competitive equilibrium model of human capital accumulation affected by product quality where the supply innovation process is affected by the size and composition of demand. We identify the model using panel data that track the path of innovations along with a cohort of individuals. We apply the model to drugs invented to combat the HIV (human immunodeficiency virus) epidemic, during which frequent, incremental innovations in medication were punctuated by sporadic breakthroughs. In this application products differ by their efficacy and their propensity to cause side effects. Our biennial data on four American cities track a replenished panel of individuals for over twenty years, from when drugs were not only ineffective but also caused debilitating side effects, to when the market matured. We use our estimates to quantify the magnitude of the technological externality in the market for HIV drugs.

Our results reveal that individuals' preferences tilt the path of innovation towards treatments with fewer side effects, away from the invention of more effective treatments. Moreover, individuals have a strong distaste for experimentation thus slowing the diffusion of new, superior products as well as the development of future treatments in clinical trials. As a measure of the externality, we compute the marginal increase in aggregate welfare generated by a planner who sends the marginal person to clinical trials at the competitive equilibrium. Since the marginal person does not want to join a trial, he loses a little more than $600 when he is forced to participate. However, because trial participation spurs innovation by pushing up the expected quality and the expected number of new products, the net social gain is about $2,000 per individual. We find that the Pigouvian subsidy needed to attain the planner's trial share is about $16,000 per individual. Our results indicate that providing monetary incentives for trial participation can improve welfare by accelerating the progress of innovation.

This study contributes to a literature on dynamic demand under uncertainty. Following Petrin (2002), each product in our model is a bundle of characteristics — in our case, efficacy and side effects.[1] Moreover, similar to Gowrisankaran and Rysman (2012), we allow product characteris-

[1] Studies pioneering the 'characteristics approach' include Stigler (1945), Lancaster (1966) and Rosen (1974).

tics to have dynamic impacts on consumers. In our model, rational individuals face the uncertainty associated with changing their current market treatment for two reasons. First, medical or commercial intermediaries may assign them a better market treatment. Second, participating in clinical trials might give them access to products that are not yet commercialized. Although experimental products may be worse than the state of the art, clinical trials may provide early access to lifesaving technologies. Modeling trial participation as a rational choice relates our work to Chan and Hamilton (2006), who model the decision to remain in a clinical trial to maintain access to good HIV medicine.

A key departure from the literature on dynamic demand with experimentation is that we explicitly model how these decisions drive innovation and thus future products, which individuals forecast when making their current decisions. Several papers have demonstrated that market size affects the speed of innovation. For example, Finkelstein (2004) shows that policies promoting vaccine use accelerate the development of vaccines and Acemoglu and Linn (2004) relate market size (potential aggregate demand) to pharmaceutical innovation. Also in the medical context, Dranove, Garthwaite, and Hermosilla (2014) identify a "social value" of pharmaceutical innovation, showing that Medicare Part D spurred the development of some drugs. A common theme in this literature is that if consumer behavior drives innovation, it follows that a demand externality arises. Waldfogel (2003) uses the term "preference externalities" to describe the mechanism through which market shares can influence products, thus benefitting individuals with similar tastes.[2] Bolton and Harris (1999) argue that a free-riding problem emerges if experimentation accelerates innovation. In our context, if clinical trials provide social benefits by spurring innovation, rational individuals may choose to participate less than is socially optimal.

We complement previous work by Goettler and Gordon (2011) and Igami (2017) who estimate structural econometric models to analyze the relationship between competition and innovation. Goettler and Gordon (2011) develop a model in which market structure (monopoly versus duopoly) affect innovation in the market for microprocessors.[3] In both models the state of the art in each period becomes the starting point for future innovations. Igami (2017) studies the market for hard disk drives as it transitions from one product generation to the next (5.25- to 3.5-inch). Firms in Igami's model have perfect foresight over exogenously evolving demand, and play a dynamic game in which innovation amounts to introducing the single new product generation. We simplify the supply side process and as a result are able to account for richer consumer characteristics and richer innovation paths. On the consumer side we allow for substantial heterogeneity on observables

[2]Demand externalities have been discussed in a variety of scenarios, including sorting into neighborhoods (Bayer and McMillan, 2012) and the emergence of food deserts (Allcott, Diamond, and Dubé, 2017). In the context of obesity, Bhattacharya and Packalen (2012) provide evidence that individual efforts to prevent obesity can shrink the market size for obesity treatments, which slows technological progress.

[3]They find that the presence of a second firm can slow innovation because no firm expects to capture all profits.

(health status, race, age, education, previous consumption and labor participation). On the supply side, our simplified process supports multidimensional products, multiple product entry, variable changes in technology (both incremental innovations and breakthroughs) and new products that are not necessarily technological improvements.[4] All of these factors play a critical role in determining demand by heterogeneous consumers.

We also contribute to research on structural estimation by providing a simulation-based econometric method to estimate models of endogenous innovation. Methodologically, our empirical strategy builds on Hotz and Miller (1993), Hotz et al. (1994) and Altuğ and Miller (1998) in using conditional choice probabilities (henceforth, *CCPs*) and forward simulation techniques to incorporate how individuals form expectations about future innovations. In our context, the choice set that individuals face evolves stochastically as a function of the innovation process which contains two components: a systematic component endogenous to aggregate choices and unexpected innovation shocks. The systematic component, which emerges endogenously from consumer demand, is captured by a non-stationary reference point describing the current state of technology. The technology paths simulated for this innovation process are based on the CCPs, which reflect rational consumer beliefs about the future.

The remainder of this paper is organized as follows. Section 2 provides a brief historical background, describes our data set, and motivates the model structure with patterns in the data. Section 3 specifies the model. Section 4 provides the parametric specification, analyzes identification and describes the estimation strategy. Section 5 presents parameter estimates and model predictions about the likelihood of technological progress. Section 6 studies the evolution of technology as well as consumer welfare under alternative regimes. Section 7 concludes.

2 Data

Our empirical application focuses on the market for HIV treatments which came into existence around 1984 with the beginning of the HIV pandemic, causing over 613,000 deaths in the U.S. by 2008.[5] HIV infection reduces the ability of the immune system to fight off routine infections, a condition known as AIDS (acquired immunodeficiency syndrome). In developed countries, where access to medication is widespread and often subsidized, technological advancement had transformed HIV infection into a manageable condition with treatments whose side effects are fairly

[4]This is a feature in our data and an equilibrium that emerges naturally in models where individuals are not fully informed about new product characteristics (Miller, 1988).

[5]For comparison, over the same period in the U.S., there were 508,000 homicides and U.S. deaths in World War II were just under 420,000. Currently, there are roughly 50,000 new infections and 13,000 deaths per year in the U.S. that are attributed to HIV/AIDS. Globally, the number of deaths due to HIV/AIDS stands at roughly 35,000,000.

mild. This was not always the case. In the early years of the epidemic, available treatments were not only largely ineffective, but also had uncomfortable, painful and even deadly side effects. Over time many innovations appeared, most of them small, and some worse than existing technology — being more toxic without being more effective. In the mid-nineties, a new set of treatments collectively known as HAART (highly active anti-retroviral treatment) was introduced, transforming HIV from a virtual death sentence into a chronic condition.[6] Within two years, mortality rates fell by over 80% among HIV infected (HIV+) men (Bhaskaran et al., 2008). However, HAART also involved drugs that were highly toxic, driving some people to refrain from using them to avoid often intolerable side effects. Innovations occurring after the mid-nineties had fewer side effects, but were generally no more effective than earlier versions of HAART.

2.1 The MACS Data Set

We use public data from the Multi-center AIDS Cohort Study (MACS). The MACS is an ongoing longitudinal investigation (beginning in 1984) of HIV infection in men who have sex with men (MSM) conducted at four sites: Baltimore, Chicago, Pittsburgh and Los Angeles. At each semi-annual visit, survey data are collected on HIV+ men's treatment decisions, out-of-pocket treatment expenditures, and physical ailments (which can reflect drug side effects), along with sociodemographic information, such as labor supply, income, race, and education. In addition, blood tests are administered at each visit to objectively measure health status. Our main objective measure of immune system health is the *CD4 count*, defined as the number of white blood cells per cubic millimeter of blood. Absent HIV infection, a normal count ranges between 500 and 1500. For HIV+ individuals, a count below 500 indicates that the immune system has begun to deteriorate. However, such individuals may remain asymptomatic. When the CD4 count drops below about 300, a patient is said to suffer from AIDS and his immune system becomes unable to fight off routine infections, which compromises his survival probability.[7] Few data sets have a continuous, precise measure of underlying health, additional data on physical health outcomes, and detailed treatment data along with information on economic outcomes, meaning the MACS data set is uniquely well-suited for an analysis of demand-pull innovation in the market for medical treatments.

The full MACS data set we start with contains information on 6,972 subjects at 49 semi-annual visits for a total of 111,271 observations in the form of subject-visit dyads. We limit our attention to HIV+ individuals, leaving us with 47,753 observations. Due to lack of data on gross income and

[6] There is no vaccine or cure for HIV or AIDS, but HAART is the current standard treatment. In general, 1996 is marked as the year when two crucial clinical guidelines that comprise HAART came to be commonly acknowledged. First, protease inhibitors (made widely available towards the end of 1995) would be an effective HIV treatment. Second, several anti-retroviral drugs taken simultaneously could indefinitely delay the onset of AIDS.

[7] The CD4 cutoff below which AIDS occurs varies between 200 and 350.

out-of-pocket expenditures at earlier visits, we use two samples, a larger sample (20,142 observations) covering visits 6 to 49 which only includes health status and product usage, and a smaller sample (16,851 observations) that starts at visit 14 (roughly, late 1990) containing all variables. The construction of both samples is described in Appendices A and B. The smaller sample comprises 1,719 males, 68 percent white, 22 percent black and the rest Hispanic[8]; 86 percent received some secondary education or more, and 23 percent attended graduate school. Underscoring the gravity of HIV infection, about 40 percent of the HIV+ subjects we observe at least once die prior to the end of the sample period.

Table 1 shows that the share of observations with positive physical ailments is 0.43 and the average CD4 count is 475, in the smaller sample. The share of observations with positive labor supply is 0.63. There is substantial variation in labor supply; 74 percent (68 percent) of unique individuals are observed working (not working) at least once.[9] The share of observations with positive market product consumption and trial product consumption are 0.65 and 0.07, respectively. There is also variation in treatment consumption; 83 percent of unique individuals are observed using a market product at least once and 24 percent opt for early access by participating in a clinical trial at least once during the sample period, suggesting a willingness to experiment with products of uncertain quality.

TABLE 1: Summary Statistics: Subjects-Visits. Visits 14-47 (1990-2007)

	Sample	Pre Haart	Post Haart
Obs	16851	6972	9879
Ailments	0.43	0.45	0.41
Market Product	0.65	0.49	0.76
Trial Product	0.07	0.09	0.05
Work	0.63	0.70	0.58
Age	44.48	40.89	47.01
	(8.03)	(6.99)	(7.75)
CD4	475	407	524
	(297)	(298)	(287)
Gross Income	17567	19036	16531
	(8787)	(8733)	(8677)
Out-of-pocket Expenditures	266	179	327
	(706)	(598)	(767)

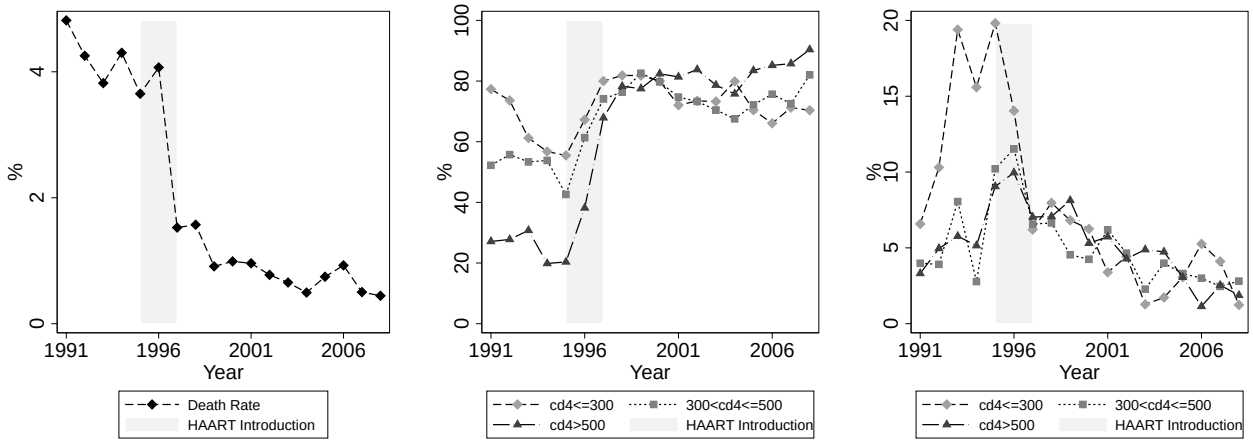
Notes: Standard deviation in parentheses. Gross income and out-of-pocket expenditures are semestral and measured in real dollars of 2000. Pre HAART era corresponds to visit $<= 24$ or roughly before 1996.

[8]Race is an important demographic in this market because participation in clinical trials for new treatments has been shown to be lower among African-Americans, which may reflect different costs associated with treatments or differences in expected health outcomes (Harris et al., 1996).

[9]This is consistent with results in Papageorge (2016) who studies labor supply and medication usage with the MACS data.

2.2 Key Empirical Patterns

Individuals respond to technological change. A distinguishing feature of the market for HIV treatments is that innovations in product quality have life-saving effects. Figure 1(a) shows that prior to the introduction of HAART, death rates were much higher despite a multitude of new treatments becoming available. After HAART, death rates plunge, and continue to fall until 2007, as smaller innovations occurred that made drugs incrementally more effective and less toxic. Table 1 above shows that improvements in survival coincide with improvements in immune system health as measured by the CD4 count. Improvements in health and survival occur as our sample ages and becomes less likely to participate in the labor market (12 percent points less after 1995), which is reflected in the reduction of unconditional average semestral gross income from about $19,036 in the pre-HAART era to $16,531 after 1995.



(a) Survival (b) Market Treatment (c) Trial Treatment

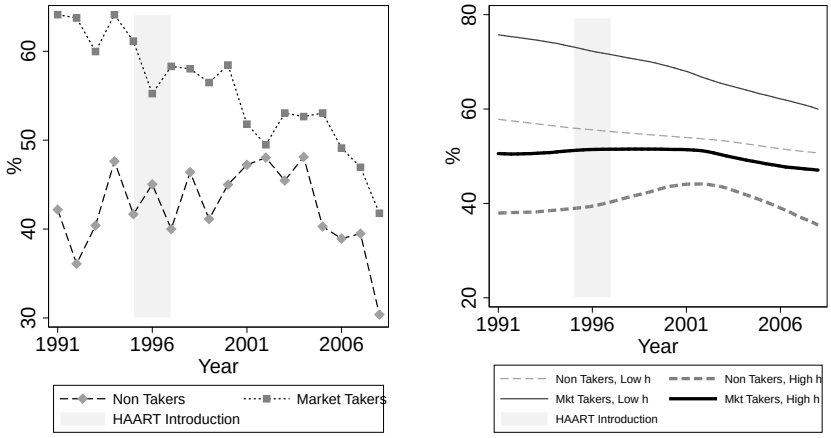
FIGURE 1: Survival and Consumer Demand over Time

Notes: Left panel shows the probability of dying between periods t and $t+1$ conditional on surviving until t. More than 1500 surveyed individuals died for AIDS-related causes during our analysis period. The middle and right panels show consumption by health status.

Table 1 above also shows that improvements in product quality, as measured by the introduction of HAART treatments, induce individuals to consume more HIV treatments. The share of individuals consuming a market product went from 0.49 in the pre-HAART era to 0.76 after HAART was introduced, and individuals out-of-pocket expenditures went from $179 to $327 per semester. Figure 1(b) shows that consumption of commercialized HIV treatment increased and converged across health levels in response to the introduction of more efficacious products. Beyond commercialized treatment, individuals in this market often have the option to consume experimental products in clinical trials. The most dramatic feature of Figure 1(c) is the spike in trial treatment around the time HAART was introduced. Early trial participation is driven largely by individuals with low CD4 counts, suggesting that less healthy individuals may be more willing to experiment with new

products of uncertain qualities.[10] Once efficacious treatments are available, trial participation is no longer driven by sick people willing to face uncertainty in exchange for early access to a product of potentially higher quality.

Product characteristics are multidimensional. Some individuals at risk did not get treated even after life-saving innovations were discovered. Figure 1(b) showed that treatment consumption only climbs to 80% after the introduction of HAART. In part, this happens because products are costly, but also because individuals consider their quality of life, trading product efficacy and side effects. Figure 2(a) shows that individuals who consume a market product suffer more ailments and Figure 2(b) shows that this result holds after controlling for underlying immune system health. Regardless of whether they are healthy or not individuals who consume a market treatment suffer more ailments. Moreover, as products become less toxic over time the gap in ailments between those who are treated and those who are not decreases.



(a) By Treatment Usage (b) By Treatment Usage and CD4

FIGURE 2: Physical Ailments by Treatment Usage and CD4

Notes: Figure contains mean of ailments indicator over time. "Mkt Takers" refers to individuals consuming a market treatment. "High h" refers to individuals with CD4 counts of 250 and above.

The number of new products fluctuates over time. We define a product (or treatment) as a combination of single-product components (see Appendix A). This means that both AZT and the combination of AZT+3TC+Saquinavir are examples of products in our framework. This definition results from noting that the interactions between components matter, and hence the sum of effects

[10] In the years just prior to HAART introduction, the drugs that comprise HAART, including protease inhibitors, marked a substantial improvement over drugs available on the market. In those years, trial participation gave individuals early access to much better products. This relates to the idea of *beta testing* in markets where some consumers are willing to experiment with new products with high potential quality.

of consuming each drug individually does not equal the effect of a treatment formed by the sum of the drugs. Additionally, this definition corresponds to the nature of the market where large treatment innovations such as HAART are themselves combinations of product components. By this definition 86 products were introduced to the market over the sample period with substantial variation in the number of new treatments introduced each period.[11] Figure 3 shows that the unconditional probability of observing more than one product being introduced in a given period is more than 30%, suggesting that product introduction has an intensive margin.

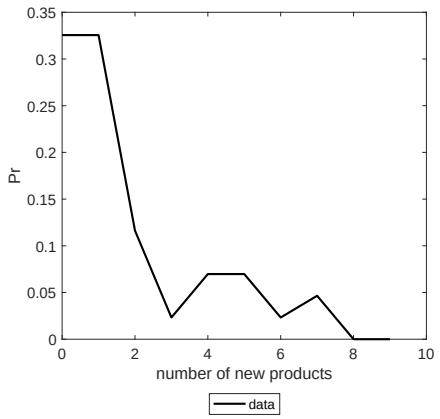
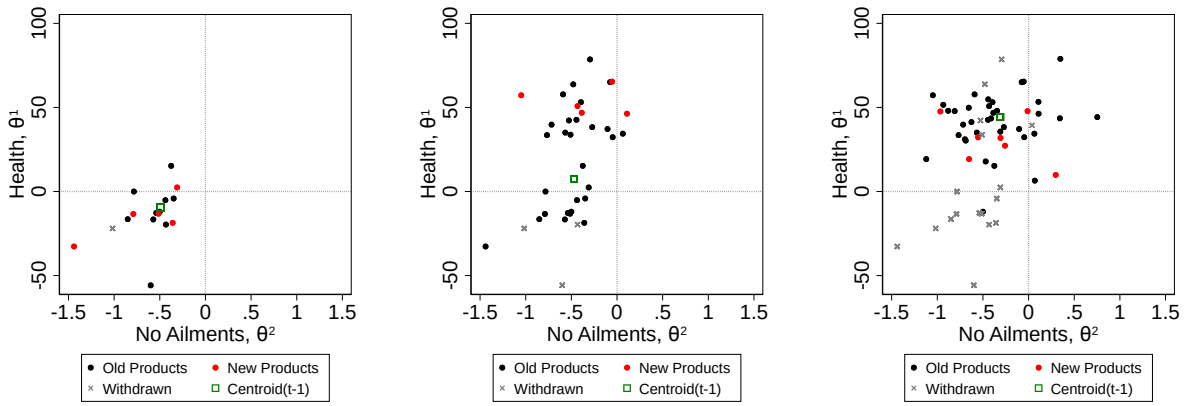


FIGURE 3: Empirical Distribution of Number of New Products.

The characteristics of new products reflect current technology and demand. Figure 4 plots treatment characteristics (effectiveness and lack of side effects) for different periods in our sample, indicating new, old, and withdrawn products as well as the lagged *centroid*, defined as the share-weighted average of product characteristics. New products are introduced around the centroid suggesting that future technologies are based on prevalent technologies today. Over time, the path of technology advances first on the efficacy dimension and then on the side effects dimension. This seem consistent with consumer demand. As we showed in Table 1, after the mid-1990's average CD4 count rises and survival becomes less of a concern. This opens an opportunity for profit-seeking firms to diversify on the side-effects dimension of treatments.

Heterogeneity in product quality increases over time. Substantial variation in the number of new products (Figure 3) and in the size and direction of innovations relative to prevalent technologies (Figure 4) lead to increasing heterogeneity in product quality as the market matures. Figure 5 shows innovation and diffusion of new products over time using a heat map—dark colors corresponds to low (or zero) market share and warmer colors indicate higher market shares. Early on

[11] Table S1 in Appendix A presents our market products including the individual drugs they are composed of as well as their entry and exit time.



(a) Treatments: 1990(1) (b) Treatments: 1996(1) (c) Treatments: 2001(1)

FIGURE 4: Treatment Evolution

Notes: Figure shows snapshots of the evolution of the state of the product market at the different stages. Products are two-dimensional. On the x-axis is a measure of a treatments ability to not cause side effects. On the y-axis is a measure of its contribution to underlying health. Dimensions are measured in different scales. Incumbent products are shown in black. New products are shown in red. Withdrawn products are shown as x. The green square is a measure of the prevalent technology in the previous period.

there are a few products with high shares. As time passes new products strip market share from incumbents and less popular products exit. Low market shares are common in the years following HAART introduction around 1995, when many new treatments were introduced, most of which were effective, but with strong side effects. As the market matured, efficacious treatments with fewer side effects entered the market, increasing market concentration once again.

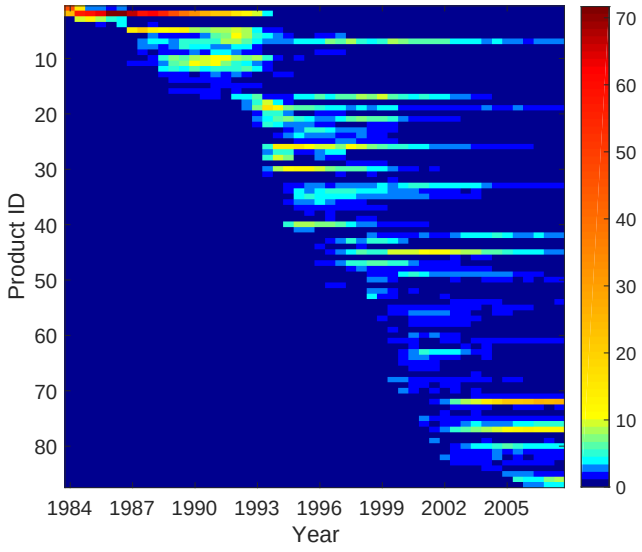


FIGURE 5: Diffusion of Products Over Time

Notes: HIV treatments from 1984 to 2008. Each id—or row—represents a product. Color indicates the share of the market that the product captures. Shares are conditional on individuals who consumed a product.

3 A Model of Demand-Pull Innovation

The dynamic framework of equilibrium product choice that we develop to explain innovation incorporates essential elements of the empirical patterns in Section 2 gleaned from our data on the HIV epidemic. Individuals in the targeted market can either forego medical treatment or consume a unit of a particular product, potentially experimental, offered by suppliers. Products have multiple characteristics, providing current utility (relief) and also human capital (health). Individuals purchasing a new product select a cluster of products with similar characteristics and are allocated a specific product by an intermediary. They can repeat their most recent purchase, but rely on intermediaries when changing treatment. They face the most uncertainty when consuming experimental products in clinical trials, and form rational expectations over upcoming innovations (future choice sets), aligning their actions accordingly. For example, consumers may choose to avoid switching costs associated with using new treatments if they expect better ones to emerge soon. The supply of innovations depends on the size and composition of aggregate product demand (drugs that lose market share being withdrawn from the market), as well as on the choices of consumers trying out experimental products or products that have just been introduced to the market. The market was created with the discovery of latent demand following the HIV outbreak, and matured when consumer choice sets stabilized (the set of distinct prescriptions to treat HIV and their characteristics approaching invariance). Our model reflects these features.

3.1 Supply

In this section we describe the pieces that constitute the supply side of our model. We define products, product characteristics, entry and exit. Then we explain how intermediaries shape consumer choice sets.

Product entry and exit. The market develops over a discrete number of periods $t \in \{0, \ldots, T\}$, where $T \leq \infty$. Each period new products are introduced to the market, and some older ones are withdrawn. Let $k \in \mathbb{N}$ denote a distinct market product, and denote by $\underline{t}_k$ and $\bar{t}_k$ the dates in which the product is introduced to and withdrawn from the market, respectively.[12] Let the set of market products at t be $\mathbf{P}_t \equiv \{k : \underline{t}_k \leq t < \bar{t}_k\}$. Each product k has an immutable set of characteristics $\theta_k \in \Theta \subset \mathbb{R}^K$, where K denotes the number of characteristics describing a product — for instance, efficacy and side effects. In addition to the market products in $\mathbf{P}_t$, an experimental product (such as a clinical trial) is made available for consumption every period. The experimental product is denoted by the letter e and its characteristics are denoted θ_{et}.

[12] A given k can be thought of as a UPC (Universal Product Code) or a specific prescription.

Over time, the market presence of a product reflects its level of adoption. Let s_{kt} be the share of product k in period t and s_{et} be the share of the experimental product available at t. Future innovations reflect the characteristics of the most popular products today. We begin formalizing this idea by defining a summary statistic of the state of technology called the *centroid* for innovation at t, denoted by ω_t, and given by:

$$\omega_t \equiv \sum_{k \in \mathbf{P}_{t-1}} \tilde{s}_{kt-1}\theta_k, \quad \text{where} \quad \tilde{s}_{kt-1} \equiv \frac{s_{kt-1}}{\sum_{k' \in \mathbf{P}_{t-1}} s_{k't-1}} \tag{1}$$

The centroid is a share-weighted average of the characteristics of the products available last period—computed only over buyers—and it is the baseline around which new products emerge in the model. Both the characteristics of the experimental product available at t and the characteristics of new market products available at t are stochastic functions of the centroid and the level of previous aggregate experimentation measured by s_{et-1}. Concretely, these characteristics are drawn from the distribution $g_\theta(\theta|\cdot, s_{et-1})$, where the first conditioning variable is ω_t for the experimental treatment and ω_{t-1} for new market products.[13] The reason why the conditioning centroid in $g_\theta(\theta|\cdot, s_{et-1})$ differs for market and experimental products follows from the fact that, whereas new products appearing at t were developed in laboratories at $t-1$ and therefore depend on ω_{t-1}, the experimental product at t is available at laboratories at t who are experimenting around ω_t.

The number of products entering the market in period t, denoted N_t, is distributed according to $g_N(N_t|\kappa_{t-1}, s_{et-1})$, where κ_{t-1} is the magnitude of previous innovations defined as:

$$\kappa_{t-1} \equiv \sum_{r=1}^{K} \delta_r \cdot \max_{\{k:\, t_k = t-1\}} \left\{ \theta_k^r - \omega_{t-2}^r \right\} \tag{2}$$

given a vector of scaling weights $\{\delta_r\}_{r=1}^K$. κ_{t-1} measures the distance (in characteristics space) between products at time $t-1$ and the centroid around which they were drawn. The distribution of the number of new products captures two ideas. First, more experimentation by product makers can be conducted if larger proportions of consumers try experimental products in clinical trials. Second, large breakthroughs tend to be followed by a relatively large number of new products; this may occur if breakthroughs spur innovative activity as firms attempt to capture market share.

The market share of product k can be decomposed by new $\underline{s}_{kt}$ and repeat $\bar{s}_{kt}$ consumers (where

[13] In our model, the characteristics of the experimental treatment at $t-1$ do not affect the distribution of new product characteristics at t. Although such specifications may seem intuitive if, for instance, better trial products lead to better new market products, in our empirical application the relation between past experimental product characteristics and the characteristics of current market products is statistically insignificant.

$\underline{s}_{kt} + \bar{s}_{kt} = s_{kt}$) and define the conditional share for new consumers as:

$$\tilde{\underline{s}}_{kt-1} \equiv \frac{\underline{s}_{kt-1}}{\sum_{k' \in \mathbf{P}_{t-1}} \underline{s}_{k't-1}} \tag{3}$$

Exit follows a rule described by the dyad $\{\underline{s}, \bar{s}\}$. When $\tilde{\underline{s}}_{kt-1}$ falls below the critical number $\underline{s}$ the product is no longer available for new consumers, and so only repeat consumers can purchase it. Additionally, when $\tilde{s}_{kt-1}$, defined in (1), falls below the critical number $\bar{s}$ the product is withdrawn altogether.

Intermediation. Consumer choices are constrained and guided by intermediaries (such as doctors and retailers) who limit consumer choice sets and make selections on their behalf. More specifically, intermediaries perform two roles in our model: grouping individual products into clusters that form consumer choice sets and, once an individual chooses a cluster, selecting individual products from clusters for their consumption. Lacking data on intermediaries, we model both roles as part of the supply function.

In any given period market products with similar characteristics are clustered into J groups of products. Let $\mathbf{P}_{jt}$ denote the products assigned to cluster j in period t. We assume a protocol or rule, denoted by $C\left(\{\theta_k\}_{k \in \mathbf{P}_t}\right)$, which is common knowledge and uniquely assigns every product on the market at t to a particular cluster. As a consequence of entry and exit, under rule $C(\cdot)$ two products that belong to the same cluster in one period might belong to different clusters in the next. A consumer opting for a new product does not choose a specific product, but a cluster. Also well known is a second rule that determines how products are selected by intermediaries from the particular cluster a consumer chooses. We denote by $q_{kjt}\left(k \left| \mathbf{P}_{jt} \right.\right)$ the probability that the intermediary selects product $k \in \mathbf{P}_{jt}$ when a consumer chooses cluster $j \in \{1, \ldots, J\}$ at t, and we denote by $f_j\left(\theta \left| \mathbf{P}_{jt} \right.\right)$ the distribution of characteristics induced onto the j^{th} cluster:

$$f_j\left(\theta \left| \mathbf{P}_{jt} \right.\right) = \sum_{k \in \mathbf{P}_{jt}} q_{kjt}\left(k \left| \mathbf{P}_{jt} \right.\right) I\{\theta_k = \theta\} \tag{4}$$

3.2 Demand

The individual chooses medical treatment to maximize expected discounted lifetime utility. In making decisions, he observes his current state which includes individual-specific variables, such as health, along with market-level variables. He uses market-level variables to form expectations over the future path of innovation. In specifying his problem, we discuss the choice set, flow utility and stochastic processes governing outcomes and state transitions. Then, we define the individual's value function and highlight the technological externality.

Choice sets. Each period individuals choose one of J clusters by selecting $j \in \{1,\ldots,J\}$, an experimental product available in trial by setting $j = J+1$, the specific market product they consumed in the previous period if it still remains on the market by setting $j = J+2$, or they refrain from making a purchase altogether by setting $j = 0$. Let the choice indicator d_{jit} be equal to one if individual i makes choice j at t, and zero otherwise, with $\sum_{j=0}^{J+2} d_{jit} = 1$, and let the assignment indicator $\tilde{d}_{kit}$ be equal to one if individual i consumed product $k \in \mathbf{P}_t$ at t, and zero otherwise. In addition, define r_{it} as an indicator function that takes the value of one if the individual consumed a market product in the last period and that product has not been withdrawn from the market. Formally,

$$r_{it} = \sum_{k \in \mathbf{P}_{t-1}} \tilde{d}_{kit-1} \mathbf{I}\{\tilde{s}_{kt-1} \geq \bar{s}\} \tag{5}$$

Hence the individual's choice j belongs to the set $\{0,1,\ldots,J+1+r_{it}\}$. In other words, if $r_{it} = 0$ his choice does not include the option to purchase the same product he consumed in the previous period. For notational convenience, in what follows we drop the individual indicator i.

We denote product characteristics at t by $\theta_{jt} \in \Theta$, where $\theta_{0t} \equiv 0$ and $\theta_{J+1,t} \equiv \theta_{et}$. They affect current utility, survival and health, and are measured relative to the value of non consumption. Thus, if $r_t = 1$, the product characteristics that a consumer faces when he makes a repeat purchase are:

$$\theta_{J+2,t} = d_{J+2,t-1}\theta_{J+2,t-1} + \sum_{j=1}^{J} d_{jt-1}\theta_{jt-1} \tag{6}$$

If the consumer selects one of the clusters $j \in \{1,\ldots,J\}$ then θ_{jt} is drawn from $f_j\left(\theta \left| \mathbf{P}_{jt} \right.\right)$; that is, θ_k is selected by the intermediary with probability q_{kjt}. If the consumer selects the experimental product, its characteristics θ_{et} are distributed $g_\theta(\theta|\omega_t, s_{et-1})$.

Human Capital, Outcomes, Preferences and Survival. Product choices affect consumer well-being in three ways: through the accumulation of health, current utility and survival. Let $h_t \in \mathbf{H}$ denote the individual's health, which evolves as a controlled Markov process with transition function $f_h(h_{t+1}|h_t, \theta)$. Thus the probability density function for his health is:

$$f_h\left(h_{t+1} \left| h_t, \sum_{j=0}^{J+1+r_t} d_{jt}\theta_{jt} \right.\right) \tag{7}$$

Let $y_{1t} \in \{0,1\}$ denote lack of physical ailments, $y_{2t} \in \{0,1\}$ labor supply, $y_{3t} \in \mathbb{R}_{\geq 0}$ gross income, and $y_{4t} \in \mathbb{R}_{\geq 0}$ out-of-pocket expenditures. Health and product characteristics affect the vector of

outcomes $y_t \equiv (y_{1t}, \ldots, y_{4t}) \in \mathbf{Y}$ through the conditional distribution:

$$f_y\left(y_t \left| h_t, \sum_{j=0}^{J+1+r_t} d_{jt}\theta_{jt} \right.\right) \tag{8}$$

Individuals draw current utility directly from their health and outcomes as well as from their choices. In addition, at every period they receive idiosyncratic, alternative-specific preference shocks ε_{jt}. Current utility $u_j(h_t, y_t)$ is a real-valued mapping from $\mathbf{H} \times \mathbf{Y}$ denoting the systematic part of current utility, and $\varepsilon_t \equiv (\varepsilon_{0t}, \ldots, \varepsilon_{J+1+r_t,t})$ is an independent and identically distributed random variable with probability density function $f_\varepsilon(\varepsilon_t)$. Individuals are forward looking and discount future utility by $\beta \in (0,1)$. Let b_t be an indicator for being alive at period t, and let $f_b(h_t)$ be the probability of living at t, which depends on the individual's health.[14] The consumer's discounted lifetime utility stream is:

$$\sum_{t=0}^{\infty} \sum_{j=0}^{J+1+r_t} \beta^t b_t d_{jt} \left[u_j(h_t, y_t) + \varepsilon_{jt}\right] \tag{9}$$

3.3 Optimization and equilibrium

At the beginning of period t the density $g_N(N_t \,|\, \kappa_{t-1}, s_{et-1})$ determines the number of new market products, independent draws from $g_\theta(\theta|\omega_{t-1}, s_{et-1})$ determine their characteristics, and market products are withdrawn following the $\{\underline{s}, \bar{s}\}$ rule. Then intermediaries cluster the products according to $C\left(\{\theta_k\}_{k\in\mathbf{P}_t}\right)$ and apply a market sharing rule within clusters using $q_{kjt}(k\,|\mathbf{P}_{kt})$. The characteristics of the trial product are drawn from $g_\theta(\theta|\omega_t, s_{et-1})$. Individuals do not know ex-ante which product within a cluster they will be allocated, they only know the distribution of characteristics for each cluster. They observe their independently distributed idiosyncratic disturbances ε_t and choose $j \in \{0, 1, \ldots, J + r_t\}$. In this way the market share of each product s_{kt} is determined in period t.

We assume that each individual has zero measure in the population, and divide the state variables into those pertaining to the aggregate economy and those specific to the individual. The aggregate state variables for the model, contained in z_t, are the products remaining on the market at t, $\{\theta_k\}_{k\in\mathbf{P}_t}$, the centroid for innovation ω_t, the magnitude of innovations first available at the current period κ_t, the previous share of experimentation s_{et-1}, and the demographics of the consumer population, including $\theta_{J+2,t-1}$, described by the joint distribution $\mathscr{F}_t$. The state variables for individual i include his health h_{it}, recent usage $\theta_{J+2,it-1}$, other demographics a_{it}, and idiosyncratic preference-shocks ε_{it}, as well as the aggregate state z_t.

[14] Thus, $f_b(h_t) = \Pr(b_t = 1|h_t)$, and the probability at t of surviving up to period t' given a sequence $\{h_r\}_{r=t+1}^{t'}$ is $\prod_{r=t+1}^{t'} f_b(h_r)$.

Our analysis is based on individual behavior within a rational expectations equilibrium. Define $z_{it} \equiv (h_{it}, \theta_{J+2,it-1}, a_{it}, z_t)$, let $d_t^e \equiv \left(d_{0t}^e, \ldots, d_{J+1+r_t,t}^e\right)$ be the optimal choice vector solving the consumer's maximization problem in equilibrium, where $d_{jit}^e \equiv d_{jt}(z_{it}, \varepsilon_{it})$, and let:

$$V(z_{it}) \equiv E\left\{ \sum_{\tau=t}^{\infty} \sum_{j=0}^{J+1+r_t} \beta^{\tau-t} d_{ji\tau}^e b_{i\tau} \left[u_j(h_\tau, y_\tau) + \varepsilon_{ji\tau}\right] \middle| z_{it} \right\} \tag{10}$$

denote the ex-ante current value function as of period t. Bellman's principle implies the decentralized equilibrium choices solve:

$$\max_{j \in \{0,1,\ldots,J+1+r_t\}} E\left\{u_j(h_{it}, y_{it}) + \varepsilon_{jit} + \beta V(z_{it+1}) \middle| z_{it}, d_{jit} = 1\right\} \tag{11}$$

The value function is standard but embedded in (10) is the expectation over the aggregate process of innovation which determines future choice sets. The process of innovation generates an externality because the aggregate demand partly determine the course of innovation and product development, which in turn affects the supply side and hence demand in future periods. In a decentralized equilibrium individuals do not take into account the consequences of their actions (such as their experimentation or their adoption of products with certain characteristics) on the future payoffs of other individuals.

4 Empirical Implementation

Leaving aside the role of technological progress, the identification of our model is standard. The model is estimated with a simulated methods of moments CCP estimator, adapted to account for aggregate shocks arising from product development, and pre-estimation of the state transitions. This section analyzes identification, explains the equations at the heart of the estimation, presents the parameterization we adopt as model primitives—and to conduct counterfactuals—and describes each estimation step in sequence.

4.1 Identification

On the supply side, the primitives of the model include the product characteristics themselves $\theta_k = \left(\theta_k^1, \theta_k^2\right)$, the transition functions determining the number of market products entering the market $g_N(N_t \,|\, \kappa_{t-1}, s_{et-1})$ and the characteristics of new and experimental products $g_\theta(\theta|\omega, s_{et-1})$, the distributional characteristics of each market cluster $j \in \{1, \ldots, J\}$, $f_j\left(\theta \,\middle|\, \mathbf{P}_{jt}\right)$, and the exit rule $\{\underline{s}, \bar{s}\}$. On the demand side, the primitives include the flow utility function $u_j(h_t, y_t)$, the probability

density function of the idiosyncratic preference shocks $f_\varepsilon(\varepsilon)$, the subjective discount factor β, the transition function for human capital $f_h(h_{t+1}|h_t,\theta)$, the outcomes mapping $f_y(y_t|h_t,\theta)$, and the survival probability $f_b(h_t)$.

Our data essentially span the epoch between the time at which pharmaceutical companies began supplying treatments that had some chance of being effective until the time at which that product market matured; that is, $t \in \{0,\ldots,T\}$. Over this entire time phase we track a (replenished) panel of individuals $i \in \{1,\ldots,I\}$ including their consumption sequence $d_{it} = (d_{0it},\ldots,d_{J+1+r_{it},it})$, health h_{it}, death b_{it}, outcomes y_{it} and demographic background a_{it}.We also observe the history of each market product k—and therefore the history of clusters under rule $C\left(\{\theta_k\}_{k\in\mathbf{P}_t}\right)$—including the date at which it was introduced to the market $\underline{t}_k$, when it was withdrawn $\bar{t}_k$, and its product share in the target or potential market s_{kt}, decomposed by new $\underline{s}_{kt}$ and repeat $\bar{s}_{kt}$ consumers.

The transition function for human capital $f_h(h_{t+1}|h_t,\theta)$, the outcomes mapping $f_y(y_t|h_t,\theta)$, and product characteristics θ_k, are identified using the cross sectional features of the consumer panel. In particular, we assume that future underlying health h_{t+1} and physical ailments y_{1t} are a function of current health, product characteristics, and idiosyncratic shocks. We assume that product characteristics enter linearly in both processes, which delivers product characteristics—relative to the no product alternative—as the coefficients of indicator variables of product usage. We rule out individual-specific treatment effects because our sample is not large enough to back out product-specific distributions of treatment effects for the 80 plus products we observe. The survival probability $f_b(h_t)$ is also identified off the cross sectional features of the consumer panel.

Both the distribution of characteristics of new and experimental products as well as the distribution of the number of new products are identified using the product panel. We observe the menu of products introduced from the discovery of latent demand when the pandemic starts (around 1985) to when the market has matured (around 2007). This provides us with 43 observations from the equilibrium distribution of the number of new products $g_N(N_{t+1}|\kappa_t,s_{et})$ and 94 observations from the equilibrium distribution of product characteristics $g_\theta(\theta|\omega,s_{et})$. The characteristics distributions $f_j\left(\theta\left|\mathbf{P}_{jt}\right.\right)$ of all clusters are identified directly by the proportions of each product within its cluster. The exit rule is identified by the aggregate data on new $\underline{s}_{kt}$ and repeat $\bar{s}_{kt}$ consumers for each product. Recalling the definition of $\tilde{\underline{s}}_{kt}$ and $\tilde{s}_{kt}$ in (1):

$$\underline{s} = \min_{k,t}\{\tilde{\underline{s}}_{kt}\} \quad \text{and} \quad \bar{s} = \min_{k,t}\{\tilde{s}_{kt}\} \tag{12}$$

Finally, following much of the literature in discrete choice, we assume the discount factor β and the choice disturbance density $f_\varepsilon(\varepsilon)$, and parameterize the latter as Type 1 Extreme value.[15]

[15]We estimated the model for values of $\beta \in \{0.8, 0.9, 0.95\}$ and found that 0.95 delivered the lowest value of the criterion function.

We also assume that individuals know the aggregate processes generating products, product characteristics, and within-cluster distributions. Hence, identification of $u_j(h_t, y_t)$ follows from the general arguments of Magnac and Thesmar (2002), more specifically covered in the framework of Arcidiacono and Miller (2017).

4.2 Specification

Time t evolves in half-year periods, corresponding to the frequency of MACS data collection. On the supply side, products have $K = 2$ characteristics, efficacy denoted θ^1 and propensity not to generate further physical ailments (or side effects) denoted θ^2. The distribution of characteristics of experimental products and new products appearing at t, $g_\theta(\theta|\cdot, s_{et-1})$, is derived from the following process:

$$\theta_k - \omega_t \mathbf{I}\{k = et\} - \omega_{t-1}\mathbf{I}\{k \in \mathbb{N}, \underline{t}_k = t\} = \phi_0^\nu + \phi_1^\nu \cdot s_{et-1} + \nu_k \tag{13}$$

where $\mathbf{I}\{k = et\}$ indicates whether product k is the experimental product at t and $\mathbf{I}\{k \in \mathbb{N}, \underline{t}_k = t\}$ indicates whether product k is a newly introduced market product. Equation (13) specifies new market products and trial products as displaced innovations around the centroid.[16] The magnitude and direction of the innovation depend on the previous share of experimentation in clinical trials and an independent and identically distributed innovation shock ν_k drawn from $f_\nu(\nu)$ with $E[\nu_k|s_{et-1}, \omega_t, \omega_{t-1}] = 0$, which is estimated non parametrically. (See Section 4.3).

We assume that the distribution of the number of new products which enter the market at t, $g_N(N_t \,|\, \kappa_{t-1}, s_{et-1})$, is a negative binomial that permits dispersion in the mean. That is:

$$\begin{aligned} N_t &\sim \textit{Poisson}\left(\mu_{t-1}^*\right); \qquad \mu_{t-1}^* \sim \textit{Gamma}\left(1/\alpha_{t-1}^N, \alpha_{t-1}^N \mu_{t-1}\right) \\ \mu_{t-1} &= \exp(\phi_1^N \kappa_{t-1} + \phi_2^N s_{et-1}); \qquad \alpha_{t-1}^N = \exp(\phi_3^N + \phi_4^N \kappa_{t-1}) \end{aligned} \tag{14}$$

where κ_{t-1} is defined in (2) and the scaling weights are given by the maximum innovations observed in the data:[17]

$$\delta_r^{-1} \equiv \max_{k:\, \underline{t}_k = \tau - 1, \forall \tau > 0} \left\{ \theta_k^r - \omega_{\tau-2}^r \right\}, \text{ for } r \in \{1, 2\} \tag{15}$$

We relax the exit rule $\{\underline{s}, \bar{s}\}$ defined in Section 3.1 as follows: the condition for rule $\underline{s}$ to apply (i.e. exit for new consumers) must be satisfied during three consecutive periods and the condition

[16] We tested the hypothesis that new products are displaced innovations around the centroid, i.e. that the centroid's coefficient is 1, and we could not reject the null hypothesis.

[17] The scaling weights account for the fact that different characteristics may be measured in different scales.

for rule $\bar{s}$ to apply (i.e. exit for all consumers) must be satisfied during two consecutive periods. The number of consecutive periods for each exit rule are chosen so that a single period of low demand does not lead to a premature exit, which is what we observe in the data. This relaxation adds two state variables to the aggregate state of the problem, $\mathscr{E}^1_{t-1}$ and $\mathscr{E}^2_{t-1}$, which are indicators of to what extent the conditions for exit are binding:

$$\mathscr{E}^1_{kt} = \mathbf{I}\left\{\underline{\tilde{s}}_{kt-1} < \underline{s}\right\}\left(\mathscr{E}^1_{kt-1} + \mathbf{I}\left\{\underline{\tilde{s}}_{kt-1} < \underline{s}\right\}\right); \quad \mathscr{E}^2_{kt} = \mathbf{I}\left\{\tilde{s}_{kt-1} < \bar{s}\right\}\left(\mathscr{E}^2_{kt-1} + \mathbf{I}\left\{\tilde{s}_{kt-1} < \bar{s}\right\}\right) \quad (16)$$

where $\underline{\tilde{s}}_{kt-1}$ and $\tilde{s}_{kt-1}$ are defined in (3) and (1), respectively, and $\mathscr{E}^1_{kt_k} = \mathscr{E}^2_{kt_k} \equiv 0$. Exit for new consumers binds when $\mathscr{E}^1_{kt} = 3$ and exit for all consumers binds when $\mathscr{E}^2_{kt} = 2$.

Intermediation is characterized by the clustering rule $C\left(\{\theta_k\}_{k\in\mathbf{P}_t}\right)$ and the distribution of characteristics for a given cluster $f_j\left(\theta\left|\mathbf{P}_{jt}\right.\right)$. We specify the clustering rule as the solution to a k-means clustering algorithm so that at every period t the clusters $j = 1,\ldots,J$ are chosen to minimize:[18]

$$C\left(\{\theta_k\}_{k\in\mathbf{P}_t}\right) = \sum_{j=1}^{J}\sum_{k\in\mathbf{P}_t}\mathbf{I}\{k\in j\}\left\|\theta_k - \theta^c_j\right\|^2, \qquad \theta^c_j \equiv \frac{\sum_{k\in\mathbf{P}_t}\mathbf{I}\{k\in j\}\,\theta_k}{\sum_{k\in\mathbf{P}_t}\mathbf{I}\{k\in j\}} \quad (17)$$

where $\sum_{j=1}^{J}\mathbf{I}\{k\in j\} = 1$ for all $k\in\mathbf{P}_t$. The distribution of characteristics for a given cluster follows from (4) and $q_{kjt}\left(k\left|\mathbf{P}_{jt}\right.\right)$, which is given by:

$$q_{kjt}\left(k\left|\mathbf{P}_{jt}\right.\right) = \frac{\exp\left(x^w_{kt}\gamma^w\right)}{\sum_{k\in j}\exp\left(x^w_{kt}\gamma^w\right)} \quad (18)$$

where x^w_{kt} includes a constant term, the ranking (within its cluster) of the characteristics of the product, the number of members in the cluster, whether the product is new, and several interactions. The vector of parameters γ^w is obtained from a nonlinear regression of within cluster shares $s_{kt|j}$ such that:

$$\mathbb{E}\left[s_{kt|j}|x^w_{kt}\right] = \exp\left(x^w_{kt}\gamma^w\right), \qquad s_{kt|j} \equiv \frac{s_{kt}}{\sum_{k'\in j}s_{k't}} \quad (19)$$

On the demand side, and again omitting the individual subindex i to reduce notation, individual demographics a_t contain age (in half year increments), race/ethnicity (black, Hispanic, white), and education level (high school, some college, college or more than college). Individual health $h_t \in \mathbf{H} \subset \mathbb{R}_{\geq 0}$ is measured by the CD4 count.[19] The production function for health that determines $f_h\left(h_{t+1}\left|h_t,\theta\right.\right)$ captures non linearities in the effects of current health (e.g. as the CD4 count

[18] See Duda and Hart (1973) and Andrew W. Moore's *K-means and Hierarchical Clustering* tutorial at http://www.cs.cmu.edu/∼awm/tutorials.html. (See Appendix B for more details.)

[19] CD4 ranges from 0 to 2915 in our analytic sample with a median of 448. Healthy CD4 counts are those above 500 units per mm^3 and typically range between 500 and 1,500.

reaches AIDS levels), and is given by the following polynomial:

$$h_{t+1} = \sum_{s=0}^{5} \gamma_s^h h_t^s + \sum_{j=0}^{J+1+r_t} d_{jt}\theta_{jt}^1 + \varepsilon_t^h \tag{20}$$

where $\mathbb{E}[\varepsilon_t^h | h_t, \theta] = 0$ and ε_t^h is drawn from the nonparametric distribution $f_{\varepsilon^h}(\varepsilon^h)$. We augment the conditional distribution for outcomes $f_y(y_t | h_t, \theta, a_t)$ to include individual demographics a_t and non linearities in the effects of current health with the equation system:

$$\Pr[y_{1t} = 0 | h_t, \theta] = \left(1 + \exp\left(\sum_{s=0}^{5} \gamma_s^x h_t^s + \sum_{j=0}^{J+1+r_r} d_{jt}\theta_{jt}^2\right)\right)^{-1} \tag{21}$$

$$\Pr[y_{2t} = 1 | x_t^l] = \left(1 + \exp\left(x_t^l \gamma^l\right)\right)^{-1} \tag{22}$$

$$y_{3t} = x_t^m \gamma^m + \eta + \varepsilon_t^m \tag{23}$$

$$y_{4t} = x_t^o \gamma^o + \varepsilon_t^o \tag{24}$$

The probability of suffering physical ailments y_{1t} in (21) depends on previous health and the side-effects characteristics of the product consumed. Since labor supply is not the main purpose of this paper, we do not model y_{2t} as a choice to avoid further complications. However, labor supply may be affected by treatment choices, e.g., through health status. Moreover, labor supply affects income and therefore utility. We treat labor supply as a state variable that individuals know at the beginning of the period before making their treatment decision. The transition probability of labor market participation is given by (22) where $x_t^l = [1, h_t, \ldots, h_t^4, a_t, y_{2t-1}]$. Gross income y_{3t} is governed by the process in (23) where $x_t^m = [1, h_t, \ldots, h_t^7, a_t, y_{1t}, y_{2t}]$, η is an exogenous individual-specific productivity and ε_t^m are iid income shocks that the individual observes before making their treatment choice.[20] The individual state z_{it} also contains η. Out-of-pocket expenditures for health y_{4t} are determined by (24) where $x_t^o = [1, h_t, \ldots, h_t^6, a_t, y_{1t}, y_{2t}, \{d_{jt}\}_{j=0}^{J+1+r_t}]$ and ε_t^m are iid $Normal(0, \sigma_o^2)$. Expenditures increase from purchasing a treatment but may also increase due to underlying health and physical ailments. Out-of-pocket expenditures are censored at zero, which is why we model them separately from gross income. Since we do not directly observe prices, and in order to simplify the problem, (24) assumes a constant cost of consuming an experimental product as well as a constant cost of consuming a market product.[21]

The probability of being alive during the current period $f_b(h_t)$, is augmented to include physical

[20] We do not need to make parametric assumptions on these shocks because they enter linearly in the payoffs from choosing all alternatives and therefore do not affect choices.

[21] End-users customarily pay a standardized deductible that is a fraction of the brochure price of the drug paid by the insurance company. Median out-of-pocket drug costs are about $300 every six months for a regime of drugs that would cost the insurance company between $5,000 and $15,000.

ailments and demographics, and is determined by $x_t^d = [1, h_t, \ldots, h_t^5, a_t, y_{1t-1}]$, through the density

$$f_b(h_t) = \left(1 + \exp\left(x_t^d \gamma^d\right)\right)^{-1} \tag{25}$$

The systematic part of current utility is a mapping from health, outcomes, choices and demographics given by:

$$u_j(h_t, y_t) = \alpha_m(y_{3t} - y_{4t}) + \alpha_{xp} y_{1t} d_{0t} + \alpha_{jh} h_t + \alpha'_{ja} a_t \tag{26}$$

The first term on the right-hand-side of (26) is net income $y_{3t} - y_{4t}$, so that α_m captures consumption utility. The second term captures the utility cost of physical ailments. $y_{1t} = 1$ indicates that the individual does not suffer from physical ailments, and the interaction with d_{0t} captures how distaste for ailments can vary depending on whether or not a treatment is being consumed. We normalize the utility cost of ailments while using a treatment to zero. Hence, α_{xp} represents the differential distaste for ailments for individuals who are not taking a treatment. The last two terms of (26) capture choice-specific utility associated with health, and demographics—age and race. Since the only relevant differences among clusters—both within a period and over time—are given by the distribution induced by the members of each cluster, $f_j\left(\theta \left| \mathbf{P}_{jt} \right.\right)$, we assume that parameters α_{jh} and α'_{ja} do not vary across clusters. Nevertheless, individuals do derive different choice-specific systematic utility from consuming a trial product or from choosing the same market product they consume in the last period. This captures how experimentation in treatment choices can imply additional costs or benefits. In the case of clinical trials, utility parameters capture, for example, the fear of trying an experimental drug or preferences for altruism since trial participation may help future patients. In the case of continuing to use the same product, utility parameters may capture a preference for certainty, which could help to explain consumer reluctance to switch even when better products enter the market.

Health affects lifetime utility through its impact on future health and survival as well as through the probability of suffering physical ailments, and physical ailments affect utility directly and through earnings. In addition, health affets flow utility directly with an effect measured by α_{jh}. This captures differences in the time and psychic costs of finding a trial slot by health if, for example, doctors are more willing to encourage experimentation or if trial slots are more readily available for sicker patients. It also captures how individuals may be more willing (or encouraged by intermediaries) to choose from a cluster when in poor health.[22] Finally, the preference shocks ε_{jt} are iid Type I Extreme Value across individuals and alternatives and over time.

[22] We normalize α_{0h} and α'_{0a} to zero. Therefore, α_{jh} captures the additional effect—relative to those who do not use a product—that health and demographics have on the utility of continuing with with the same product, or experimenting with a new treatment, either through cluster use or trial participation.

4.3 Estimation

This section summarizes our estimation procedure. First, we obtain estimates of product characteristics, transition functions and outcome equations. Then we estimate CCPs and simulate paths of future choices, states and technology that we use to form moments and deliver estimates of the utility parameters with a closed-form solution GMM estimator. Our simulation accounts for the endogenous evolution of the aggregate state as well as aggregate innovation shocks. We obtain standard errors using subsampling with 100 subsamples. Detailed explanations of some of the steps below are found in Appendix B. Our estimation procedure follows the steps below:

1. *Products*. We define one single trial product per period as the one used by those individuals joining a clinical trial. Given these definitions, we estimate product characteristics together with the health and no-ailment processes ((20) and (21)).

2. *Clusters*. Based on product characteristics from step 1, we cluster products at every period using a k-means algorithm determined by the clustering rule in (17). Then, using the characteristics of the products in each cluster and product shares, we obtain the distribution of characteristics induced onto the j^{th} cluster using (18) and (19). We use the first two moments to describe each distribution.

3. *Innovation*. We back out centroids for innovation for each period. (See (1)) using product characteristics from step 1.) Then we estimate the innovation equation (13) and use the residuals to non-parametrically estimate the distribution of innovation shocks $f_\nu(\nu)$. Finally, we use the number of new products per period to estimate the distribution of the number of new products using (2) and (14).

4. *Outcomes and survival*. We estimate processes for labor supply, income, out-of-pocket expenditures, and survival using (22), (23), (24), and (25).

5. *Utility function*. We estimate the utility parameters in (26) using a GMM estimator and moment conditions that equate the log odds ratio of current conditional choice probabilities with an expression involving utility parameters and simulated future CCPs, states and choices (Hotz et al., 1994; Altuğ and Miller, 1998). In order to obtain these moments we estimate flexible parametric CCPs that control for the aggregate state z_t as well as individual-specific state variables. (See Appendix B.) We then forward-simulate paths of future technology as well as individual choices and states that serve as inputs to the simulated continuation value. We break dependency between observations by simulating aggregate technology paths for every observation $\{i,t\}$: we first simulate a collection of aggregate paths describing product evolution (this requires simulating all individuals' behavior for each aggregate path); then

for every observation $\{i,t\}$ we simulate individual choices and transitions taking as given a group of randomly selected aggregate paths.

5 Parameter Estimates

Figure 6 plots observed treatment choices over time along with those generated by the model given the state at every point in time. The estimated model captures the key trends, including the rise in repeated usage as treatments improve over time and the decline in the share of individuals not consuming a product. It also captures relatively well the share of individuals experimenting—either by buying from a cluster or by consuming a trial product. The rest of this section presents the estimates of the supply (Section 5.1) and demand side processes (Section 5.2) as well as model predictions about the likelihood of technological progress (Section 5.3).

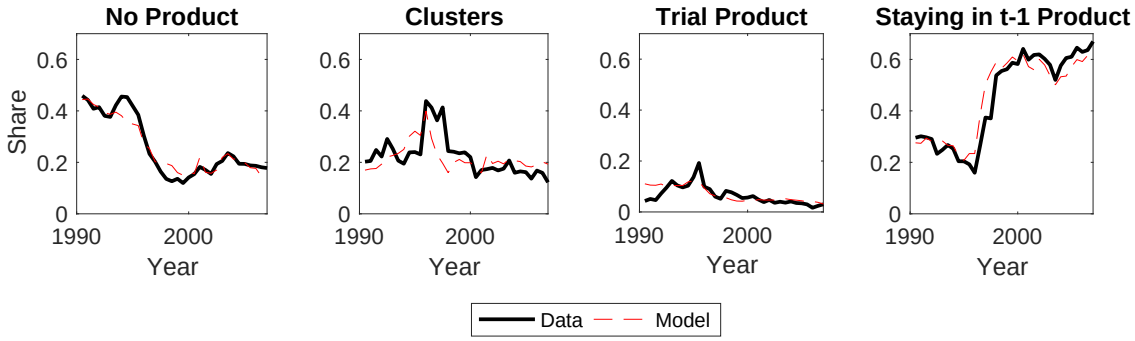


FIGURE 6: Goodness of Fit
Notes: Simulated and empirical choice shares over time.

5.1 Supply

Innovation. The process of innovation in (13) specifies the characteristics of new products as displaced innovations around the centroid. Table 2 reports the estimated coefficients of the systematic part of innovations. Our estimates suggest that new drugs would be worse on average (relative to the centroid) if no individuals participated in trials. With increase consumption of trial products, on average the quality of products introduced in the future improves in both dimensions of quality. Since average trial participation in our sample is 7 percent (Table 1), the estimates in Table 2 imply that new products were on average more efficacious than the prevalent technology—as measured by the centroid—but did not offer fewer side effects.[23]

[23]Expected health innovations are positive for lagged trial shares above 5.6 percent, and expected innovations on the ailments dimension are positive for lagged trial shares above 7.7 percent.

TABLE 2: Innovation Components

	Health Innovation			*Ailments Innovation*		
	coef.	est.	se	coef.	est.	se
s_{et-1}	ϕ^{ν}_{11}	433.11	(19.95)	ϕ^{ν}_{12}	1.93	(0.34)
Constant	ϕ^{ν}_{01}	-24.14	(1.47)	ϕ^{ν}_{02}	-0.15	(0.03)

Notes: Estimates from (13). In parentheses, standard errors computed using subsampling with 100 subsamples.

The stochastic part of every innovation in (13) is an aggregate innovation shock drawn from the non-parametric distribution $f_{\nu}(\nu)$ showed in Figure 7. The distribution, centered on $(0,0)$, is unimodal. Conditional on trial participation, most innovations are small improvements. Because products are multidimensional, it is possible for innovation shocks to generate products that are more efficacious, but also more toxic, causing worse ailments via side effects. However, $f_{\nu}(\nu)$ in Figure 7 exhibits a positive correlation of 0.24 between the two quality dimensions of an innovation shock; unexpected improvements in efficacy tend to be accompanied by fewer side effects.

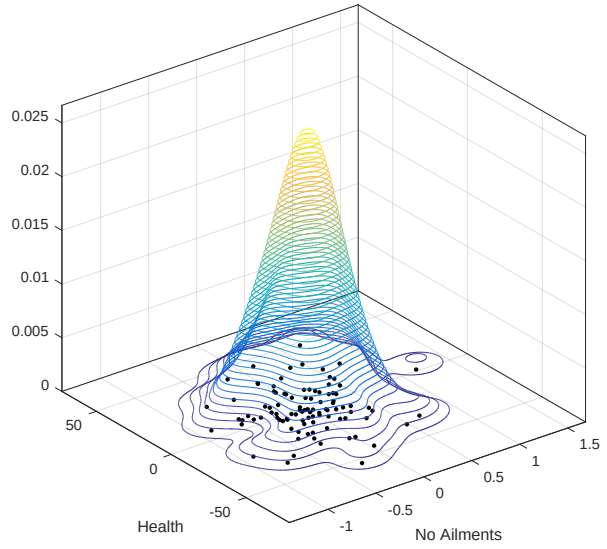


FIGURE 7: The Distribution of Innovation Shocks, $f_{\nu}(\nu)$.
Notes: $f_{\nu}(\nu)$ is estimated non-parametrically off the residuals from (13).

Estimates of the distribution of the number of new products are shown in Table 3. Large positive innovations in previous periods lead to a higher number of new products in expectation. The magnitude of previous innovations also reduces the dispersion around the number of new products that enter. Both patterns are consistent with firms vying for market share following breakthroughs by producing similar products. The share of individuals opting to participate in a clinical trial in the prior period also increases the likelihood of more products entering the market. Our interpretation is that as more consumers select trial products, firms increase their experimental activity. This

leads to an increase in the quantity of viable new treatments that can be introduced into the market. The estimated distribution fits the data well. (See Figure S2 in Appendix C.)

TABLE 3: Distribution of Number of New Products, F_N

$\ln \mu$				$\ln \alpha$			
variable	coef.	est.	se	variable	coef.	est.	se
κ_{t-1}	ϕ_1^N	0.432	(0.246)	*Constant*	ϕ_3^N	-0.206	(0.451)
s_{et-1}	ϕ_2^N	6.177	(2.462)	κ_{t-1}	ϕ_4^N	-1.019	(0.626)

Notes: Model is specified in (14). κ_{t-1} measures the magnitude of previous innovations. $E[N_t] = \mu_{t-1}$ and $Var[N_t] = \mu_{t-1}(1 + \alpha_{t-1}^N \mu_{t-1})$. In parentheses, standard errors computed using subsampling with 100 subsamples.

Intermediation. Whenever individuals decide to switch to a different market product they select one of J clusters of products and are assigned a product from the chosen cluster according to the assignment probabilities specified in (18) and (19). The point estimates of the intermediation process are laid out in Table S2 in Appendix C. Summarizing, products with better side effects relative to other products in the cluster have higher within-cluster shares. Also, the shares of products ranking lower in both efficacy and side effects disproportionately decrease with the number of products within the cluster.

5.2 Demand

Transitions and outcomes. Processes for health and ailments are estimated jointly with product characteristics. (See Appendix B.) We relegate estimated treatment characteristics as well as point-estimates of the health and ailments processes to Tables S3 and S4 in Appendix C. Instead, in the left panels of Figure 8 we present the impact profile of health at the beginning of the period (h_t) on future health (h_{t+1}) and ailments (y_{1t}). While the production function for health displays low concavity, the production function for ailments is very non-linear. The panels suggest that in the region where CD4 counts are below 250, changes in health generate much larger shifts in the log odds ratio of suffering ailments. The reason is that HIV infection has a gradual negative impact on immune system health, as measured by CD4 count. However, the impact of CD4 count on ailments is not gradual. It is virtually non-existent until CD4 count has dropped below about 250 and AIDS-related symptoms emerge.

We relegate the point estimates for all other outcomes processes (gross income, out-of-pocket expenditures, labor supply) and survival to Tables S5 to S8 in Appendix C. Instead, we plot the relationships between health, outcomes and survival in the remaining panels of Figure 8. Although health displays relationships with other outcomes in the expected direction, these relationships are non-linear. As explained above, this is due to large changes in physical health once the AIDS

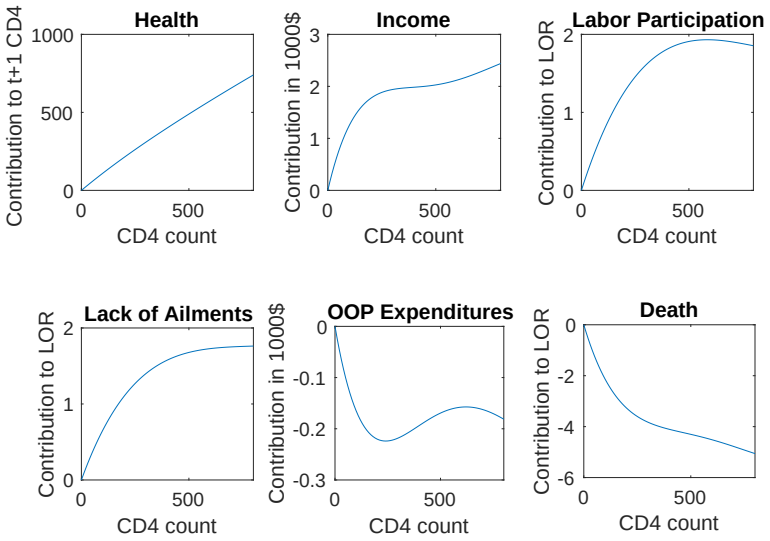


FIGURE 8: Effect of Current Health on Future Health and Outcomes

Notes: CD4 Count measured in hundreds of cells per microliter. LOR stands for log odds ratio. OOP stands for out-of-pocket. Semestral income and expenditures measured in thousands of dollars of 2000.

threshold is reached. These relationships underscore the importance of modeling the relationship between health and outcomes in a non-linear fashion for HIV+ individuals.

Beyond the relationships between outcomes y_t and health h_t discussed in the previous paragraphs, several key patterns emerge. (See point estimates in Tables S5 to S8 in Appendix C.) Individuals who do not suffer ailments have higher gross income, since their productivity is likely to be higher. Income is concave in age and increases with employment and education, though racial minorities earn less on average. Out-of-pocket expenditures increase with age, minorities spend less, and more educated males spend more. Similarly, individuals that suffer ailments face higher expenditures over and above treatment costs, perhaps because they are managing other health conditions. Employment increases expected expenditures, which may reflect different pricing schemes for public versus private insurance. The log odds ratio of working versus not working increases with age until about age 40, after which point it decreases. The log odds ratio of labor force participation increases with education. Moreover, there is strong persistence in employment, reflected by a large increase in employment odds for individuals who worked in the previous period. Estimates also imply that the log odds ratio of death decreases with age until about age 35 and then increases. The likelihood of death is lower for black males and for males who are not suffering ailments.

Utility. Estimates of the utility function are reported in Table 4. Individuals gain positive utility from income net of out-of-pocket expenditures (including treatment costs), which captures con-

sumption utility. Moreover, a lack of physical ailments enters positively into the flow utility.[24] Prior literature has shown that even in the context of a deadly infection (HIV) individual treatment choices reflect a distaste for side effects (Chan and Hamilton, 2006; Papageorge, 2016). The positive estimate of α_{xp} implies that the cost of ailments is larger when individuals are not consuming a treatment. This finding is consistent with the idea that the utility cost of ailments from side effects of medical treatment may be less than the cost of ailments due to illness.

TABLE 4: Utility Parameters, u_t

coef.	variable					est.	se
α_m	$NetIncome_t$ $(y_{3t} - y_{4t})$					0.057	(0.057)
α_{xp}	$NoAilments_t \cdot NoProduct_t$ $(y_{1t} d_{0t})$					1.019	(1.767)
		Cluster $j = 1, \ldots, J$		*Trial* $j = J+1$		*Repeat* $j = J+2$	
coef.	variable	est.	se	est.	se	est.	se
α_{ja1}	*White*	-3.546	(0.744)	-1.468	(0.280)	0.502	(0.567)
α_{ja2}	*Black*	-4.190	(0.762)	-2.553	(0.334)	0.276	(0.613)
α_{ja3}	*Hispanic*	-3.967	(0.958)	-1.585	(0.356)	0.707	(0.454)
α_{ja4}	Age_t	0.043	(0.011)	0.032	(0.005)	0.009	(0.007)
α_{jh}	$h_t/10^3$	-2.021	(0.423)	-2.461	(0.203)		

Notes: Estimation of (26). Discount factor $\beta = .95$. $J = 3$. $NoProduct_{it}$ indicates whether he did not consume a product. h_t is defined as the number of white blood cells per cubic millimeter of blood. In parentheses, standard errors computed using subsampling with 100 subsamples.

Utility parameters for treatment choices are interacted with race and age. We find that using treatments is costly for all individuals, with higher costs accruing to African Americans and Hispanics.[25] African Americans face a particularly high penalty for consuming a trial product, a finding that is consistent with a broad literature investigating historical reasons why African Americans are reluctant to participate in clinical trials (Harris et al., 1996; Alsan and Wanamaker, 2018). However, age helps to mitigate the utility costs of treatment: older agents have more contact with the medical community, or become accustomed to taking medications. We also find that better health leads to larger utility costs of experimentation. This is consistent with more frequent contact with doctors among less healthy patients, who may thus face lower costs of switching to new or experimental treatments. In the case of trials, there may be more slots available for sicker patients if a goal is to test drugs on patients who most need them. Finally, the utility of remaining on a treatment is positive, although insignificant; if the individual is suffering ailments repeated consumption is preferred over taking no treatment at all and over choosing a cluster or trial treatment. This result underscores the idea that individuals are reluctant to experiment with new drugs.

[24]Table S9 in Appendix C shows that if the estimated ancillary parameters of the CCPs were the true parameters, both net income and physical ailments would be highly significant. However, both net income and physical ailments parameters become insignificant once the standard errors are corrected for the two stage procedure. Our final specification in (26) was determined by the statistical significance of results before the computationally intensive correction of standard errors in the last stage of estimation.

[25]Recall that the non pecuniary benefits from no treatment are normalized to zero.

5.3 The Likelihood of Technological Progress

According to our model the observed path of innovation is a draw from a stochastic process that relates the supply of new treatments to consumer behavior. We assess the likelihood of the observed innovation path using the estimated innovation process (Section 5.1). Starting at two different initial states, the first semester of 1991 and the second semester of 1996, we simulate 100 innovation paths spanning until the end of our sample in 2008, and compare them against the realized path of innovation. While the state at 1991 captures the market prior to the introduction of breakthroughs when health was declining among HIV+ men, the state at 1996 captures the market after the introduction of HAART, which reversed the trend in average health for HIV+ men. We compare realized and simulated average consumer health, ailments and survival probabilities.[26] This exercise reveals what rational individuals expected as the market for HIV drugs matured.

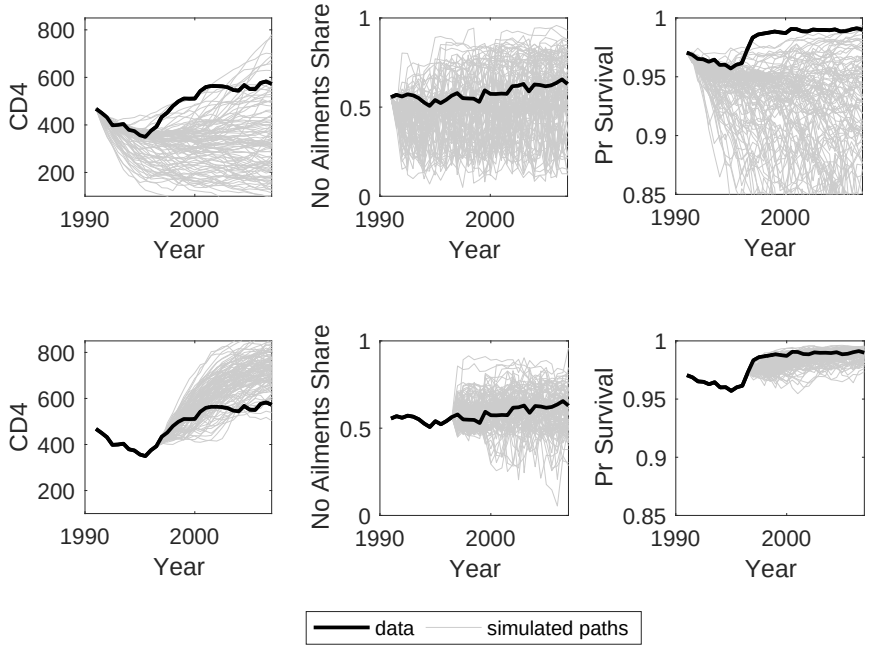


FIGURE 9: Distribution of Technology Paths: Individuals
Notes: 100 simulated paths conditional on the state of the world at 1991 and 1996.

The top-left panel of Figure 9 displays the evolution of average CD4 count using 1991 as the starting point. Although most of the mass of simulated paths is below the realized path of average health, the gap between the realized path and its simulated counterparts is small before 1996. Once the breakthrough in technology is introduced around 1996 the realized path diverges from most of the simulated paths. If we instead use 1996 as the starting point (bottom-left) the realized

[26] We also compare the simulated and realized paths of prevalent technology—the centroid—and the share of market product consumers in Figure S3 in Appendix C.

path underperforms; innovations did not substantially raise average health after 2000 even though steady improvements in average health were likely. The simulated probability, evaluated at 1991 and 1996, of a path with higher average health than the realized path by the end of the sample period (2008) is 9 percent and 97 percent, respectively.[27]

In the middle and right panels of Figure 9 we repeat the exercise with the share of individuals with no physical ailments and the average probability of survival, respectively. Although the realized path of no physical ailments seems to be above the majority of paths starting at 1991, it is located rather in the middle of the mass of simulated paths regardless of initial state. This is because there were no significant breakthroughs in the ailments dimension. Hence, rational individuals expected the no-ailments share to evolve approximately as it did. The simulated probability, evaluated at 1991 and 1996, of a path with higher no-ailments share than the realized path by 2008 is 45 percent and 40 percent, respectively. The top-right panel suggests that the future discovery of a breakthrough that would increase survival the way HAART did was given a very low probability at 1991. At 1996, even though survival had improved, individuals continued to expect lower survival rates than the realized path. The simulated probability, evaluated at 1991 and 1996, of a path with higher survival rate than the realized path by 2008 is 2 percent and 34 percent, respectively.

6 Technology and Welfare under Alternative Regimes

The evolution of technology, and ultimately consumer welfare, is affected by demand externalities arising from the innovation process. To quantify the importance of these externalities we investigate several alternative regimes in which the supply process is left unchanged. We conduct two types of experiments and compare them against the benchmark. In Section 6.1 we analyze how technology would evolve if consumers had less influence over the process of innovation, restricting the role of demand pull as a factor determining the equilibrium. In Section 6.2 we explore how social welfare would improve if demand externalities were internalized in a limited way.

6.1 Demand Pull: How Consumer Choices Affect Technology

We investigate two ways in which the process of innovation is detached from demand. The first experiment assumes a scientific body determines exit exclusively on the basis of product quality, and at the entry margin innovation is based on average product characteristics. The second experiment eliminates the effect of repeat purchase on innovation. For each experiment we present

[27] Although we choose the end of the sample period to define the probabilistic event, similar exercises can be undertaken by counting the amount of paths that are above the realized path at any given period after the initial period.

results averaging over 500 simulated paths starting at the first semester of 1991.

Exogenous scientific intervention. In the first alternative regime innovation is independent of consumer demand. We redefine the centroid to be a simple average of products in the market — as opposed to a share-weighted average — and take as given our estimates of the supply process in Section 5.1.[28] Thus, new product characteristics are no longer dependent on product demand. We also separate product exit from demand by adopting two alternative exogenous exit rules designed to resemble the actions of scientific authorities tasked with keeping only the best products on the market. The first rule, denoted *frontier*, removes from the market all products that are not on the technological frontier. This rule provides an upper bound for how fast innovation can move. The second rule, denoted *inverse frontier*, captures expert intervention in a less draconian way. Under this regime, the product exit rate equals the average product exit rate computed over the baseline simulations in Section 5.3. The exit rate pins down the number of products to be withdrawn and the worst products are dropped from the market independent of demand.[29]

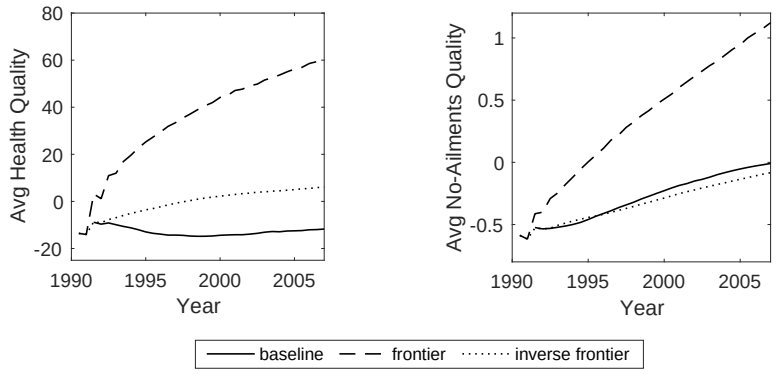


FIGURE 10: Alternative Regimes: Exogenous Scientific Intervention

Notes: Evolution of the average quality of products in the market under alternative regimes (500 simulations per regime) conditional on the state of the world at the first semester of 1991. The *baseline* is the estimated model of demand-pull innovation. In both the *frontier* and *inverse frontier* regimes the centroid is not driven by demand as it is a simple average of products on the market. In the *frontier* regime all products inside the quality frontier are exogenously withdrawn. In the *inverse frontier* regime the exit rate of products equals the average exit rate in the baseline simulations and the worst products are withdrawn.

Figure 10 shows that under the first regime (frontier) innovation is more rapid, leading to much better products on both dimensions of quality. In contrast, the path of product quality is not as different from the baseline under the second regime (inverse frontier). In the rational expectations equilibrium individuals avoid using the very worst products. Consequently, removing these products has little impact on the centroid, and hence on subsequent innovations. Nevertheless,

[28] For simulation, we need a path of trial participation $\{s_{et}, s_{et+1}, \ldots, s_{eT}\}$ to feed into the distribution of the number of new products in (14) and the innovation process in (13). We use the average path of trial participation over 500 simulations using the baseline model. (See Section 5.3.)

[29] Selection of worst products in the inverse frontier regime is explained in Appendix C.

our estimates imply that the inverse frontier regime does lead to somewhat higher average health quality.

Eliminating the effect of repeat purchase. Since consumers dislike changing treatment, they face a tradeoff between old and new technologies, and are more likely to repeat purchase if prior treatment offers higher qualities than current clusters—where multiple qualities are balanced through individual preferences. Our second counterfactual regime assigns individuals to alternatives in the choice set in the same proportions as the benchmark (including trial products and no treatment), but makes repeat consumption of old technologies random. Thus the preferences and characteristics of repeat consumers do not guide the direction of innovation. Figure 11 shows that eliminating the effects of repeat custom improves health and survival, but leads to more physical ailments. The reason is that individuals prefer medical treatments with fewer side effects despite the detrimental impact on their survival. The new rule tilts the path of innovation towards more efficacious treatments with greater side effects.

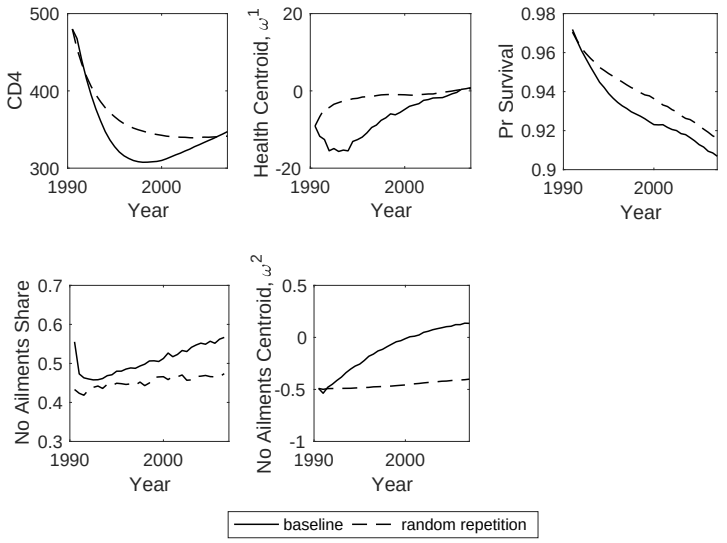


FIGURE 11: Alternative Regimes: Eliminating the Effect of Repeat Purchase.
Notes: Average paths computed over 500 simulations that are conditional on the state of the world at 1991. The *baseline* is the estimated model of demand-pull innovation. The baseline solid lines in Figure 11 are the averages of the grey lines in Figure 9 and Figure S3 in Appendix C. Individuals in the alternative regime are assigned alternatives using the unconditional shares from the baseline model as assignment probabilities.

6.2 Targeting Demand Externalities in Innovation

Together with our results in Section 5, our simulations in Section 6.1 underscore three reasons why scientific intervention may not increase welfare: individuals care about quality of life and face high costs of changing treatments or experimenting with trial products. The optimal planner policy in-

corporating the externality is a mapping from consumer characteristics into treatment alternatives, but solving for this mapping is intractable given the size of the state space. In this section we study a temporary policy change lasting only one period before reverting to the competitive equilibrium. This allows us to compute continuation values using the CCPs estimated in Section 4.3. Nevertheless, the policy has a long term impact because it affects the state variables of the competitive equilibrium resuming next period.

Mandated treatment. Our first policy experiment assigns individuals to alternatives based on health and previous treatment. The population is split into four groups with high or low health and who are or are not potential repeat customers (those who bought a market product in the previous period). For each of the groups the planner either assigns one of the alternatives in the choice set to all members of the group, or he assigns the competitive equilibrium individual-specific allocation. We solve the problem in the first semester of 1991 by computing average simulated lifetime utility under all possible allocation rules (1,764 total).[30]

TABLE 5: Mandated Treatment

				Groups, Group Shares and Assignment			
		% Gain/Loss over CE		High H No Repeat	High H Repeat	Low H No Repeat	Low H Repeat
	Average Welfare ($1000)	High H	Low H	0.50	0.26	0.05	0.19
	351.61	2.3	-2.8	0	6	6	0
	351.28	2.3	-3.2	0	6	0	0
Top Rules	350.84	2.1	-3.1	0	5	0	0
	350.82	2.1	-2.9	0	5	6	0
	350.63	1.0	1.2	0	6	0	6
	⋮						
Competitive Equilibrium	346.11	-	-	6	6	6	6
	⋮						
	167.97	-54.2	-40.9	1	4	6	4
	167.96	-53.3	-44.6	1	4	1	4
Bottom Rules	167.33	-53.5	-44.9	1	4	3	4
	167.24	-53.5	-45.1	3	4	2	4
	165.90	-54.4	-43.1	1	4	4	4

Notes: Planner's problem solved at 1991. *High H* (*Low H*) individuals have $CD4 > (\leq) 250$. *Repeat* (*No Repeat*) costumers can (cannot) repeat their prior period market product. Population shares shown on top of each group label. Numbers 1 to 3 correspond to clusters and numbers 0, 4, 5, and 6 stand for no treatment, trial product, repeat consumption, and the competitive equilibrium allocation, respectively.

Table 5 presents the top and bottom five assignment rules. In the worst rules the planner imposes experimentation on healthy patients who dislike it most, and discards information contained in the competitive equilibrium allocation, often assigning individuals to low quality clusters, thereby incurring switching costs. In the best rules the planner improves technology by relying on healthy potential repeat customers because their previous choices incorporate product quality information, and treats hardly anyone else because average product quality is low in 1991. The top rule increases average welfare by 1.6% but decreases equity. Relative to the competitive equilib-

[30] We simulate aggregate lifetime utility 200 times for each each of the $7^2 * 6^2 = 1,764$ possible assignment rules.

rium allocation, healthy individuals ($CD4 > 250$) gain 2.3% in average lifetime utility while the unhealthy lose 2.8%. However, the fifth top rule increases average welfare by almost as much as the top rule and both health groups gain.

Optimal experimentation with trial products. Our results in Section 5.1 show that the share of trial product consumption increases the expected number of new products as well as their quality. Our second policy experiment focuses on the externality arising from individually rational agents who do not internalize the effects of their experimentation with trial products on the welfare of other individuals through their impact on the evolution of technology. The planner assigns alternatives based on all components of the individual state, but he can only assign one of two alternatives: the trial product or the competitive equilibrium allocation (excluding trial products). Facing a tradeoff between innovation and individual experimentation costs, the planner chooses a cutoff for experimentation s^*_{et} such that the gain in average welfare from allocating the next individual to the trial product is no longer positive. We solve the problem at the first semester of 1991 and again at the second semester of 1996.[31]

TABLE 6: Optimal Experimentation with Trial Products

Planner trial share s^*_{et}	0.100	0.185
Competitive equilibrium trial share s_{et}	0.102	0.092
Average lifetime utility at planner solution	346	360
Average lifetime utility at competitive equilibrium	346	354
Increment in trial share for marginal person sent to trials at s_{et}	0.001	0.002
Individual loss for marginal person sent to trials at s_{et}	-0.178	-0.628
Social gain from sending marginal person to trials at s_{et}	-1133	1051
Flat subsidy per trials participant to attain s^*_{et}	-	16.0

Notes: Planner's problem solved at the first semester of 1991 and the second semester of 1996. Monetary values in $1,000s.

Results are presented in Table 6. In 1991, the planner's experimentation share is approximately the same as the competitive equilibrium share (s_{et}). The costs of increased experimentation outweigh the benefits of new drugs in a time when individuals are very sick, no good treatments have been invented and previous innovations have been small. By 1996, large innovations had occurred, further innovations were therefore more probable, and consumers' health was improving fast. At this period the planner's experimentation share doubles the competitive equilibrium share and yields an average welfare of $360,000, about 2% higher than average welfare at the competitive equilibrium. Figure 12 illustrates the planner's problem at 1996, it shows that trial participation shares up to 9 percent points above s_{et} generate welfare gains that outweigh individual losses due to experimentation.[32] Average welfare drops precipitously for trial shares beyond s^*_{et} because the

[31] We discretize the trial share from 0 to 1 in increments of 0.005 units and simulate aggregate lifetime utility 1000 times for each value.

[32] The solid line in Figure 12 yields from applying a fifth degree local smoothing polynomial over the original less smooth Figure S4 in Appendix C. We use this smoothed version to evaluate marginal gains.

new individuals being assigned to trials face larger losses relative to their optimal choice, and the innovation benefits to additional experimentation are not large enough.

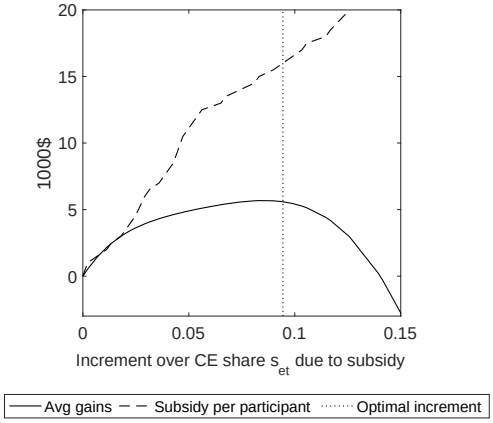


FIGURE 12: Optimal Assignment to Clinical Trials with a Flat Subsidy

Notes: On the x-axis are increments in trial share over the competitive equilibrium (CE) share s_{et}. The solid line represents average gains in welfare over the CE allocation. The dashed line indicates the subsidy per participant necessary to decentralize a given increment. The dotted line indicates the planner's optimal increment over s_{et}. Year is 1996.

To measure the magnitude of the externality we obtain the derivative of average welfare with respect to the trials share, evaluated at the competitive equilibrium share s_{et}. We achieve this by assigning the marginal consumer to a trial and computing the net social benefit.[33] Focusing on year 1996, we find that the marginal consumer loses roughly \$600 (Table 6). However, because trial participation spurs innovation by raising the expected quality and the expected number of new products, the net social gain is over \$2,000 per person. In our sample of 445 individuals in 1996, this means that a \$600 loss from raising trial participation by 1 person (about 0.22 percentage points) leads to a welfare gain of roughly \$1,000,000.

Although these results suggest a substantial externality associated with trial product experimentation, governments with the authority to assign individuals to clinical experimentation may target groups that do not belong to their coalition.[34] Figure 12 shows the flat Pigouvian subsidy necessary to decentralize a given trials share. The subsidy that attains the planner's optimal trial share s^*_{et} is about \$16,000 per participant. This subsidy represents a large reallocation of utility because all trial product consumers are paid, including those who would consume the trial product without the subsidy. Besides, the subsidy must be large enough to induce the marginal, highest-cost person into a trial. However, decentralizing s^*_{et} with a subsidy not only increases efficiency but also equity. As opposed to most top mandated treatment policies (Table 5), the subsidy decreases the gap in lifetime utility between the sickest individuals ($h_t < 200$) and everyone else by ten percent.

[33]The marginal consumer would otherwise choose something else, but he faces the smallest lifetime utility loss from assignment to a trial.

[34]The infamous Tuskegee experiment is an example of this (Harris et al., 1996; Alsan and Wanamaker, 2018).

Additionally, equity does not increase at the expense of the healthy as the lump sum transfer (about $3,000) to pay for the subsidy is below their welfare gains under s^*_{et}. Equity increases because the sickest individuals benefit the most from faster innovation and because they are more likely to consume trial products. In other words, the subsidy reduces technological free-riding undertaken by healthy individuals.

7 Conclusion

We provide a framework to assess how consumer choices affect technological progress. In our case, aggregate consumer demand affects not only the speed of innovation, which has been studied in a number of contexts, but can also tilt the path of innovation in cases where product quality is multi-dimensional. We apply our framework to study consumer behavior and innovation in the market for HIV drugs. In this context, we capture several mechanisms through which consumer demand affects innovation, including experimentation with new drugs by participating in clinical trials, thereby accelerating innovation. By joining a trial, individuals gain access to experimental products that may be high-quality breakthroughs, but may also be less efficacious or painfully toxic. Additionally, consumer decisions can bend the technological path if firms avoid innovating around unpopular products. We show that consumer behavior can slow the process of innovation and bend it towards less efficacious products that hamper survival probabilities.

Because individuals do not internalize the consequences of their product choices on other consumer's welfare, an externality arises through their impact on technological progress. Our estimates show that a constrained planner can increase average welfare by at least two percent (approximately $6,000 per individual), and that providing incentives for trial participation can improve social welfare. Although demand pull has been widely recognized as a source of innovation, the externality it creates had not been previously quantified. Our demand-pull framework could be applied to other industries and integrated with models focusing on the supply side.

References

Acemoglu, Daron and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119 (23):1049–90.

Allcott, Hunt, Rebecca Diamond, and Jean-Pierre Dubé. 2017. "The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States." NBER WP 24094.

Alsan, Marcella and Marianne Wanamaker. 2018. "Tuskegee and the Health of Black Men." *Quarterly Journal of Economics* 133 (1):407–455.

Altuğ, Sumru and Robert A. Miller. 1998. "The Effect of Work Experience on Female Wages and Labour Supply." *Review of Economic Studies* 65 (1):45–85.

Arcidiacono, Peter and Robert A. Miller. 2017. "Identifying Dynamic Discrete Choice Models off Short Panels." Mimeo, Carnegie Mellon University.

Bayer, Patrick and Robert McMillan. 2012. "Tiebout Sorting and Neighborhood Stratification." *Journal of Public Economics* 96 (11):1129–1143.

Bhaskaran, K., O. Hamouda, M. Sannes, F. Boufassa, A.M. Johnson, Lambert P.C., K. Porter, and CASCADE Collaboration. 2008. "Changes in the Risk of Death after HIV Seroconversion Compared with Mortality in the General Population." *JAMA* 300 (1):51–59.

Bhattacharya, Jay and Mikko Packalen. 2012. "The Other Ex Ante Moral Hazard in Health." *Journal of Health Economics* 31 (1):135–146.

Bolton, Patrick and Christopher Harris. 1999. "Strategic Experimentation." *Econometrica* 67 (2):349–374.

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

Dranove, David, Craig Garthwaite, and Manuel Hermosilla. 2014. "Pharmaceutical Profits and the Social Value of Innovation." NBER WP 20212.

Duda, Richard O. and Peter E. Hart. 1973. *Pattern Classification and Scene Analysis*. Wiley.

Finkelstein, Amy. 2004. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Quarterly Journal of Economics* 119 (2):527–564.

Goettler, Ronald L. and Brett Gordon. 2011. "Does AMD Spur Intel to innovate More?" *Journal of Political Economy* 119 (6):1141–1200.

Gowrisankaran, Gautam and Marc Rysman. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy* 120 (6):1173–1219.

Harris, Yvonne, Philip B Gorelick, Patricia Samuels, and Isaac Bempong. 1996. "Why African Americans May Not Be Participating in Clinical Trials." *Journal of the National Medical Association* 88 (10):630.

Hicks, John R. 1932. *The Theory of Wages*. London: Macmillan.

Hotz, V. Joseph and Robert A. Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies* 60 (3):497–529.

Hotz, V. Joseph, Robert A. Miller, Seth Sanders, and Jeffrey Smith. 1994. "A Simulation Estimator for Dynamic Models of Discrete Choice." *Review of Economic Studies* 61 (2):265–289.

Igami, Mitsuru. 2017. "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981-1998." *Journal of Political Economy* 125 (3):798–847.

Jovanovic, Boyan and Glenn M. MacDonald. 1994. "Competitive Diffusion." *Journal of Political Economy* 102 (1).

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2):132–157.

Magnac, Thierry and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70 (2):801–816.

Miller, Robert A. 1988. "Innovation and Reputation." *Journal of Political Economy* 96 (4):741–765.

Papageorge, Nicholas W. 2016. "Why Medical Innovation is Valuable: Health, Human Capital, and the Labor Market." *Quantitative Economics* 7 (3):671–725.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110 (4):705–729.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1):34–55.

Scherer, Frederic M. 1982. "Demand-Pull and Technological Invention: Schmookler Revisted." *Journal of Industrial Economics* :225–237.

Schmookler, Jacob. 1966. *Invention and Economic Growth*. Harvard University Press.

Stigler, George J. 1945. "The Cost of Subsistence." *Journal of Farm Economics* 27 (2):303–314.

Waldfogel, Joel. 2003. "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets." *RAND Journal of Economics* 34 (3):557–568.

For Online Publication Supplement to
"Innovation and Diffusion of Medical Treatment"

A Data Appendix

Data collection for the Multi-Center AIDS Cohort Study started in 1984 with 4,954 men enrolled.[35] Two more enrollments have taken place: one in 1987-1991 (668 additional men) and another in 2001-2003 (1,350 additional men). We only use data from the first two enrollments. Since data is semi-annual each period t corresponds to 6 months. Below we describe the main variables we use in our study:

Health (h_{it}): at every visit individuals undertake a physical examination that includes a blood sample which provides a measure of underlying health status: the individual's CD4 count. We denote as h_{it} the CD4 count at of the individual at the start of period t. According to the official U.S. government's website for HIV:[36]

> The CD4 count is [...] a snapshot of how well your immune system is functioning. CD4 cells (also known as CD4+ T cells) are white blood cells that fight infection. [...] These are the cells that the HIV virus kills. As HIV infection progresses, the number of these cells declines. When the CD4 count drops below 200 [cells per microliter] due to advanced HIV disease, a person is diagnosed with AIDS. A normal range for CD4 cells is about 500-1,500.

Ailments (y_{1it}): starting at visit 4, individuals are asked about physical symptoms. We focus on unusual bruises lasting at least two weeks, unintentional weight loss of at least 10 pounds, fatigue, diarrhea, fever, night sweats, and tender/enlarged glands. The last 5 ailments must be felt for at least 3 days during the period. Although individuals are asked explicitly about side effects starting at visit 13, we choose not to use this part of the data because it lacks consistency over time and more importantly, because individuals are most likely unable to correctly distinguish between side effects and symptoms. Thus, in our model y_{1it} takes the value of 1 if an individual reports having any of the problems mentioned above.

[35] Data in this manuscript were collected by the Multi-Center AIDS Cohort Study with centers (Principal Investigators) at The Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa P. Jacobson), Howard Brown Health Center, Feinberg School of Medicine, Northwestern University, and Cook County Bureau of Health Services (John P. Phair, Steven M. Wolinsky), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles R. Rinaldo). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute. UO1-AI-35042, 5-MO1-RR-00052 (GCRC), UO1-AI-35043, UO1-AI-35039, UO1-AI-35040, UO1-AI-35041. Website located at http://www.statepi.jhsph.edu/macs/macs.html.

[36] See https://www.hiv.va.gov/patient/diagnosis/labs-CD4-count.asp

Labor supply (y_{2it}): whether the individual worked full time (35 hours or more per week) during period t.

Income (y_{3it}): starting at visit 14, individuals answer the question "*Which of the following categories describes your annual individual gross income before taxes?*" For visit 14, categories are brackets that increase every $10,000, the last category being censored at "$70,000 or more." For visits 15 to 35 the brackets are censored at $50,000 and for visits 36 to 41 the brackets are censored at $60,000. We censor at $50,000 to obtain a uniform question over time. Then we assign the middle point to individuals in the bracket. For the highest bracket we assign the upper limit ($50,000). We divide gross income by two since our periods are half-years. Gross income as well as out-of-pocket expenditures (below) are in constant dollars of 2000.

Out-of-pocket expenditures (y_{4it}): starting at visit 14, individuals are asked a version of the following question "*Please, estimate the TOTAL out-of-pocket expenses that you or other personal sources (your lover, family or friends) paid for prescription medications since your last visit*." This question is open so values are not categorized.

Demographics (a_{it}): individuals are either white, black or Hispanic, and their age increases by half a year every period.

A.1 Products and Product Components

Starting at visit 6 individuals are asked about their medication. From visit 13 forward, as the number of treatments available increase, they answer separate survey modules for antiretroviral drugs (ARVs) and non antiretroviral drugs (NARVs). We focus on ARVs since these are the drugs used to treat HIV infection. Below we provide the empirical definition of trial and market products that we use in the paper.

Trial Products. Individuals are asked to name specifically which drugs they took as well as whether or not they took the drug as part of a research study. In the original data, some of the reported drugs are themselves coded as trials. We regard these instances as individuals participating in trials. If an individual consumes one of his drugs as part of a trial we regard the individual as consuming a trial product in that period.

Market Products. We define a market product as a combination of components where no component is consumed in trial. This definition generates $1,835$ products. We reduce the number of market products using the following algorithm:

1. We start with the set of treatments that have more than 40 observations in the sample and denote this the set of "core market products."[37] Our core market products are listed in Table

[37] We tried different criteria for the minimum number of observations and product classification did not change

S1 which shows that there are 70 core market products overall with at most five components. Out of 20,142 subject-visit observations of individuals taking market products, 13,767 are covered by treatments classified as core market products.

2. We code the remaining 6,375 observations of non-core market products as core market products using the steps below. Each step sequentially assigns the remaining observations that were not assigned in previous steps.

   (a) Non-core market product k is assigned to core market product k' if k' is the core market product with the highest number of components that is contained by k. Of the remaining 6,375 observations of non-core market products, this rule assigns 2,963 uniquely and leaves 3,412 with unassigned (1,647 that were assigned to multiple core market products plus 1,765 that were not assigned to any core market product).

   (b) If assigned to multiple core market products in step (a):

      i. First, we use the past history of the individual. If at period t the individual is consuming non-core market product k'' that was assigned to both core market products k and k' in step (a), and he was observed consuming core market product k in period $t-1$, then his treatment at t is recoded as k. We repeat this procedure until no further gains are obtained. Out of the remaining 1,647 observations assigned to multiple core market products, 428 are assigned uniquely in this step.

      ii. Second, we use the future history of the individual. If at period t the individual is consuming non-core market product k'' that was assigned to both core market products k and k' in step (a), and he was observed consuming core market product k' in period $t+1$, then his treatment at t is recoded as k'. We repeat this procedure until no further gains are obtained. Out of the remaining 1,219 observations assigned to multiple core market products, 274 are assigned uniquely in this step.

      iii. Third, we use the core market product with the highest share at t. If at period t the individual is consuming non-core market product k'' that was assigned to both core market products k and k' in step (a), and $s_{kt} > s_{k't}$, then his treatment at t is recoded as k. This final step assigns uniquely the remaining 945 observations assigned to multiple core market products.

   (c) If not assigned to a core market product in step (a): we regard all 1,765 observations as "fringe treatments" since they do not contain any core market product. We aggregate all fringe treatments that appear at period t into one single "fringe mix," and assign to it

substantially. Since our definition of core market products can miss treatments appearing near the end of the time period studied, we select the core products using all periods but exclude the last 4 periods from estimation.

all users consuming this product over time. We only consider fringe mixes that have at least 40 users. This reduces the number of observations by 345 (which represents 1.6% of the number of observations of individuals using a treatment). This aggregation leads to 16 fringe mixes that we pool with the set of core market products, which amounts to a total of 86 market products overall. (See Table S1.)

3. In the paper we specified that a treatment gets withdrawn from the market altogether when its share falls below $\bar{s}$ for 2 consecutive periods. However, in the data, a treatment may have a share below $\bar{s}$ for more than 2 consecutive periods and then reappear again. 78 out of 86 core market products have unique spells without "reappearance." We regard the remaining treatments with multiple spells as measurement error and follow the next procedure to ensure that treatments have unique spells without reappearance. For every core market product k with reappearance:

   (a) We identify all spells that treatment k has in the data. This is, we identify the first spell and all reappearances.

   (b) From those spells we select the one that contains the period t' in which s_{kt} was the highest. We drop all observations of individuals consuming market product k in other spells.

   Out of 19,797 (20,142 minus 345 from step 2(c)) observations of individuals taking market products, this smoothing procedure drops 42 observations leaving 19,755 observations of individuals taking market products. Supporting the importance of the spells selected by this procedure, the maximum share in the selected spell is on average about 24 times larger that the maximum share in other spells of the same market product.[38] Table S1 includes entry and exit dates implied by this spell smoothing procedure.

B Estimation Appendix

B.1 Product Characteristics

We estimate product characteristics using the larger sample (visits 6 to 49) thereby using all data available on previous health, individual treatment usage, and subsequent health and ailments. Es-

[38] In addition to this procedure we tried (i) selecting the spell with the highest average share and (ii) selecting the spell with the highest sum of shares. All criteria result in very similar entry and exit dates.

APPENDIX TABLE S1: Market Products

Market Product	Entry	Exit	Market Product	Entry	Exit
AZT	1987 S1	-	ddI , d4T, Nevirapine	1997 S2	-
Interferons (α and/or β), AZT	1987 S2	1995 S2	ddI , 3TC, Nelfinavir	1997 S2	-
AL-721 egg lecithin	1987 S2	1991 S2	ddI , d4T, Efavirenz	1998 S2	2008 S1
AZT, Acyclovir	1989 S2	2000 S1	3TC, Abacavir, Efavirenz	1998 S2	-
Acyclovir	1989 S2	2000 S1	AZT, Nevirapine, 3TC, Abacavir	1999 S1	-
AZT, Acyclovir, ddI	1990 S1	1997 S1	AZT, 3TC, Abacavir, Efavirenz	1999 S1	-
Acyclovir, ddI	1990 S1	2000 S1	AZT, 3TC, Efavirenz	1999 S1	-
AZT, ddC	1990 S1	2001 S2	AZT, 3TC, Abacavir	1999 S1	-
AZT, ddI	1990 S1	2004 S2	d4T, 3TC, Efavirenz	1999 S1	2006 S1
ddI	1990 S1	-	Nevirapine, 3TC, Abacavir	1999 S2	-
AZT, ddC, Acyclovir, ddI	1991 S1	1997 S1	d4T, 3TC, Kaletra	2001 S1	2006 S1
AZT, ddC, Acyclovir	1991 S1	1999 S2	3TC, Kaletra, Abacavir	2001 S2	-
AZT, ddC, ddI	1991 S1	1995 S2	AZT, 3TC, Kaletra	2001 S2	-
ddC, Acyclovir	1991 S1	1997 S2	AZT, 3TC, Kaletra, Abacavir	2002 S1	-
ddC	1991 S1	1999 S1	3TC, Abacavir, Efavirenz, Tenofovir	2002 S1	-
d4T	1993 S1	-	AZT, 3TC, Abacavir, Tenofovir	2002 S1	-
AZT, Acyclovir, 3TC	1994 S2	2000 S1	AZT, 3TC, Kaletra, Tenofovir	2002 S1	-
AZT, 3TC	1995 S1	-	Nevirapine, 3TC, Tenofovir	2002 S1	2007 S1
Acyclovir, d4T, 3TC	1995 S2	2000 S1	3TC, Kaletra, Tenofovir	2002 S1	-
AZT, 3TC, Saquinavir	1996 S1	2005 S1	Kaletra, Efavirenz, Tenofovir	2002 S1	-
d4T, 3TC	1996 S1	-	3TC, Efavirenz, Tenofovir	2002 S1	-
AZT, 3TC, Saquinavir, Ritonavir	1996 S2	-	AZT, 3TC, Kaletra, Abacavir, Tenofovir	2002 S2	-
AZT, Acyclovir, 3TC, Indinavir	1996 S2	2000 S1	ddI , Kaletra, Tenofovir	2002 S2	-
Acyclovir, d4T, 3TC, Indinavir	1996 S2	2000 S1	ddI , Efavirenz, Tenofovir	2002 S2	-
AZT, 3TC, Ritonavir, Indinavir	1996 S2	2006 S2	Abacavir, Efavirenz, Tenofovir	2002 S2	-
d4T, 3TC, Ritonavir, Indinavir	1996 S2	2006 S2	Kaletra, Abacavir, Tenofovir	2002 S2	-
d4T, 3TC, Saquinavir, Ritonavir	1996 S2	2004 S2	3TC, Ritonavir, Abacavir, Atazanavir	2003 S2	-
ddI , d4T, Indinavir	1996 S2	2004 S2	Efavirenz, Tenofovir, Emtricitabine	2003 S2	-
d4T, 3TC, Indinavir	1996 S2	2008 S1	Ritonavir, Efavirenz, Tenofovir, Emtricitabine, Atazanavir	2004 S1	-
AZT, 3TC, Indinavir	1996 S2	-	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir	2004 S1	-
d4T, Nevirapine, 3TC	1997 S1	-	ddI , Ritonavir, Tenofovir, Atazanavir	2004 S1	-
AZT, Nevirapine, 3TC	1997 S1	-	Ritonavir, Tenofovir, Emtricitabine, Atazanavir	2004 S1	-
AZT, 3TC, Nelfinavir	1997 S1	-	Nevirapine, Tenofovir, Emtricitabine	2004 S1	-
ddI , d4T, Nelfinavir	1997 S1	2005 S2	Kaletra, Tenofovir, Emtricitabine	2004 S2	-
d4T, 3TC, Nelfinavir	1997 S2	-	Ritonavir, Tenofovir, Emtricitabine, Lexiva	2005 S1	-
Fringe Mixes					
Isoprinosine, Ribavirin, Interferons (α and/or β)	1987 S1	1992 S1	Nevirapine, 3TC, Ritonavir, Kaletra, Tenofovir	2003 S1	-
Interferons (α and/or β), 3TC, Saquinavir, Indinavir, Efavirenz	1997 S1	2007 S1	3TC, Ritonavir, Kaletra, Abacavir, Tenofovir, Atazanavir	2004 S1	-
Nevirapine, 3TC, Saquinavir, Ritonavir, Indinavir	1997 S2	2006 S2	Ritonavir, Tenofovir, Emtricitabine, Atazanavir, Lexiva	2004 S2	-
Nevirapine, 3TC, Saquinavir, Ritonavir, Nelfinavir	1998 S1	2006 S2	Saquinavir, Ritonavir, Tenofovir, Emtricitabine, Atazanavir	2005 S1	-
Nevirapine, Saquinavir, Ritonavir, Abacavir, Efavirenz	1999 S1	2005 S2	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir, Lexiva	2005 S2	-
Nevirapine, Ritonavir, Nelfinavir, Abacavir, Efavirenz	1999 S2	-	Saquinavir, Ritonavir, Abacavir, Tenofovir, Emtricitabine	2007 S1	-
Nevirapine, Ritonavir, Kaletra, Abacavir, Efavirenz	2001 S2	2008 S2	3TC, Ritonavir, Tenofovir, Emtricitabine, Raltegravir	2008 S1	-
Nevirapine, 3TC, Nelfinavir, Abacavir, Tenofovir	2002 S2	-	Ritonavir, Tenofovir, Emtricitabine, Darunavir, Raltegravir	2008 S2	-

Notes: Entry and exit dates implied by the smoothing of spells in Step 3 of the algorithm used to reduce market products in Section A.1. S1 and S2 indicate the semester within a year. Many products had not exited by the end of the sample. For *Fringe Mixes* we only include the 5 or 6 most used products in the mix.

timation equations follow from (20) and (21):

$$h_{t+1} = \sum_{s=0}^{5} \gamma_s^h h_t^s + \sum_{k \in \mathbf{P}_t} \tilde{d}_{kt} \theta_k^1 + d_{J+1,t} \theta_{et}^1 + \varepsilon_t^h \tag{S1}$$

$$\Pr[y_{1t} = 0 | h_t, \theta] = \left(1 + \exp\left(\sum_{s=0}^{5} \gamma_s^x h_t^s + + \sum_{k \in \mathbf{P}_t} \tilde{d}_{kt} \theta_k^2 + d_{J+1,t} \theta_{et}^2 \right) \right)^{-1} \tag{S2}$$

Along with estimates of product characteristics, (S1) and (S2) provide parameter vectors γ^h and γ^x that describe the health transition in (20) and the process for physical ailments in (21).

B.2 Clusters

In our empirical implementation we assume there are J clusters every period. We implement the following version of the k-means algorithm. At every period t:

1. Select the products for which the $\underline{s}$ rule has not been applied. In other words, select products that are still available for new consumers at t. Denote this set of products $\mathbf{A}_t$.

2. In order to keep comparability, re-scale the characteristics of all products available for clustering at t by computing:
$$\tilde{\theta}_k^r = \frac{\theta_k^r}{\max_{k \in \mathbf{A}_t} \left| \theta_k^r \right|}, \text{ for } r = 1, 2 \tag{S3}$$
Thus, by construction $\tilde{\theta} \in [-1,1] \times [-1,1]$.

3. Select the first J centroids using the scaled characteristics $\tilde{\theta}$ of J randomly selected products from $\mathbf{A}_t$.

4. Allocate all remaining products $k \in \mathbf{A}_t$ to clusters sequentially. At each step select for allocation the product whose scaled characteristics $\tilde{\theta}_k$ are closest to one of the existing clusters. Assign product k to the closest cluster and update the centroid of the cluster. Repeat this process until all products in $\mathbf{A}_t$ are assigned to a cluster.

5. Taken the centroids as given, reallocate all products to their closest centroid.

6. Calculate the value of the clustering rule $C\left(\{\theta_k\}_{k \in \mathbf{P}_t}\right)$ in (17) for the current allocation.

7. Repeat 200 times steps 3 to 6 using the scaled characteristics $\tilde{\theta}$ of different groups of J randomly selected products in $\mathbf{A}_t$ as initial centroids. The allocation with the lowest value of $C\left(\{\theta_k\}_{k \in \mathbf{P}_t}\right)$ is chosen.[39]

[39] In estimation, whenever we simulate clusters we only repeat the process 50 times.

B.3 Innovation

According to (13), the characteristics of new products and trial products are displaced innovations about the centroid (current or previous), and depend on previous trial participation and a draw from the distribution of innovation shocks $f_V(v)$. To estimate (13) and $f_V(v)$ we use all periods in the MACS data with relevant information on treatment consumed, health and ailments (1986 to 2008). Over the time span in our data, and given our definition of products, we observe 86 realized innovations from newly introduced market products and 22 realized innovations from trials products. Consistent with our definition of market products, we only consider trial products that entail at least 40 users. We do not impose that innovations vectors cannot be strictly negative. In other words, relative to the centroid, inferior products with lower quality in both dimensions (health and ailments) may be introduced.[40]

B.4 Utility Parameters

We estimate the utility parameters in (26) using a GMM estimator and moment conditions that equate the log odds ratio of current conditional choice probabilities with a representation of the differences in conditional value functions in terms of utility parameters and future CCPs, states and choices (Hotz et al., 1994; Altuğ and Miller, 1998). Below we explain this step of the estimation process in more detail.

B.4.1 Moment Condition

Our moment conditions appeal to well-known results following from our assumption that the taste shocks ε_{jit} are iid Extreme Value Type I distributed (Hotz and Miller, 1993). They rely on differences between the log odds ratio and an alternative representation of differences in conditional value functions $(v_j(z_{it}) - v_0(z_{it}))$ in terms of future conditional choice probabilities, choices, states and utility parameters. Recalling the definition of $V(z_{it})$ in (10), the conditional value function of choosing alternative j at period t is:

$$v_j(z_{it}) = E\left\{u_j(h_{it}, y_{it}) + \beta V(z_{it+1}) \,\middle|\, z_{it}, d_{jit} = 1\right\} \tag{S4}$$

Let $p_{jit}(z_{it})$ be the probability that individual i chooses option j at time t conditional on his state z_{it}. Let $\psi_{jit}(z_{it})$ be the expected value of the j^{th} taste shock conditional on alternative j being

[40] This is consistent with what we observe in the data, and theoretical reasons why this may happen have been provided in the literature (Miller, 1988).

optimal, and let γ be the Euler constant. Since the joint distribution of ε_t is Extreme Value Type-I:

$$\psi_j(z_{it}) \equiv E_\varepsilon\left[\varepsilon_{jit}|z_{it}, d^e_{jit}=1\right] = \gamma - \ln\left(p_{jit}(z_{it})\right) \tag{S5}$$

Define $E_j\{\cdot\}$ as the expectation conditional on $d_{jit}=1$. Dropping the individual subindex i for simplicity, using (S5), we can write the conditional value function in (S4) in terms of future utility flows induced by all available alternatives, weighted by the future probabilities of those alternatives being chosen and corrected by the fact that the alternative may not be optimal. Notably, the weighted average of corrected flow payoffs of a given period must be discounted by the probability of survival up to that period conditional on today's state and choice. Letting T^* be an arbitrary period with $t < T^* \leq T$, he alternative representation of the conditional value function is given by:

$$\begin{aligned}
v_{jt}(z_t) &= E_j\{u_j(h_t,y_t)|z_t\} + \beta E_j\{V(z_{t+1},\varepsilon_{t+1})|z_t\} \\
&= E_j\{u_j(h_t,y_t)|z_t\} + \beta E_j\left\{ b_{t+1} E_\varepsilon\left\{ \sum_{j'=0}^{J+1+r_{t+1}} d^e_{j't+1}\left[u_{j'}(h_{t+1},y_{t+1}) + \psi_{j'}(z_{t+1})\right]\right\}\middle| z_t\right\} \\
&\quad + \beta^2 E_j\{b_{t+2}V(z_{t+2},\varepsilon_{t+2})|z_t\} \\
&= E_j\{u_j(h_t,y_t)|z_t\} + \beta E_j\left\{ b_{t+1} \sum_{j'=0}^{J+1+r_{t+1}} p_{j't+1}(z_{t+1})\left[u_{j'}(h_{t+1},y_{t+1}) + \psi_{j'}(z_{t+1})\right]\middle| z_t\right\} \\
&\quad + \beta^2 E_j\{b_{t+2}V(z_{t+2},\varepsilon_{t+2})|z_t\} \\
&= E_j\{u_j(h_t,y_t)|z_t\} + \beta E_j\left\{ b_{t+1} \sum_{j'=0}^{J+1+r_{t+1}} p_{j't+1}(z_{t+1})\left[u_{j'}(h_{t+1},y_{t+1}) + \psi_{j'}(z_{t+1})\right]\middle| z_t\right\} \\
&\quad + \beta^2 E_j\left\{ b_{t+1}b_{t+2} \sum_{j'=0}^{J+1+r_{t+2}} p_{j't+2}(z_{t+2})\left[u_{j'}(h_{t+2},y_{t+2}) + \psi_{j'}(z_{t+2})\right]\middle| z_t\right\} \\
&\quad + \beta^3 E_j\{b_{t+1}b_{t+2}V(z_{t+3},\varepsilon_{t+3})|z_t\} \\
&= E_j\{u_j(h_t,y_t)|z_t\} + \sum_{\tau=1}^{T^*}\beta^\tau E_j\left\{ \left(\prod_{r=1}^{\tau} f_b(h_{it+r})\right) \sum_{j'=0}^{J+1+r_{t+\tau}} p_{j't+\tau}(z_{t+\tau})\left[u_{j'}(h_{t+\tau},y_{t+\tau}) + \psi_{j'}(z_{t+\tau})\right]\middle| z_t\right\} \\
&\quad + \beta^{T^*+1} E_j\left\{ \left(\prod_{r=1}^{T^*+1} f_b(h_{it+r})\right) V(z_{t+T^*+1},\varepsilon_{t+T^*+1})\middle| z_t\right\}
\end{aligned} \tag{S6}$$

Let $w(z_{it})$ be a vector of instruments orthogonal to the difference between the log odds ratio and the alternative representation. Hence, we can form the following moment conditions:

$$\mathbb{E}\left\{ w(z_{it}) \otimes \begin{bmatrix} \ln\left(\frac{p_{0it}(z_{it})}{p_{1it}(z_{it})}\right) + v_{1it}(z_{it}) - v_{0it}(z_{it}) \\ \vdots \\ \ln\left(\frac{p_{0it}(z_{it})}{p_{J+1+r_{it},it}(z_{it})}\right) + v_{J+1+r_{it},it}(z_{it}) - v_{0it}(z_{it}) \end{bmatrix} \right\} = 0. \tag{S7}$$

B.4.2 Conditional Choice Probabilities

The individual's choice set $\{0,1,\ldots,J+1+r_{it}\}$ includes the following alternatives: no product, one of J clusters, a trial product, and last-period's product (if $r_{it}=1$). The probability that an individual chooses one of the alternatives depends on the individual and aggregate elements of his state, where the aggregate state is given by $z_t = \{\{\theta_k\}_{k\in\mathbf{P}_t}, \omega_t, \kappa_t, s_{et-1}, \mathscr{F}_t\}$. In estimation we include ω_t, κ_t and s_{et-1} directly into the CCPs and characterize other components of z_t as follows. The set of products available is $\{\theta_k\}_{k\in\mathbf{P}_t}$ is characterized by the distribution of product characteristics of all clusters. We use the first two moments of these distributions in estimation. The distribution of consumer characteristics $\mathscr{F}_t$ is controlled for using a set of non parametric moments denoted $\tilde{\mathscr{F}}_t$.[41] Let m_{jit} be the moments describing the distribution of characteristics induced by alternative j for individual i at period t—mean vector and the variance matrix. Effectively, m_{jit} is heterogeneous across individuals only when $j=J+2$, i.e. when the individual decides to purchase the same product he consumed last period. Let $m_{jit}m_{jit}$ denote a vector of interactions between the elements of m_{jit}. Let $\tilde{x}_{it}$ and $\tilde{z}_{it}$ be subsets of the individual-specific components of the state.[42] Let $\omega_t m_{jit}$ denote a vector of interactions between the centroid and the elements of m_{jit}. Similarly, let $m_{jit}\tilde{z}_{it}$ be a vector of interactions between the components of m_{jit} and individual-specific state components and let $\omega_t m_{jit}\tilde{z}_{it}$ be defined in a similar fashion. Our flexible CCPs are given by:

$$p_{jit} = \frac{\exp\left(I_{jit}\right)}{\sum_{j'=0}^{J+1+r_{it}} \exp\left(I_{j'it}\right)} \tag{S8}$$

where

$$I_{0it} \equiv 0 \tag{S9}$$

$$I_{jit} \equiv \gamma_J \tilde{x}_{it} + \beta_0 m_{jt} + \beta_1 m_{jt} m_{jt} + \beta_2 \omega_t m_{jt} + \beta_3 m_{jt}\tilde{z}_{it} + \beta_4 \omega_t m_{jt}\tilde{z}_{it} + \beta_5 m_{jt}\tilde{\mathscr{F}}_t + \beta_6 \kappa_t + \beta_7 s_{et-1}, \quad j=1,\ldots,J \tag{S10}$$

$$I_{J+1,it} \equiv \gamma_{J+1}\tilde{x}_{it} + \beta_0 m_{J+1,t} + \beta_1 m_{J+1,t} m_{J+1,t} + \beta_3 m_{J+1,t}\tilde{z}_{it} + \beta_5 m_{J+1,t}\tilde{\mathscr{F}}_t + \beta_6 \kappa_t + \beta_7 s_{et-1} \tag{S11}$$

$$I_{J+2,it} \equiv \gamma_{J+2}\tilde{x}_{it} + \beta_0 m_{J+2,it} + \beta_1 m_{J+2,it} m_{J+2,it} + \beta_2 \omega_t m_{J+2,it} + \beta_3 m_{J+2,it}\tilde{z}_{it} + \beta_4 \omega_t m_{J+2,it}\tilde{z}_{it} + \beta_5 m_{J+2,it}\tilde{\mathscr{F}}_t + \beta_6 \kappa_t + \beta_7 s_{et-1} \tag{S12}$$

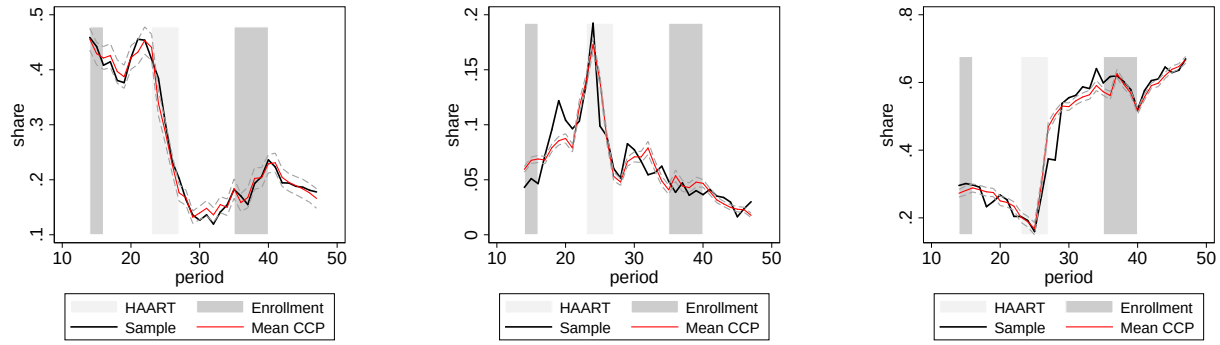
Although the characteristics of the choice sets are non stationary due to product entry and exit, by interacting our time-varying regressors $\tilde{z}_{it}$ with the characteristics of the choice for individual i, m_{jit}, we are able to control for the state of the world inside the CCPs.[43] This procedure gives us CCPs for any simulated world as long as our observed worlds cover the space of possible worlds. Additionally, we include in the CCPs ancillary coefficients that are unrelated to the state of technology, denoted γ in (S10) to (S12), which capture stationary taste differences between alternatives. Because, conditional on cluster characteristics, all clusters are equivalent to "trying a new market product," we impose $\gamma_j = \gamma_J =$ for any $j=1,\ldots,J$.

[41] We specify these moments as shares of people with different sets of characteristics.

[42] $\tilde{z}_{it}$ includes h_{it-1}, a_{it-1}, b_i, y_{2it} while $\tilde{x}_{it}$ includes a constant, a_{it-1}, b_i.

[43] Because some of the components of m_{J+1t} are linear functions of ω_{t-1} (see (13)) we avoid collinearity by not including terms $\omega_t m_{J+1,t}$ and $\omega_t m_{J+1,t}\tilde{z}_{it}$ in (S11).

Figure S1 displays the mean predicted conditional choice probability using (S8) over time against the correspondent share of the population who chose the alternative.[44]



(a) No Product, $j = 0$ (b) Trial Product, $j = J+1$ (c) Repeat, $j = J+2$

APPENDIX FIGURE S1: Average CCPs

Notes: The figure shows the average estimated conditional choice probability against the share of people choosing the alternative. Dashed lines represent 95% confidence intervals around the predicted CCPs. Three periods of special relevance are highlighted in the Figure: two periods during which enrollment into the sample was undertaken and the period in which products belonging to the HAART class were introduced.

B.4.3 Simulation

In order to form the sample analog of the moment condition in (S7) we obtain a simulated version of the conditional value function in (S6) truncated at T^* for every observation $\{i,t\}$ and alternative $j \in \{0,1,\ldots,J+1+r_{it}\}$. We select $T^* = 10$ so that the product $\beta^{T^*+1}\prod_{r=1}^{T^*+1} f_b\left(h_{it+r}\right)$ approaches zero, eliminating further differences in conditional value functions beyond T^*. Let S denote the number of simulated paths for each $\{j,i,t\}$ and let the superscript s indicate that a quantity is simulated. For individual i and alternative j at period t we write the simulated counterpart of the truncated value function as

$$\bar{v}_{jit}\left(z_{it}\right) \equiv \frac{1}{S}\sum_{s=1}^{S}\left\{u_j\left(h_{it},y_{it}^s\right) + \sum_{\tau=1}^{T^*}\beta^{\tau}\left(\prod_{r=1}^{\tau} f_b\left(h_{it+r}^s\right)\right)\sum_{j'=0}^{J+1+r_{t+\tau}} d_{j'it+\tau}^s\left[u_{j'}\left(h_{it+\tau}^s,y_{it+\tau}^s\right)+\psi_{j'}\left(z_{it+\tau}^s\right)\right]\right\} \tag{S13}$$

Each future path depends on the current individual state z_{it}, and hence on the current aggregate state z_t, and the current choice j. We first simulate as many aggregate paths at t as there are individuals at period t. Overall this yields IT paths of technological innovation. Then, because individuals are atomistic, for each observation $\{i,t\}$ and alternative j we generate sequences of future choices and payoffs taking as given $S = 20$ artificial technological paths chosen at random

[44] We also explore the fit of our CCP estimates comparing the relative shares that clusters received in reality against the predictions from our estimated CCPs. We ranked the three clusters at every period by the share they received and compare this ranking against the ranking obtained from our estimated CCPs. Predicted ranks match observed ranks in about 80% of the periods.

from the set of I simulated technological paths that start at date t.[45] This serves two purposes. It maintains the assumption, needed for consistency of the estimator, that the sample draws from the moment conditions—the contribution from each observation—are independent from each other, and it prevents simulation errors in technology paths from propagating across all observations.

Simulation of Aggregate State. Taking as given the current aggregate state z_t we create as many simulated aggregate state paths $\{z^s_{t+\tau}\}_{\tau=1}^{T^*}$ as there are individuals at every t. In other words, we repeat the algorithm below to generate I simulated aggregate paths for every period t:

1. Let $\tau = 1$.

2. *Supply*. Simulate a number of new products at $t+\tau$, $New^s_{t+\tau}$, using the entry process in (14). If $New^s_{t+\tau} > 0$, for each simulated new product draw simulated characteristics using (13). Simulate the characteristics of the trial product using (13). Obtain $\kappa^s_{t+\tau}$ using (2) and (15). For all incumbent products, apply the exit rule $\{\underline{s}, \bar{s}\}$ taking into account the extent to which it binds according to (16). From the simulated set of products in $\mathbf{P}^s_{t+\tau}$ that have not yet satisfied the $\underline{s}$ exit rule, form clusters following the clustering rule in (17). Obtain the distribution of characteristics of each cluster using (18) and (19). For $\tau > 1$ compute the simulated centroid $\omega^s_{t+\tau}$ using (1).

3. *Demand*. For all individuals i' at t: If $\tau = 1$, define $h^s_{i't+1} \equiv h_{i't+1}$ and $d^s_{i't} \equiv d_{i't}$, otherwise, simulate $h^s_{i't+\tau}$ using (20). Draw a simulated labor state $y^s_{2i't+\tau}$ using (22). Compute deterministic transitions (e.g. age). Using $z^s_{i't+\tau}$, and hence $z^s_{t+\tau}$, and (S8) to (S12) compute simulated CCPs $p^s_{ji't+\tau}\left(z^s_{i't+\tau}\right)$ for every alternative $j \in \{0, 1, \ldots, J+1+r^s_{it+\tau}\}$ and draw a decision $d^s_{i't+\tau}$. Obtain the simulated share of trial participation $s^s_{e,t+\tau}$ and the nonparametric representation of the simulated distribution of consumer characteristics $\tilde{\mathscr{F}}^s_{t+\tau}$.

4. *Cycle back*. If $\tau = T^*$ end the loop. Otherwise, let $\tau = \tau + 1$ and go back to step 2.

Simulation of Individual Paths. For every observation $\{i,t\}$ and every alternative $j \in \{0, 1, \ldots, J+1+r_{it}\}$ we generate S sequences of future states, choices and outcomes $\{z^s_{it+\tau}, d^s_{it+\tau}, y^s_{it+\tau}\}_{\tau=1}^{T^*}$ taking as given a subset of S simulated aggregate paths—that start at t—chosen at random without replacement. We follow the steps below:

1. Let $\tau = 1$.

[45] Notice that we could rely on Hotz et al. (1994) and set $S = 1$ and obtain consistency of our estimator. However, we choose $S = 20$ after trying different values for robustness.

2. *Demand*. Same as above but only for individual i. When j is not equal to the observed choice for $\{i,t\}$, we also simulate health at the beginning of period $t+1$. For this we back out the realized health residual using (S1) and use (20) to simulate health h^s_{it+1} under counterfactual choice j. Additionally, we compute the simulated one-period-ahead survival probability $f_b\left(h^s_{it+\tau}\right)$.

3. *Outcomes*. Only for individual i: Simulate (lack of) ailments using (21) and the relevant distribution of product characteristics implied by the simulated choice $d^s_{it+\tau}$. Simulate income using (23) and out-of-pocket expenditures using (24).[46]

4. *Cycle back*. If $\tau = T^*$ end the loop. Otherwise, let $\tau = \tau + 1$ and go back to step 2.

When simulating a path following an alternative j that is not the observed choice for $\{i,t\}$, we obtain current-period simulated payoffs $u_j\left(h^s_{it}, y^s_{it}\right)$ by simulating current income, out-of-pocket expenditures and ailments conditional on the counterfactual choice j at t.

B.4.4 Estimator

Let $j=0$ be the base alternative, and let δ_{it} be an indicator of whether individual i is in the data at period t. The simulated sample analog of the moment condition in (S7) is

$$\frac{1}{\sum_i \sum_t \delta_{it}} \sum_{i=1}^{I} \sum_{t=1}^{T} \delta_{it} w(z_{it}) \otimes \begin{bmatrix} \ln\left(\frac{p_{0it}(z_{it})}{p_{1it}(z_{it})}\right) + \bar{v}_{1it}(z_{it}) - \bar{v}_{0it}(z_{it}) \\ \vdots \\ \ln\left(\frac{p_{0it}(z_{it})}{p_{J+1+r_{it},it}(z_{it})}\right) + \bar{v}_{J+1+r_{it},it}(z_{it}) - \bar{v}_{0it}(z_{it}) \end{bmatrix} = 0 \tag{S14}$$

Denote Λ as the $M-$dimensional vector of parameters of the utility function. Following Hotz et al. (1994) we estimate Λ as the vector that minimizes the following objective function:

$$\left((IT)^{-1} \sum_{i=1}^{I} \sum_{t=1}^{T} \delta_{it} w(z_{it}) \otimes A_{it}(z_{it}, \Lambda)\right)' W_n \left((IT)^{-1} \sum_{i=1}^{I} \sum_{t=1}^{T} \delta_{it} w(z_{it}) \otimes A_{it}(z_{it}, \Lambda)\right) \tag{S15}$$

$$A_{it}(z_{it}, \Lambda) \equiv \begin{bmatrix} \ln\left(\frac{p_{0it}(z_{it})}{p_{1it}(z_{it})}\right) + \bar{v}_{1it}(z_{it}) - \bar{v}_{0it}(z_{it}) \\ \vdots \\ \ln\left(\frac{p_{0it}(z_{it})}{p_{J+2it}(z_{it})}\right) + \bar{v}_{J+2it}(z_{it}) - \bar{v}_{0it}(z_{it}) \end{bmatrix} \tag{S16}$$

[46] Even though individuals know their idiosyncratic income shocks ε^m_{it} we do not need to simulate these shocks as they are iid, have mean zero, and enter linearly in the flow utility, which results in them averaging out to zero in the moment condition.

where W_n is a square weighting matrix. Using the linear structure of the utility function in (26) we collect and factor terms in order to write the jth component of the vector $A_{it}(z_{it},\Lambda)$ as the linear form

$$\tilde{y}_{jit} - \tilde{x}'_{jit}\Lambda \tag{S17}$$

Define Y as a vector with $(J+2)IT$ rows that stacks all $\tilde{y}_{jit}$, and X as a $(J+2)IT \times M$ matrix that stacks all $\tilde{x}_{jit}$. Define Z as the $IT \times R$ matrix whose columns contain the R instruments orthogonal to the difference between the log odds ratio of current conditional choice probabilities and the alternative representation of the differences in conditional value functions.[47] Thus

$$Y = \begin{bmatrix} \tilde{y}_{1,1,1} \\ \tilde{y}_{1,1,2} \\ \vdots \\ \tilde{y}_{1,I,T-1} \\ \tilde{y}_{1,I,T} \\ \vdots \\ \tilde{y}_{J+2,1,1} \\ \tilde{y}_{J+2,1,2} \\ \vdots \\ \tilde{y}_{J+2,I,T-1} \\ \tilde{y}_{J+2,I,T} \end{bmatrix}, \; X = \begin{bmatrix} \tilde{x}_{1,1,1,1} & \cdots & \tilde{x}_{1,1,1,M} \\ \tilde{x}_{1,1,2,1} & \cdots & \tilde{x}_{1,1,2,M} \\ \vdots & & \vdots \\ \tilde{x}_{1,I,T-1,1} & \cdots & \tilde{x}_{1,I,T-1,M} \\ \tilde{x}_{1,I,T,1} & \cdots & \tilde{x}_{1,I,T,M} \\ \vdots & & \vdots \\ \tilde{x}_{J+2,1,1,1} & \cdots & \tilde{x}_{J+2,1,1,M} \\ \tilde{x}_{J+2,1,2,1} & \cdots & \tilde{x}_{J+2,1,2,M} \\ \vdots & & \vdots \\ \tilde{x}_{J+2,I,T-1,1} & \cdots & \tilde{x}_{J+2,I,T-1,M} \\ \tilde{x}_{J+2,I,T,1} & \cdots & \tilde{x}_{J+2,I,T,M} \end{bmatrix}, \; Z = \begin{bmatrix} w(z_{11})_1 & \cdots & w(z_{11})_R \\ w(z_{12})_1 & \cdots & w(z_{12})_R \\ \vdots & & \vdots \\ w(z_{IT})_1 & \cdots & w(z_{IT})_R \end{bmatrix} \tag{S18}$$

Finally, let $\mathbf{I}_{[J+2]}$ be a $(J+2)$-dimensional identity matrix and define $\tilde{Z} \equiv \mathbf{I}_{[J+2]} \otimes Z$. Then we can write the objective function in (S15) as

$$\left((IT)^{-1}\tilde{Z}'(Y - X\Lambda)\right)' W_n \left((IT)^{-1}\tilde{Z}'(Y - X\Lambda)\right) \tag{S19}$$

Equation (S19) is a linear arrangement so we can obtain a close form solution for $\hat{\Lambda}$ as the optimal GMM estimator. It entails first and second stage estimators given by

$$\hat{\Lambda}^{1S} = \left(X'\tilde{Z}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\tilde{Z}'Y\right), \qquad \hat{\Lambda}^{2S} = \left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'X\right)^{-1}\left(X'\tilde{Z}\hat{S}^{-1}\tilde{Z}'Y\right) \tag{S20}$$

where

$$\hat{S} = \frac{1}{I^*}\tilde{Z}'D\tilde{Z}, \qquad I^* = IT(J+1) + \sum_{i=1}^{I}\sum_{t=1}^{T} r_{it} \tag{S21}$$

accounts for the fact that some individuals cannot repeat their previous consumption (for instance, if the product was withdrawn), and D is the $I(J+2)$ square diagonal matrix with diagonal elements $\hat{u}^2_{jit} = \left(y_{jit} - x'_{jit}\hat{\Lambda}^{1S}\right)^2$. As instruments we use initial health h_{it}, lagged labor state y_{2it-1}, income

[47] Hence W_n is a $(J+2)R$-dimensional square matrix.

fixed effect η_i, race, education indicators, and age a_{it}, the centroid ω_t and the lagged share of trial participation s_{et-1}, as well as interactions between these variables. The variance-covariance matrix of the second stage estimator is

$$\hat{V}^{2S} = I^* \left(X' \tilde{Z} \hat{S}^{-1} \tilde{Z}' X \right)^{-1} \tag{S22}$$

B.5 Standard Errors

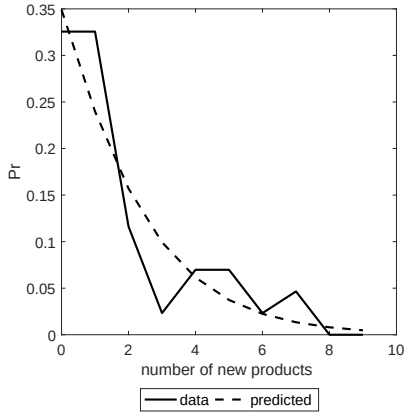
The uncorrected standard errors for our utility parameters yield from the variance-covariance matrix in (S22). In order to obtain corrected standard errors we undertake subsampling taking as given the following objects obtained from the full sample: the definition of products (i.e. what their components are, for instance, AZT or AZT + ddI), their corresponding entry and exit dates, and the exit thresholds $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ specified in Section 3. We draw $R = 100$ subsamples containing a proportion $\tilde{p} = 0.9$ of the individuals in the sample drawn without replacement, and estimate all parameters in the model using each subsample. This includes estimating product characteristics, parameters governing transition and outcome processes, and simulating forward paths of technology to obtain utility parameters. For any parameter γ with estimated value $\hat{\gamma}_r$ from the r^{th} subsample, the subsampling standard errors are obtained as

$$se(\hat{\gamma}) \approx se(\hat{\gamma}_r) \cdot \sqrt{\tilde{p}} \tag{S23}$$

where $se(\hat{\gamma}_r)$ is estimated as the standard deviation of the R quantities $\hat{\gamma}_r$.

C Results Appendix

C.1 Estimates



APPENDIX FIGURE S2: Distribution of Number of New Products

Notes: Model is specified in (14). Figure shows the empirical distribution of the number of new products and the average over time of the predicted probabilities using the estimated parameters in Table 3.

APPENDIX TABLE S2: Within Cluster Share Function

variable	coef. (γ^w)	se
Ailments Rk	-0.427	(0.124)
Ailments Rk $\times$ *Health Rk*	0.074	(0.020)
Health Rk^2	-0.029	(0.008)
Ailments Rk^2	-0.019	(0.006)
NP	-0.509	(0.048)
Health Rk $\times$ *NP*	0.046	(0.009)
Ailments Rk $\times$ *NP*	0.063	(0.010)
Ailments Rk $\times$ *Health Rk* $\times$ *NP*	-0.007	(0.002)
New	-0.352	(0.508)
New $\times$ *NP*	0.027	(0.404)
Constant	0.786	(0.121)

Notes: Parameters estimates from (18) and (19). *Rk* stands for the rank of the characteristic compared to other treatments within a cluster. *NP* is the cluster size. *New* indicates whether the product just entered the market. In parentheses, standard errors computed using subsampling with 100 subsamples.

APPENDIX TABLE S3: Product Characteristics

Market Product	Ailments, θ^2 coeff	se	Health, θ^1 coeff	se	Market Product	Ailments, θ^2 coeff	se	Health, θ^1 coeff	se
AZT	-0.500	(0.020)	-12.004	(0.736)	ddI , d4T, Nevirapine	0.753	(0.175)	44.240	(3.781)
Interferons (α and/or β), AZT	-0.600	(0.061)	-55.796	(3.102)	ddI , 3TC, Nelfinavir	-0.810	(0.083)	47.816	(6.848)
AL-721 egg lecithin	-0.433	(0.087)	-19.655	(3.917)	ddI , d4T, Efavirenz	-0.626	(0.078)	41.280	(2.772)
AZT, Acyclovir	-0.539	(0.050)	-12.752	(1.670)	3TC, Abacavir, Efavirenz	0.108	(0.047)	53.341	(1.501)
Acyclovir	-0.783	(0.047)	-0.017	(2.678)	AZT, Nevirapine, 3TC, Abacavir	0.038	(0.131)	39.379	(3.369)
AZT, Acyclovir, ddI	-0.851	(0.037)	-16.474	(1.497)	AZT, 3TC, Abacavir, Efavirenz	0.348	(0.080)	78.914	(3.549)
Acyclovir, ddI	-0.348	(0.043)	-4.159	(2.479)	AZT, 3TC, Efavirenz	0.342	(0.079)	43.526	(3.073)
AZT, ddC	-0.439	(0.029)	-5.155	(1.309)	AZT, 3TC, Abacavir	-0.442	(0.078)	54.824	(3.175)
AZT, ddI	-0.571	(0.061)	-16.615	(2.488)	d4T, 3TC, Efavirenz	-0.346	(0.069)	47.978	(3.876)
ddI	-0.375	(0.071)	15.263	(2.587)	Nevirapine, 3TC, Abacavir	-0.470	(0.099)	17.866	(12.148)
AZT, ddC, Acyclovir, ddI	-0.789	(0.115)	-13.351	(7.73)	d4T, 3TC, Kaletra	-0.310	(0.123)	35.611	(5.199)
AZT, ddC, Acyclovir	-0.514	(0.086)	-13.186	(2.168)	3TC, Kaletra, Abacavir	-0.934	(0.124)	51.570	(5.325)
AZT, ddC, ddI	-1.440	(0.047)	-32.700	(1.801)	AZT, 3TC, Kaletra	-0.655	(0.140)	49.838	(3.967)
ddC, Acyclovir	-0.310	(0.093)	2.415	(4.370)	AZT, 3TC, Kaletra, Abacavir	0.298	(0.234)	9.855	(9.404)
ddC	-0.358	(0.084)	-18.630	(3.389)	3TC, Abacavir, Efavirenz, Tenofovir	-0.308	(0.070)	31.845	(3.848)
d4T	-0.717	(0.054)	39.776	(2.210)	AZT, 3TC, Abacavir, Tenofovir	-0.652	(0.074)	19.273	(5.651)
AZT, Acyclovir, 3TC	-0.527	(0.096)	42.267	(3.394)	AZT, 3TC, Kaletra, Tenofovir	-0.552	(0.067)	32.227	(2.681)
AZT, 3TC	0.064	(0.051)	34.398	(1.875)	Nevirapine, 3TC, Tenofovir	-0.258	(0.163)	27.246	(4.619)
Acyclovir, d4T, 3TC	-0.509	(0.100)	33.792	(4.664)	3TC, Kaletra, Tenofovir	-0.092	(0.082)	51.672	(2.709)
AZT, 3TC, Saquinavir	-0.271	(0.052)	38.283	(1.992)	Kaletra, Efavirenz, Tenofovir	-0.966	(0.100)	47.617	(2.684)
d4T, 3TC	-0.104	(0.112)	37.173	(4.070)	3TC, Efavirenz, Tenofovir	-0.011	(0.108)	47.790	(5.468)
AZT, 3TC, Saquinavir, Ritonavir	-0.591	(0.085)	57.776	(10.571)	AZT, 3TC, Kaletra, Abacavir, Tenofovir	-0.738	(0.141)	19.980	(4.226)
AZT, Acyclovir, 3TC, Indinavir	-0.479	(0.056)	63.734	(2.201)	ddI , Kaletra, Tenofovir	-0.276	(0.112)	18.396	(4.015)
Acyclovir, d4T, 3TC, Indinavir	-0.295	(0.108)	78.559	(3.665)	ddI , Efavirenz, Tenofovir	-0.420	(0.117)	2.381	(2.505)
AZT, 3TC, Ritonavir, Indinavir	-0.567	(0.102)	35.032	(6.629)	Abacavir, Efavirenz, Tenofovir	-0.762	(0.140)	39.457	(3.150)
d4T, 3TC, Ritonavir, Indinavir	-0.767	(0.049)	33.510	(3.321)	Kaletra, Abacavir, Tenofovir	-0.820	(0.198)	14.891	(2.601)
d4T, 3TC, Saquinavir, Ritonavir	-0.444	(0.085)	42.631	(5.409)	3TC, Ritonavir, Abacavir, Atazanavir	-0.061	(0.039)	26.850	(1.181)
ddI , d4T, Indinavir	-0.048	(0.137)	32.286	(3.981)	Efavirenz, Tenofovir, Emtricitabine	0.118	(0.082)	54.798	(2.464)
d4T, 3TC, Indinavir	-0.395	(0.096)	53.128	(4.546)	Ritonavir, Efavirenz, Tenofovir, Emtricitabine, Atazanavir	0.306	(0.053)	83.823	(1.706)
AZT, 3TC, Indinavir	-0.075	(0.066)	65.041	(2.809)	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir	-0.403	(0.163)	38.313	(10.521)
d4T, Nevirapine, 3TC	-0.386	(0.052)	46.846	(2.962)	ddI , Ritonavir, Tenofovir, Atazanavir	0.049	(0.108)	47.800	(2.837)
AZT, Nevirapine, 3TC	0.109	(0.087)	46.275	(4.061)	Ritonavir, Tenofovir, Emtricitabine, Atazanavir	0.138	(0.104)	53.028	(3.940)
AZT, 3TC, Nelfinavir	-0.432	(0.072)	50.776	(3.924)	Nevirapine, Tenofovir, Emtricitabine	-0.205	(0.079)	37.227	(2.303)
ddI , d4T, Nelfinavir	-1.049	(0.060)	57.227	(3.672)	Kaletra, Tenofovir, Emtricitabine	-0.183	(0.093)	46.723	(5.990)
d4T, 3TC, Nelfinavir	-0.881	(0.134)	48.018	(9.588)	Ritonavir, Tenofovir, Emtricitabine, Lexiva	-0.372	(0.116)	30.226	(3.328)
Fringe Mixes									
Isoprinosine, Ribavirin, Interferons (α and/or β)	-1.017	(0.110)	-21.950	(6.644)	Nevirapine, 3TC, Ritonavir, Kaletra, Tenofovir	-1.265	(0.113)	45.683	(4.934)
Interferons (α and/or β), 3TC, Saquinavir, Indinavir, Efavirenz	-0.054	(0.243)	65.353	(5.179)	3TC, Ritonavir, Kaletra, Abacavir, Tenofovir, Atazanavir	-0.465	(0.077)	28.440	(2.687)
Nevirapine, 3TC, Saquinavir, Ritonavir, Indinavir	0.068	(0.134)	6.457	(7.335)	Ritonavir, Tenofovir, Emtricitabine, Atazanavir, Lexiva	-0.612	(0.142)	42.050	(3.579)
Nevirapine, 3TC, Saquinavir, Ritonavir, Nelfinavir	-0.689	(0.156)	30.293	(7.841)	Saquinavir, Ritonavir, Tenofovir, Emtricitabine, Atazanavir	-0.665	(0.120)	31.824	(3.879)
Nevirapine, Saquinavir, Ritonavir, Abacavir, Efavirenz	-1.121	(0.161)	19.278	(4.112)	3TC, Ritonavir, Abacavir, Tenofovir, Atazanavir, Lexiva	-0.210	(0.078)	26.678	(5.890)
Nevirapine, Ritonavir, Nelfinavir, Abacavir, Efavirenz	-0.697	(0.099)	31.044	(4.027)	Saquinavir, Ritonavir, Abacavir, Tenofovir, Emtricitabine	0.072	(0.142)	32.865	(4.856)
Nevirapine, Ritonavir, Kaletra, Abacavir, Efavirenz	-0.410	(0.174)	43.495	(5.757)	3TC, Ritonavir, Tenofovir, Emtricitabine, Raltegravir	0.032	(0.094)	33.352	(2.728)
Nevirapine, 3TC, Nelfinavir, Abacavir, Tenofovir	-0.467	(0.109)	27.893	(3.250)	Ritonavir, Tenofovir, Emtricitabine, Darunavir, Raltegravir	-0.221	(0.067)	47.736	(2.929)

Notes: Product characteristics are estimated as indicators for treatment usage in (S1) and (S2). In parentheses, standard errors computed using subsampling with 100 subsamples. For *Fringe Mixes* we only include the 5 or 6 most used products in the mix.

APPENDIX TABLE S4: Health Effects on Future Health and Ailments

	Ailments, γ^x		Health, γ^h	
Variables	coef.	se	coef.	se
h_t	0.008	(0.0004)	1.152	(0.013)
$h_t^2/10^3$	-0.013	(0.001)	-0.519	(0.043)
$h_t^3/10^7$	0.109	(0.017)	4.375	(0.546)
$h_t^4/10^{10}$	-0.040	(0.010)	-2.016	(0.298)
$h_t^5/10^{14}$	0.054	(0.021)	2.803	(0.546)
Constant	-0.929	(0.038)	-5.874	(1.350)

Notes: Parameters estimated using (S1) and (S2). In parentheses, standard errors computed using subsampling with 100 subsamples.

APPENDIX TABLE S5: Gross Income, y_{3t}

variable	coef. (γ^m)	se
h_t	0.018	(0.001)
$h_t^2/10^3$	-0.064	(0.007)
$h_t^3/10^7$	1.138	(0.171)
$h_t^4/10^{10}$	-1.030	(0.213)
$h_t^5/10^{14}$	4.854	(1.414)
$h_t^6/10^{18}$	-11.270	(4.712)
$h_t^7/10^{20}$	0.101	(0.062)
age_t	0.482	(0.034)
age_t^2	-0.006	(0.0004)
black	-5.534	(0.115)
hispanic	-4.167	(0.222)
some college	2.497	(0.141)
college	5.812	(0.157)
more than college	8.203	(0.151)
labor participation$_t$, y_{2t}	5.738	(0.074)
lack of ailments$_t$, y_{1t}	0.207	(0.024)
constant	-2.095	(0.801)

Notes: Estimates of (23). Random effects regression of gross-income on covariates. y_{3t} is measured in thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter. In parentheses, standard errors computed using subsampling with 100 subsamples.

APPENDIX TABLE S6: Out-of-pocket Expenditures, y_{4t}

variable	coef. (γ^o)	se
h_t	-0.002	(0.0004)
$h_t^2/10^3$	0.009	(0.002)
$h_t^3/10^7$	-0.133	(0.032)
$h_t^4/10^{10}$	0.090	(0.029)
$h_t^5/10^{14}$	-0.266	(0.118)
$h_t^6/10^{18}$	0.279	(0.181)
age_t	0.037	(0.004)
age_t^2	-0.0002	(0.0001)
black	-0.240	(0.014)
hispanic	-0.119	(0.016)
some college	0.169	(0.016)
college	0.318	(0.018)
more than college	0.336	(0.018)
market product$_t$	0.429	(0.016)
trial product$_t$	0.313	(0.021)
labor participation$_t$, y_{2t}	0.105	(0.009)
lack of ailments$_t$, y_{1t}	-0.122	(0.008)
constant	-1.459	(0.099)
σ^o	0.862	(0.027)

Notes: Estimates of (24) using a Tobit Model for data censored at 0. *market product*$_t = d_{J+2,t} + \sum_{k=1}^{J} d_{kt}$. Out-of-pocket expenditures y_{4t} are measured in thousands of real dollars of 2000. Health is given by the CD4 count measured in hundreds of cells per microliter. In parentheses, standard errors computed using subsampling with 100 subsamples.

APPENDIX TABLE S7: Labor Supply, y_{2t}

variable	coef. (γ^l)	se
h_t	0.009	(0.0003)
$h_t^2/10^3$	-0.013	(0.001)
$h_t^3/10^7$	0.075	(0.005)
$h_t^4/10^{10}$	-0.013	(0.002)
age_t	0.102	(0.009)
age_t^2	-0.001	(0.0001)
black	-0.168	(0.025)
hispanic	-0.040	(0.044)
some college	0.312	(0.031)
college	0.537	(0.029)
more than college	0.613	(0.033)
labor participation$_{t-1}$, y_{2t-1}	4.458	(0.028)
constant	-5.914	(0.190)

Notes: Estimates of the Logit model in (22). Health is given by the CD4 count measured in hundreds of cells per microliter. In parentheses, standard errors computed using subsampling with 100 subsamples.

APPENDIX TABLE S8: Death, $1 - b_t$

variable	coef. (γ^d)	se
h_t	-0.028	(0.001)
$h_t^2/10^3$	0.079	(0.005)
$h_t^3/10^7$	-1.104	(0.102)
$h_t^4/10^{10}$	0.704	(0.088)
$h_t^5/10^{14}$	-1.610	(0.285)
age_t	-0.116	(0.021)
age_t^2	0.002	(0.0002)
$black$	-0.509	(0.069)
$hispanic$	0.034	(0.076)
$some\ college$	0.060	(0.057)
$college$	-0.353	(0.053)
$more\ than\ college$	-0.512	(0.060)
$lack\ of\ ailments_{t-1},\ y_{1t-1}$	-1.140	(0.050)
$constant$	1.682	(0.474)

Notes: Estimates of the Logit model in (25). Health is given by the CD4 count measured in hundreds of cells per microliter. In parentheses, standard errors computed using subsampling with 100 subsamples.

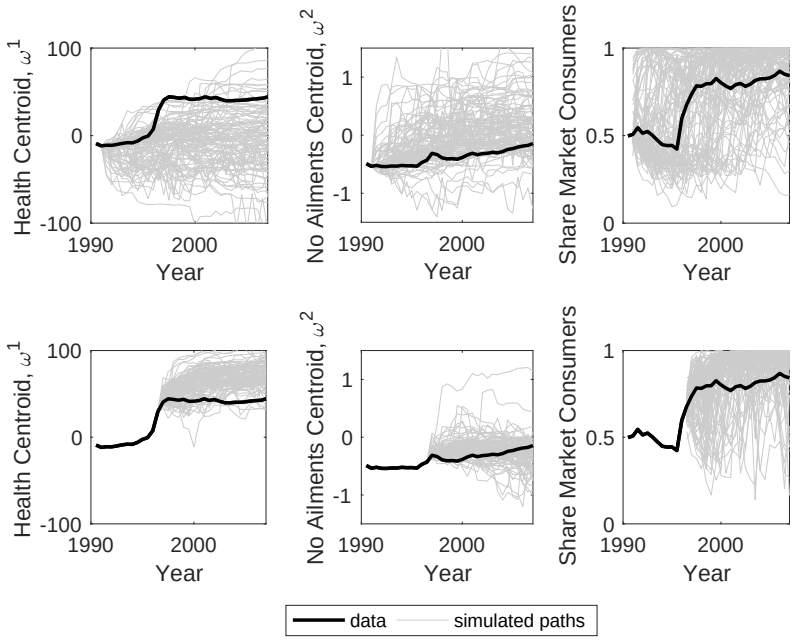
APPENDIX TABLE S9: Utility Parameters, y_{it}

coef.	variable	est.	se	unc. se
α_m	$NetIncome_t\ (y_{3t} - y_{4t})$	0.057	(0.057)	(0.010)
α_{xp}	$NoAilments_t \cdot NoProduct_t\ (y_{1t} d_{0t})$	1.019	(1.767)	(0.260)

		Cluster $j = 1,\ldots,J$			*Trial* $j = J+1$			*Repeat* $j = J+2$		
coef.	variable	est.	se	unc. se	est.	se	unc. se	est.	se	unc. se
α_{jw}	$White$	-3.546	(0.744)	(0.179)	-1.468	(0.280)	(0.136)	0.502	(0.567)	(0.130)
α_{jb}	$Black$	-4.190	(0.762)	(0.190)	-2.553	(0.334)	(0.142)	0.276	(0.613)	(0.145)
α_{jl}	$Hispanic$	-3.967	(0.958)	(0.647)	-1.585	(0.356)	(0.300)	0.707	(0.454)	(0.354)
α_{ja}	Age_t	0.043	(0.011)	(0.004)	0.032	(0.005)	(0.003)	0.009	(0.007)	(0.002)
α_{jh}	$h_t/10^3$	-2.021	(0.423)	(0.104)	-2.461	(0.203)	(0.078)			

Notes: Estimates of (26). Discount factor $\beta = .95$. $J = 3$. $NoProduct_{it}$ indicates whether he did not consume a product. h_t is defined as the number of white blood cells per cubic millimeter of blood. In parentheses, uncorrected standard errors (unc. se) computed using (S22), and corrected standard errors (se) computed using subsampling with 100 subsamples.

C.2 The Likelihood of Technological Progress



APPENDIX FIGURE S3: Distribution of Technology Paths: Technology and Consumption
Notes: 100 simulated paths conditional on the state of the world at 1991 and 1996.

C.3 Alternative Regimes

Exogenous scientific intervention. In these regimes we separate the process of innovation from demand. At the entry margin we transform the centroid to be simply the average of the characteristics of products currently available on the market, as opposed to the share weighted average in the baseline model (see (1)), and take the estimates from the supply process in Section 5.1 as given. Since $g_N(N_t \,|\, \kappa_{t-1}, s_{et-1})$ depends on s_{et-1}, we use the trial participation path resulting from averaging the simulated trial share paths from the baseline model. By following this approach we keep that part of the comparison constant relative to the baseline. At the exit margin we exogenously drop products from the market based on their quality. For this we follow one of two procedures as explained below:

↪ *Frontier*. Any product that is not in the technological frontier is dropped from the market.

↪ *Inverse frontier*. We use the exit rate path resulting from averaging the simulated exit rate paths from the baseline model. This exit rate determines the number of products n_t to be dropped. We define the inverse qualities of product k as $-\theta_k$ and the inverse frontier as the technological frontier constructed using the inverse qualities. Then we drop n_t products at

random from the inverse frontier. If n_t is larger than the amount of products in the inverse frontier, we construct the new inverse frontier and repeat the process until n_t products are dropped from the market.

Eliminating the effect of repeat purchase. In this regime we study the evolution of product quality when the process of innovation responds to demand but demand by repeat consumers is random. This regime neutralizes the dependence of the technological path on the preferences and characteristics of repeat consumers without changing the nature of the process on the supply side. We avoid spurious effects on the process of innovation, yielding from arbitrary aggregate shares (e.g. $1/G$ for a choice set of size G), by setting the unconditional shares of this alternative regime to match the unconditional shares in the baseline.

Mandated treatment. The first planner can only assign alternatives based on whether a person's health is high or low and whether the person decided to consume a market treatment last period (either by repeating his previous market product or by choosing a cluster). Hence, the planner's policy rules can be based only on four different categories. The planner can send all individuals in each of the four groups to any of the $J+2+r_{it}$ alternatives available. We nest the baseline competitive equilibrium allocation by adding one alternative to the planner's action set: the competitive equilibrium allocation. Hence, there are $J+3+r_{it}$ alternatives in the planners action set and he can base his assignment on 4 categories. Since only two of the four categories can repeat their previous market treatment (when $r_{it}=1$), this amounts to $7^2 \times 6^2 = 1,764$ policy rules. An example of a policy rule is presented in Table S10. We precompute a set of continuation values and match them to allocation rules to avoid forward simulation for each rule. We further explain these procedures below.

APPENDIX TABLE S10: Example of an Action-Constrained Planner's Policy Rule

Category		Alternatives						
Health status	Product $t-1$	Cluster 1	Cluster 2	Cluster 3	Trial	Repeat	No product	CE
high	yes							x
high	no		x					
low	yes				x			
low	no						x	

Notes: *Product* $t-1$ column indicates whether individuals in this category consumed a market product in $t-1$. *CE* column indicates that the planner assigns the competitive equilibrium allocation.

Optimal experimentation with trial products. The second planner we consider can base his policy on the entirety of the individual state but his action set has only two elements: he can give the person the trial product or he can allocate the competitive equilibrium allocation (excluding

the trial product). Policy rules for this planner are levels of trial participation and his problem also nests the competitive equilibrium allocation. For policy rules below the competitive trial share s_{et} the planner incurs a welfare costs by preventing people from rationally joining a trial. For policy rules above s_{et} he incurs a welfare costs by forcing people to join a trial who rationally avoided it. Welfare gains, if any, come from the externality via experimentation in clinical trials, which pushes innovation. We evaluate policies in increments of 0.5 percent points, which amounts to 202 policy rules. Here we also use the set of precomputed continuation values and match them to allocation rules to avoid forward simulation for each rule.

C.3.1 Continuation Values and Smoothing

We obtain continuation values for every planner rule by implement the following algorithm:

1. Create a collection, denoted $\mathscr{A}$, of 500 continuation value vectors computed for as many $t+1$ states—each row in a value vector is an individual. Each value vector $v \in \mathscr{A}$ corresponds to a $t+1$ aggregate state z^v_{t+1}.

2. For each rule n in a given planner problem, we compute each individual's current payoff and their future state, as well as the implied $t+1$ aggregate state z^n_{t+1}.

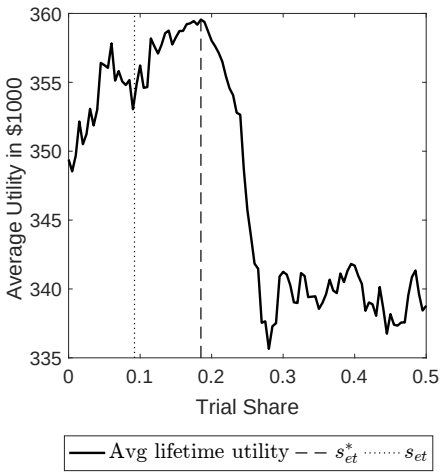
3. We match rule n to the continuation value vector $v^* \in \mathscr{A}$ corresponding to the $t+1$ aggregate state that is closest to the aggregate state induced by rule n. In other words, we match rule n to the continuation value vector v^* that solves:

$$v^* = \arg\min_{v \in \mathscr{A}} ||z^n_{t+1} - z^v_{t+1}|| \tag{S24}$$

   We use a measure of Euclidean distance that yields from discretizing the aggregate states z^n_{t+1} and z^v_{t+1} into vectors with 196 components. We scale each component of the discretized aggregate state vectors to be between zero and one by dividing over its largest value.

4. We repeat steps 2 and 3 one thousand times for every rule n and average over repetitions.

As Figure S4 shows, our method of matching continuation values generates noise around the mapping from planner rules into average consumer lifetime utility for the planner who chooses the optimal trial share s^*_t. Hence, we use a local polynomial to smooth the mapping in an interval starting at the competitive equilibrium share s_{et} and going 15 percent points above it (from 0.09 to 0.24). This produces Figure 12 and the results associated with it in Table 6.



APPENDIX FIGURE S4: Optimal Assignment to Clinical Trials

Notes: The solid line represents average lifetime utility. The dashed line indicates the planner's optimal share s^*_{et}. The dotted line represents the competitive equilibrium share s_{et}. Year is 1996.