Why Medical Innovation is Valuable: Health, Human Capital and the Labor Market*

Nicholas W. Papageorge†

February 5, 2014

Abstract: I develop a framework to assess the value of pharmaceutical innovation, taking explicit account of how side effects and the labor market affect the demand for medical treatment. In the framework, forward-looking patients do not simply maximize underlying health or longevity. Rather, in light of painful or uncomfortable side effects, they choose labor supply and medicine in an effort to jointly manage two forms of human capital: their health and their labor market experience. The framework is used to examine the treatment and employment decisions of HIV+ men before and after a medical breakthrough known as HAART. Main findings include (1) a counterfactual medical innovation that reduces side effects of existing drugs—despite no improvement to drug effectiveness—is potentially very valuable and part of its value arises since working while suffering side effects can be difficult. (2) Forward-looking, chronically ill patients optimally choose to cycle among available treatment options, favoring effective treatments despite side effects when in poor health, but switching to less effective drugs with fewer side effects (or avoiding treatment altogether) when their health improves. Since working while suffering side effects can be difficult, part of the incentive to cycle off of treatment comes from the desire to increase consumption through employment.

Key Words: Innovation, Health, Human Capital, Labor, Structural Models, HIV/AIDS.
JEL Classification: I10 J24 O31

*This paper was originally a chapter in my doctoral thesis completed at Washington University in St. Louis. Special thanks goes to my dissertation committee: Barton Hamilton, Tat Chan, Mariagiovanna Baccara, Sebastián Galiani, Juan Pantano and Robert Pollak. For helpful comments and conversations, I also thank Jorge Balat, Melanie Blackwell, Janet Currie, Hülya Ereslan, Amy Finkelstein, Stephanie Heger, Glenn MacDonald, Robert Moffitt, Harry Paarsch, Carl Sanders and Seth Sanders. All errors are my own.

†Department of Economics, Johns Hopkins University. Email: papageorge@jhu.edu.

1 Introduction

Beginning with the work of Grossman (1972), economists have envisioned health as a form of human capital that affects productivity as well as longevity and well-being. This framework has been a dominant one in the literature for assessing the value of improvements in medical technology, including innovations in drug effectiveness. However, the framework leaves out two critical factors: first, it leaves out the possibility that new drugs have more serious side effects than older drugs and, second, it does not have an explicit compliance decision on whether to take a new drug. When these two factors are added to the model, the individual can be seen as facing a tradeoff between enhancing health and suffering side effects that potentially reduce time in the labor market. An implication is that individuals make decisions about medical treatment and labor supply in an effort to jointly manage two forms of human capital: their health and their work experience. Evaluation of medical innovation is therefore incomplete if the interaction between health and the labor market is not considered. In particular, a medical innovation that lengthens life, but also has side effects that cause pain and discomfort—or make it difficult to work—may be less valuable than a treatment that does not affect longevity, but instead improves the quality of life.

In this paper, I develop a general framework to assess the value of medical innovation, taking explicit account of how side effects and the labor market affect demand for medical treatment. Though prior work has recognized various links between health and labor, the main contribution of this paper is to incorporate how these links influence patient treatment and employment decisions. In the framework, patients are not viewed as maximizing their underlying health or longevity. Rather, in light of potential side effects, patients actively manage their health capital in a way that balances the impacts of medication on the labor market and productivity with impacts on mortality and morbidity. The framework is therefore consistent with research emphasizing how individuals value healthcare because it makes them live not only longer, but also better lives (Hall and Jones, 2007). It also marks a departure from earlier work studying the value of medical innovation, which typically focuses on increases to life expectancy (Murphy and Topel, 2003, 2006) or relies on stated or elicited (as opposed to revealed) preferences to assess how medicine affects the quality of life (Lipscomb et al., 2009).

The framework centers around estimation of a dynamic model where forward-looking agents simultaneously choose medical treatment and labor supply. The model is set up in a way that allows it to capture the following two key tradeoffs. First, and consistent with earlier work linking health and labor, agents treat their health as a form of capital stock (Becker, 2007; Heckman and Cunha, 2007; Currie, 2009; Conti, Heckman, and Urzua,

2010). They choose effective treatments to invest in their health capital, but may also forego treatment to avoid painful or uncomfortable side effects, thus allowing their health capital to depreciate. The second tradeoff is the one faced by agents choosing whether or not to work. By working, agents earn income and accumulate human capital, but also forego valuable leisure time. Moreover, the structure of the model also allows it to capture various ways in which these two tradeoffs interact. Poor health can negatively affect productivity, earnings and labor supply, which encourages investments in health capital (Currie and Madrian, 1999; Cawley, 2004; Garthwaite, 2012). Further, side effects can discourage employment by raising the utility cost of work, leading some patients to avoid medicine. Finally, employment gaps, including those induced by illness or side effects, can slow the accumulation of labor market experience, reducing future income (Mincer and Polachek, 1974; Becker, 1985; Eckstein and Wolpin, 1989).

Before discussing the application, I highlight two specific features of the model and describe the benefits of each in allowing the model to capture the tradeoffs mentioned above. First, patients using medication are viewed as consuming bundles of characteristics. Second, each treatment is measured along two dimensions of quality: (i) effectiveness at improving underlying health, which governs longevity and symptoms, and (ii) propensity to cause immediate side effects. Symptoms and side effects manifest as physical *ailments* and it is these ailments that affect patient utility.[1] To summarize how these features of the model work: agents are not viewed as having preferences over specific medications or over their underlying health *per se*. Rather, they have preferences over what their underlying health delivers: a longer life and a reduction in symptoms. To improve their health, agents can take drugs, but these health investments come at the potential cost of painful or uncomfortable side effects.

A key benefit of using the 'characteristics approach' to estimate demand is that it permits assessment of potential new drugs introduced to the market, each constructed as unique, counterfactual effectiveness and side effects bundle (Petrin, 2004).[2] Another benefit is that it permits straightforward interaction of preferences over health and longevity with preferences over goods that influence the quality of life, like consumption and leisure. Exploiting the characteristic approach allows me to show, for example, that medical innovations aimed at reducing side effects of existing drugs—despite no improvement to drug effectiveness—are potentially very valuable and that part of this value arises since working while suffering side effects can be difficult and employment gaps are costly.[3]

[1] In the model, out-of-pocket costs are also included. Also, note that it would be straightforward to extend the model to incorporate further drug characteristics, like convenience.

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

[3] A common alternative to the characteristics approach is to allow patients to have preferences over specific

The major benefit to measuring drugs along multiple dimensions of quality, i.e., as having more than one characteristic, is that it highlights how, in cost-benefit analyses of competing drugs, it is not always appropriate to view one drug as strictly better or worse than another. Instead, two or more competing treatments taken at different points in time and depending on time-varying patient characteristics (including their current health status and their accumulated work experience) could be better than either treatment alone. To capture various ways that multiple drug characteristics can affect demand, the framework goes beyond studies explaining patient-level treatment demand solely through the utility generated by underlying health or by health and side effects (Crawford and Shum, 2005; Chan and Hamilton, 2006; Fernandez, 2008; Chintagunta, Jiang, and Jin, 2009). Instead, I allow time-varying patient characteristics (including current-period health, employment, age and accumulated work experience) to affect choices. I also permit unobserved heterogeneity in how drugs affect individuals (through both effectiveness and side effects) and in distaste for physical ailments, work and the interaction between the two. The result is a rich model of demand, which rationalizes observed variation in treatment choices—not only across individuals, but also for the same individuals across time. By rationalizing strong variation in demand among similarly healthy individuals, the framework therefore offers a compelling explanation for why multiple drugs of similar average effectiveness (often known as 'me-too') drugs can co-exist within a single market.

In the case of chronic illness, a patient's efforts to jointly manage health and labor market human capital become permanent fixtures in dynamic decision-making. I apply the framework developed in this paper to study the treatment and employment decisions of men suffering from a potentially severe chronic condition: infection with HIV. Focusing attention on men with HIV does not mean that findings are difficult to generalize. HIV, like many other chronic conditions (e.g: diabetes, multiple sclerosis and depression) is harmful or deadly when untreated, but can be quite manageable when treated, though at the possible cost of mild-to-severe side effects.[4] Further, according to the Centers for Disease Control and Prevention, nearly 50% of adults in the U.S. suffer from a chronic condition, about one quarter of whom experience significant limitations in daily activities like working.[5]

drugs or molecules (see, e.g., Arcidiacono et al. (2012)). Doing so can capture consumption and substitution patterns, some effects of market structure or the removal of drugs from the consumer choice set. Nonetheless, use of drug or molecule dummy variables in the utility function effectively precludes analysis of counterfactual drugs and makes it difficult to interact treatment demand with preferences over other goods.

[4]Individuals suffering from multiple sclerosis, for example, can live longer if they take one from a class of drugs containing interferons. The cost, in terms of side effects, is that patients feel like they have the flu, experiencing fatigue, fever, soreness and chills. In response, some patients choose to forego medication for limited periods of time, though this can accelerate disease progression (Kerbrat et al., 2011).

[5]For this point, see: http://www.cdc.gov/chronicdisease/resources/publications/aag/chronic.htm. In principle, the structure of the model means it could be applied to illnesses that are not chronic, but where

Several features of HIV and the AIDS epidemic make it a natural setting for examining how agents choose medication to jointly manage their health and labor market human capital. Perhaps most important, identifying this tradeoff requires strong variation in both health status and drug characteristics, including side effects. Untreated, HIV infection leads to immune system deterioration (known as AIDS) where routine infections lead to grave symptoms and death. Absent treatment, an individual newly infected with HIV lives an average of 11 years. Additionally, phases of the AIDS epidemic are distinguished by wide variability in the characteristics of available treatments. The key to identifying the model parameters is that I observe treatment and employment choices for the same individuals both before and after a medical breakthrough known as HAART.[6] A treatment introduced in 1996, HAART is credited with having transformed HIV infection from a virtual death sentence into a chronic, manageable condition, though at the cost of harsh side effects.[7]

Turning to results, I find that from the perspective of an HIV+ patient, a dynamically optimal treatment plan is not consistent with full compliance nor with strict longevity maximization. This finding stands in stark opposition to prevailing medical literature emphasizing strict adherence to the most effective medication available, despite costs like side effects (El-Sadr et al., 2006). Observed treatment choices confirm that sicker HIV+ individuals opt for effective treatments like HAART. Once in better health, however, they are less likely to choose HAART, a pattern the model rationalizes as part of a dynamically optimal plan of treatment cycling. When in poor health, agents facing low survival rates anticipate high marginal returns to investments in their health 'stock'. They respond by opting for effective treatments. Once their health improves, however, agents exploit persistence in underlying health, switching to less effective drugs to avoid side effects, allowing their health capital to depreciate. However, they maintain the option value of switching back to effective treatments once their health deteriorates. This phenomenon is henceforth referred to as *optimal treatment cycling*.[8]

effective treatment can influence labor supply. To take an extreme example, a good treatment for the flu is bed rest. People with the flu therefore face a tradeoff between working and getting better more quickly. More generally: agents do not need to be ill at all to face a tradeoff between investing in their health versus their labor market human capital. For example, many individuals face a daily choice between going to the gym or working longer hours.

[6]HAART stands for highly active anti-retroviral treatment. There is no vaccine or cure for HIV or AIDS, but HAART, introduced in 1996, is the current standard treatment.

[7] Duggan and Evans (2008) also use HAART introduction to study the affects of a medical breakthrough, though their focus is on health rather than on the influence of labor or side effects on treatment demand. It should also be noted that the impact of HAART has not been limited to HIV+ patients. First, it increased the continuation value associated with HIV-infection. Second, it lowered the infectiousness of HIV+ men. Both of these lowered the implicit price of risky sexual behavior. These effects are explored in Philipson and Posner (1993); Lakdawalla, Sood, and Goldman (2006) and Chan, Hamilton, and Papageorge (2013).

[8]Even if treatment cycling deteriorates health, it is not incongruent with an optimal treatment plan since

The model also reveals how employment decisions and the labor market interact with health. Physical ailments—either symptoms or side effects—exacerbate the utility cost of work. Accordingly, full-time employment exhibits cycles that mimic optimal treatment cycling because relatively healthy agents cycling onto to milder treatments (or avoiding treatment altogether) experience fewer side effects and return to work. In other words, while allowing their health capital to depreciate, agents invest in their labor market capital by accumulating work experience. Moreover, although HAART has side effects, it improves average health, thus reducing symptoms, so that the net effect can be an increase in employment. Accordingly, I find that, had HAART not been introduced, employment would have been up to 7.5% lower among HIV+ men in the years following its introduction in 1996.

Exploiting the characteristics approach to evaluate HIV treatment innovations, including HAART and counterfactual treatments, I find that the value of a given treatment varies widely across similarly unhealthy individuals, depending on their age and human capital along with unobserved heterogeneity in drug effectiveness, drug side effects and preferences over physical ailments. HAART, for example, is worth between $2,000 and $180,000, with higher values accruing to younger agents and those with more work experience. Moreover, I find that side effects innovations are valuable: a counterfactual version of HAART with no side effects is valued up to $160,000 over HAART.

The remainder of this paper proceeds as follows: Section 2 introduces the data and provides some background on HIV and the AIDS epidemic; Section 3 presents the model; Section 4 describes estimation; Section 5 studies the value of pharmaceutical innovation; Section 6 discusses policy experiments highlighting how drug innovation interacts with employment; and Section 7 concludes.

2 Data and Background

I use the public data set from the Multi-Center AIDS Cohort Study (MACS), an ongoing study (beginning in 1984) of the natural and treated histories of HIV− (i.e., not infected with HIV) and HIV+ homosexual and bisexual men conducted at four sites: Baltimore, Chicago, Pittsburgh and Los Angeles.[9] At each biannual visit, data are collected on: medical

it reflects how patients trade off health and other components of utility. Nonetheless, some studies cast doubt on the near-consensus in the medical literature that intermittent treatment is bad for health, which underscores the dynamic optimality of treatment cycling (Stebbing and Dalgleish, 2009).

[9]Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) 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,

treatment choices, employment decisions, labor market outcomes and health status, which includes CD4 count (a measure of immune system health) and subject reports of physical ailments, like nausea, fever and drenching sweats. As the data set is a panel, I observe behavior before and after HAART introduction, which occurred between 1995 and 1996. This permits analysis of how a medical breakthrough can affect both health and employment. I use data on HIV+ men beginning in 1990, at which point drugs with some effectiveness at combating HIV emerge and treatment data collection becomes more consistent across time. The MACS data set contains information on 769 HIV+ individuals, corresponding to 9,837 subject-visits, between 1990 and 2003 (at which point the sample period ends). Observations with missing data are dropped and the resulting analysis sample is an unbalanced panel of 8,300 observations: 743 subjects over 26 visits.[10]

Agents report all medications they have used since their previous interview. As there are dozens of medications used to fight HIV infection, I follow previous research (see, for example, Detels et al. (2001)) in creating four broad and mutually exclusive treatment categories: no treatment, mono-therapy, combo-therapy and HAART.[11] To measure accumulated human capital, I use potential experience (current age minus 25) up until the start of the AIDS epidemic (1984) and thereafter construct employment histories using observed labor supply choices.[12] I model employment choices to be dichotomous—full time or not full time—since more detailed information is available for only a subset of sample periods.

2.1 Summary Statistics

The interactions among health, medical treatment choices, side effects and employment are complex, though important dynamics emerge from summary statistics. These are presented in Table 1 for the full analysis sample and then separately for the periods before and after HAART, by health status (high or low CD4 count) and by employment status (full time

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.

[10]The full MACS data set, including pre-1990 observations and information on uninfected individuals, contains information on 5,622 subjects at 41 possible visits for a total of 98,886 subject-visits.

[11]An agent with the label "none" may take medications to fight opportunistic infections, such as pneumonia. Mono-therapy denotes a regimen consisting of a single nucleoside reverse transcriptase inhibitor (NRTI). Combo-therapy consists of several NRTIs. HAART has a more complex definition that includes several drug regimens, most of which include a protease inhibitor in combination with an NRTI or a non-nucleoside reverse transcriptase inhibitor (NNRTI).

[12]Employment histories are constructed using all available data, including observations when HIV+ agents were observed HIV−, if applicable, and observations with up to two missed subsequent visits, in which case I assume that agents engage in the same employment status as in the last observed period.

or not). In the full sample, subjects are on average about 33-years-old at the start of the AIDS epidemic (1984). Uninfected individuals generally exhibit a CD4 count (a measure of immune system functionality) of 500-1500 units per mm^3 of blood. The sample average is slightly lower—about 450—but this number obfuscates important variation: the pre- and post-HAART averages are 393 and 567, respectively. The most salient feature to capture is whether CD4 is low enough to signal loss of immune system functionality. For subsequent analysis, I therefore construct a binary variable that takes the value 0 when patient CD4 count is low enough to indicate AIDS (< 250). About one quarter of observations exhibit AIDS-level CD4. Subjects also report a number of physical ailments, which may reflect symptoms of AIDS, side effects of medications or both. I construct a second indicator variable for ailments, which takes the value 0 if agents report persistently experiencing one of the following ailments: fatigue, diarrhea, headaches, fever or drenching sweats. About 60% of subjects report that they are free of such ailments (i.e., $F = 1$). Finally, death probability is about 4% over the entire sample period.

Considering the pre-HAART and post-HAART eras separately reveals important differences (Columns 2 and 3 of Table 1). Foremost are health improvements (measured by both CD4 count, AIDS level CD4 and survival).[13] Further, better health correlates with fewer ailments (see Columns 4 and 5 of Table 1, which compare high and low CD4 count agents). Despite improved average health after HAART is introduced, the same proportion of agents reports suffering physical ailments (59%) in the pre- and post-HAART eras. This parity arises since side effects replace symptoms as agents increase treatment usage, a change that is also reflected by post-HAART increases in expenditures on treatment. In other words, HAART both mitigates and exacerbates physical ailments by effectively fighting symptom-causing illness, but simultaneously causing side effects. The model developed in the following section is designed to capture how this tension influences treatment choices.

Turning to the interactions between health and the labor market, average reported income is about \$38,000 (in 2003 dollars per year).[14] Non-wage income averages about \$26,500, which is lower than the average amount reported by workers (about \$44,000), but may seem high at first glance. It reflects that HIV/AIDS is considered a disability, which opens up the possibility of social security disability payments and private pensions, which would

[13] In the context of HIV and HAART, Goldman and Bao (2004) find that HAART use is associated with a higher likelihood of remaining employed.

[14] Income is a categorical variable reported in increments of \$10,000 where the highest income category is \$50,000 or more. To convert per-period income into dollars, I take the midpoint of each category and then divide it by 2. The highest income category is set to \$27,500 per semester, though reduced-form results are robust to higher values that account for censoring. I then use the TAXSIM version 9 .ado file developed by the National Bureau of Economic Research to calculate net income, which I then convert to 2003 dollars. Out-of-pocket treatment costs are also converted into 2003 dollars.

presumably increase with pre-disability wage income. In support of this possibility, auxiliary regressions show that HIV+ agents' non-wage income is positively correlated to their wage income in periods before they were infected with HIV. Despite high non-wage income, it is clear that agents experience a large income drop if their health or physical ailments discourage work. Indeed, there is clear evidence that both good health and freedom from physical ailments predict full-time work (see columns 6 and 7 of Table 1, which compare agents by their employment status). Further, the rate of public insurance is higher in the post-HAART era (compared to private or no insurance), which could reflect that the sample is an aging cohort and that laws governing disability change over time. Finally, public insurance is also correlated with not working and poor health. Since insurance influences out-of-pocket treatment costs, which can affect treatment choices, it will play a role in subsequent analysis.

2.2 Treatment Choices and Employment Decisions

A number of factors influence treatment choices. First, notice (again referring to summary statistics in Table 1) that a plurality of agents (45%) eschewed all medical treatments in the pre-HAART era, but a majority of agents (62%) use HAART after it is introduced. This shift is depicted in Figure 1(a), which plots treatment choices over the sample period. Not only do agents substitute HAART for other treatments, but those who refrained from using earlier, less effective treatments switched onto HAART after 1996. This dynamic suggests that agents are more willing to suffer side effects if the treatment is effective.

A puzzling feature that emerges in Figure 1(a) is that not everyone who is HIV+ uses HAART. One possibility is that after HAART is introduced, agents must learn about it before adopting the new technology. However, learning is not consistent with changes in usage over time, including an immediate and explosive increase in HAART usage after its introduction followed by a sharp leveling-off within a couple of years, after which there is a fairly constant proportion of agents who do not use HAART (about 38% after 1997). This brings up a second possibility: that a subgroup of agents simply never goes onto HAART. This hypothesis, however, is not borne out in Figure 1(b), which plots lifetime HAART usage as a proportion of the total sample over time. By the end of the sample period, nearly 90% of all agents have used HAART at least once. Nonetheless, there is no period when 90% of agents are on HAART at the same time. This discrepancy is crucial. It means that being on HAART is not an absorbing state. Rather, for these dynamics to arise, it must be the case that some agents go onto and off of HAART at different points in time. A key contribution of the framework developed in this paper is to rationalize why individuals would ever go off of HAART.

To shed light on the dynamics shown in Figures 1(a)-1(b), Table 2 presents a transition matrix for individual treatment choices. First, treatment choices are highly persistent, though more strongly so for agents in better health. Second, a small proportion of both healthy and unhealthy agents go off of HAART in any given period. About 13% of AIDS-level patients and about 6% of healthier agents go off of HAART. Once off, however, healthier agents are likely to remain off of HAART (about 87%), whereas about half of sick agents start taking medication, with about 35% going onto HAART. Further, in any given period 5% of individuals who are not on HAART develop AIDS-level CD4 counts. Given low CD4 and HAART usage, about 29% recover a higher CD4 count in each period. Taken together, these statistics suggest that many healthy agents will eventually go off of HAART and and, if so, get sick. Once sick, many go onto HAART to get well. The data therefore exhibit health-dependent cyclicality in treatment choices.

Given the post-HAART drop in death rates (see Figure 1(c)), however, it remains unclear why an HIV+ agent would ever go off of HAART given the possibility of succumbing to AIDS via CD4-count drops. It is unlikely that individuals avoid HAART due to its cost since out-of-pocket treatment costs are fairly low and exhibit low variability across treatments. Recall, however, that although HAART can ultimately decrease physical ailments by improving CD4 count, it does so at the immediate cost of inducing physical ailments via side effects. Hence, agents in good health may avoid going onto HAART in order to avoid physical ailments arising from side effects.

The data also suggest how the labor market and employment decisions interact with treatment choices, health and side effects. According to summary statistics in Table 1, poor health discourages work. Further, Figure 1(d) depicts labor supply decisions over time. HAART coincides with a break in the decreasing trend of full-time employment in the aging sample. To underline the significance of this break, I extrapolate the pre-HAART full-time employment trend until 2001.[15] This exercise suggests that a counterfactual world without HAART may have witnessed lower employment among HIV+ men. The structural model developed in the following section is designed to uncover how labor interacts with treatment innovations and health.

Table 1 also shows that agents reporting physical ailments are less likely to engage in full-time work. However, it is unclear if physical ailments have an independent effect on employment or merely capture the effect of poor health on agent choices. To explore this possibility, I present coefficient estimates from static logistic regressions where the depen-

[15] I regress pre-HAART employment decisions on age, age-squared and a linear time trend and then use these parameters to predict employment decisions in the post-HAART era, taking post-HAART age profiles as given.

dent variable is a dichotomous employment choice (Table 3). First, employment is very persistent (also apparent in in Table 4, which presents transitions into and out of the labor force). Estimates also indicate that, independently of CD4 count, physical ailments reduce labor force participation. Therefore, a health-dependent cyclicality in the labor market also emerges: healthier agents go off of HAART, which lowers side effects and thus encourages employment and raises income and consumption. However, in doing so, they face a higher probability of a drop in their CD4 count, physical ailments in the form of symptoms, and death.

2.3 HAART Introduction

In subsequent analysis, when specifying agents' future beliefs, the introduction of HAART is not anticipated by HIV+ patients.[16] Two observations justify this approach. First, HAART was not a single medication developed and improved over time such that subjects might update their beliefs and anticipate higher future efficacy. Rather, HAART introduction was abrupt and many components of HAART already existed prior to 1996. The key insight involved the union of several existing technologies, none of which was particularly effective on its own. Second, subject reports from survey questions asking about their hopefulness about the future are not consistent with anticipation of HAART. Specifically, one in a battery of questions meant to assess depression asks subjects how often in the week preceding their interview they felt hopeful about the future.[17] Figure 1(e) plots the probability that subjects answer, "all or most of the time" over time. Notice the pre-HAART flat (or even downward) trend followed by a break and reversal coinciding with HAART introduction. Importantly, if the effectiveness of HAART had been anticipated, this upward shift in hopefulness should have occurred before HAART introduction.[18]

[16] In this sense, HAART introduction is treated as a quasi-experiment, an assumption that implies the need for caution in applying the framework developed in this paper to cases where medical innovation is anticipated.

[17] Questions are from a depression screening tests known as the Center for Epidemiological Studies Depression (or CES-D) scale. See, for example, Ostrow et al. (1989), for an example of CES-D scale use with the MACS data set.

[18] One concern is that hopefulness is highly correlated with health so that the trend reversal simply reflects HAART-induced health improvements. To account for this, I control for a polynomial in CD4 count and age in a regression where the regressand is a dichotomous variable for being hopeful about the future 'most or all of the time'. Plotted residuals (Figure 1(f)) show a similar trend reversal at, but not before, HAART introduction.

3 Model

In each period, agents enjoy flow utility, which is a function of current choices and state variables. Before retirement at age 65, agents choose treatments and employment at each period. Agents are forward-looking, so their choices maximize the present discounted value of future utility. Agents retire at age 65 and cease making decisions. Period t state variables are a function of previous-period states and choices so that the dynamic programming problem can be solved using backward induction.

3.1 Choices and Flow Utility

At each period t until retirement agents choose a pair $d_{it} \equiv (d_{it}^L, d_{it}^M)$, where d_{it}^L represents the employment choice and d_{it}^M the treatment choice. In particular, the possible choices on each dimension are:

$$d_{it}^L = \begin{cases} 0 & \text{Not full time work} \\ 1 & \text{Full time work} \end{cases} \quad \text{and} \quad d_{it}^M = \begin{cases} 0 & \text{No Treatment} \\ 1 & \text{Mono-therapy} \\ 2 & \text{Combo-therapy} \\ 3 & \text{HAART (only after 1996)} \end{cases} \tag{1}$$

Note that the set of choice pairs, denoted by D_t, is time-dependent since HAART is available only after 1996. Specifically, denoting as D_t^L and D_t^M the set of labor and treatment options available at period t, respectively, $D_t \equiv D_t^L \times D_t^M$. Ailment status is given by $F_{it} \in \{0, 1\}$, where 1 signifies being free of ailments and 0 signifies suffering ailments. Flow utility is given by:

$$\begin{aligned} U(C_{it}, F_{it}, d_{it}) \;=\; & \textstyle\sum_{f=0}^{1} \mathbf{1}\{F_{it} = f\} \times \big[\\ & u\left(C_{it}, F_{it}, \gamma(F_{it})\right) + \theta_1^f + (\theta_2^f \times \mathbf{1}\{d_{it}^L = 1\}) \\ + \;& \theta_3^f \times \mathbf{1}\{d_{i,t-1}^M = 0\} \times \mathbf{1}\{d_{it}^M \neq 0\} \\ + \;& \theta_4^f \times \mathbf{1}\{d_{i,t-1}^M \neq d_{it}^M\} \times \mathbf{1}\{d_{i,t-1}^M \neq 0\} \\ + \;& \theta_5^f \times \mathbf{1}\{d_{i,t-1}^M \neq 0\} \times \mathbf{1}\{d_{it}^M = 0\} + \epsilon_{it}(d_{it}) \;\; \big] \end{aligned} \tag{2}$$

The first term on the right-hand side of equation (2) is a sum over each ailment status $F_{it} \in \{0, 1\}$ along with an indicator function. This term is multiplied with the remainder of the terms so that flow utility is health-state dependent in an indirect way. This first term on the second line of equation (2) represents individual utility over consumption (C_{it}). The marginal utility of consumption varies by ailment status and $u(\cdot)$ is a CRRA utility function

with parameter $\gamma(F_{it})$ so that

$$u\left(C_{it}, F_{it}, \gamma(F_{it})\right) = \frac{1}{1-\gamma(F_{it})} C_{it}^{1-\gamma(F_{it})}. \tag{3}$$

The second term on the second line of equation (2) represents the direct utility level effect of suffering ailments (when $f = 1$) or being free of ailments (when $f = 0$). θ_0^f is normalized to zero. The third term on the second line captures the ailment-specific utility cost of full-time work. Interacting the disutility of work with ailment status captures whether agents find it relatively more costly to be employed when suffering from symptoms or side effects.

Agents do not have preferences over CD4 count *per se*. Instead, CD4 count affects agent symptoms and longevity, but flow utility depends on day-to-day physical ailments. Period t treatment choices therefore affect intertemporal utility (through their effect on health as measured by CD4 count) and current period flow utility (through ailments induced by side effects). Both of these processes will be explained in the following section. Treatment choices also enter flow utility directly via switching costs, captured by the terms in lines 3-5 of equation (2). Finally, $\epsilon(d_{it})$ is a choice-specific utility-shifter, which captures factors that affect agent choices, but that are not observable to the econometrician. In particular, $\epsilon_{it} : D_t \to \mathbb{R}$ and I use $\epsilon(d_{it})$ to denote the utility shifter associated with choice d_{it}. Finally, $\epsilon_{it}(d_{it})$ are Extreme Value Type I distributed.[19]

Switching costs capture factors—beyond preferences over ailments and long-term health—that affect agent treatment decisions, including doctors' orders, treatment protocols and the social benefits of antiretrovirals.[20] Note that switching costs are generic, i.e., not specific to any particular treatment. Instead, agents experience a cost of starting, switching or ending treatment. Moreover, switching costs vary by ailment status. This specification of preferences amounts to a characteristics approach to modeling the demand for treatment. In other words, patients do not have preferences over a specific treatment like HAART, whereby HAART would enter the utility function as a dummy variable. This approach is crucial for evaluating counterfactual treatment innovations, each defined by the probability distribution it implies over CD4 count and ailments. The processes according to which choices and states generate ailments and consumption are described in the following section.

[19]This assumption, along with conditional independence of states and outcomes, which will be formally stated later, follows Rust (1987).

[20]Effective HIV treatments lower viral loads (the amount of virus in a patient's blood), which renders patients less infectious to HIV− sex partners.

3.2 States and Transitions

Upon entering period t, the agent learns his vector of period t state variables (denoted S_{it}), but he still faces uncertainty about ailments (F_{it}) and consumption (C_{it}), both of which are realized only after he makes his labor supply and treatment decision. Therefore, the agent evaluates expected flow utility conditional on his current choice d_{it} and his vector of period t state variables, formally:

$$\mathrm{E}[U(C_{it}, F_{it}, d_{it})|S_{it}]. \tag{4}$$

The agent's treatment and labor supply decision has a direct impact on the stochastic process generating F_{it} and C_{it}. Finally, choices and current states jointly determine period $t+1$ state variables.

State variables (S_{it}) include a vector of observables, denoted $\mathcal{X}_{it}$, and a vector of unobservable utility shifters (ϵ_{it}). Specifically, $\mathcal{X}_{it} \equiv [H_{i,t-1}, A_{i,t-1}, E_{i,t-1}, v_{t-1}]$, where:

$$\begin{aligned}
H_{i,t-1} \in \{0,1\} &: \text{High (non-AIDS) CD4 count at } t \\
A_{i,t-1} \in \{25, 25.5, 26, \ldots, 65\} &: \text{Age at } t \\
E_{i,t-1} \in \{10, 20, \ldots, 50\} &: \text{Semesters of full-time experience at } t \\
v_{t-1} \in \{1, \ldots, 15\} &: \text{Period } t \text{ dummy}
\end{aligned}$$

Recall from Section 2 that HIV infection leads to a low CD4 count, which means that the patient's immune system is compromised.[21] S_{it} also includes the unobserved, choice-specific utility shifters ($\epsilon(d_{it})$'s) defined in the previous section.

Next, the agent forms expectations on F_{it} and C_{it}, which are collected into a vector denoted $\mathcal{Y}_{it}$ so that: $\mathcal{Y}_{it} = [F_{it}, C_{it}]$.[22] I assume conditional independence of $\mathcal{Y}_{it}$, i.e., outcomes are independent of realizations of unobservable flow utility shifters. Formally:

$$\mathrm{E}[\mathcal{Y}_{it}|\mathcal{X}_{it}, d_{it}, \epsilon_{it}] = \mathrm{E}[\mathcal{Y}_{it}|\mathcal{X}_{it}, d_{it}]. \tag{5}$$

[21] To reduce the size of the state space and thereby reduce computational burden, H_{it} is a binary variable. It captures the most salient effect of CD4 count, namely, whether it is low enough to suggest AIDS (i.e., <250). Transitions between binary health states are fairly low, reflecting persistence in continuous CD4 count. However, dichotomous health does not reflect that agents with CD4 counts near the cutoff of 250 face a higher probability of switching health states. Nonetheless, once agents who suffered HIV-induced declines in their CD4 go onto HAART, despite their counts rising to non-AIDS levels, they rarely exhibit exceedingly high CD4 counts again. Therefore, although an extension of the current model would capture this difference by permitting H_{it} to take on more values, there is little evidence that such an addition would change the model implications.

[22] Note that F_{it} and C_{it} are not state variables so do not belong to S_{it}, but do affect utility. Such variables are often deemed 'payoff' or 'outcome' variables.

Ailments F_{it} evolve according to:

$$\mathrm{P}\left[F_{it}=1|X_{it}^{F};\theta^{F}\right]=\frac{\exp(X_{it}^{F}\theta^{F})}{1+\exp(X_{it}^{F}\theta^{F})} \tag{6}$$

where $X_{it}^{F}\equiv\left[H_{i,t-1},v_{i,t-1},H_{i,t-1}\times d_{it}^{M}\right]$ and θ^{F} is a vector of parameters governing the process generating ailments.

Consumption is equal to income (I_{it}) minus out-of-pocket treatment costs (p_{it}).[23] Formally,

$$C_{it}=I_{it}-p_{it}\ . \tag{7}$$

Evaluating expected consumption requires several steps since agents face uncertainty on both income and treatment costs. Agent income uncertainty reflects unanticipated shocks. For example, an agent may fall ill at some point before the end of period t and incur an income loss for missing work days. Agents also form expectations on out-of-pocket treatment costs (p_{it}). These are a function of underlying health at the end of period t (H_{it}) and period t insurance provision (N_{it}), both of which are unknown at the beginning of period t. This setup reflects that, after agents choose a treatment category at t, out-of-pocket treatment costs will depend on their (as yet unrealized) health state throughout the period. In summary, to derive expected consumption given period t choices and states, the agent must form expectations on income (I_{it}), insurance (N_{it}), CD4 count (H_{it}) and out-of-pocket treatment costs (p_{it}). Each of these stochastic processes is explained in turn. Income is modeled as

$$I_{it}\ \ =\ \ X_{it}^{I}\theta^{I}+\epsilon_{it}^{I} \tag{8}$$

where $X_{it}^{I}\equiv\left[(E_{i,t-1},E_{i,t-1}^{2},A_{i,t-1},H_{i,t-1},v_{i,t-1})\times d_{it}^{L}\right]$, $\epsilon_{it}^{I}\sim N\left(0,\sigma_{I}^{2}\right)$ and θ^{I} denotes a vector of parameters governing the income process.[24] Note that state variables affecting the income process are interacted with period t employment decisions. This reflects that an agent's current state can affect wage and non-wage income in different ways.

Insurance status (N_{it}) affects treatment costs and is also modeled as a process determined

[23] Agents in the model cannot save. The potential impact of this assumption is discussed as results are presented.

[24] Income is a function of health at the beginning of the period $H_{i,t-1}$. This modeling choices reflects the timing of income offers and employment decisions. After learning his health status, the agent faces income offers for full-time employment. Employers know agent productivity, which is a function of health and human capital. The employer does not, however, know which medications will be chosen, so the income offer is not a function of expected ailment status.

by state variables and labor supply decisions.[25] Formally,

$$P\left[N_{it}|X_{it}^{N};\theta^{N}\right]=\frac{\exp(X_{it}^{N}\theta^{N})}{1+\exp(X_{it}^{N}\theta^{N})}, \tag{9}$$

where $X_{it}^{N}=\left[H_{i,t-1},E_{i,t-1},E_{i,t-1}^{2},A_{i,t-1},A_{i,t-1}^{2},v_{i,t-1},d_{it}^{L}\right]$ and θ^{N} is a vector of parameters governing the insurance process.

Underlying health, as measured by CD4 count, is affected by treatments. The salient features to be captured are (a) whether treatment (or lack thereof) moves CD4 above or below AIDS levels and (b) possible persistence in CD4 count. First, ΔH_{it} indicates whether an agent's CD4 increased (versus either decreased or remained unchanged) between periods t and $t+1$. ΔH_{it} evolves according to

$$\mathrm{P}\left[\Delta H_{it}=1|X_{it}^{\Delta H},d_{it}^{M};\theta^{\Delta H}\right] \;=\; \frac{\exp(X_{it}^{\Delta H}\theta^{\Delta H})}{1+\exp(X_{it}^{\Delta H}\theta^{\Delta H})}, \tag{10}$$

where $X_{it}^{\Delta H}\equiv\left[H_{i,t-1},v_{i,t-1},d_{it}^{M}\times H_{i,t-1}\right]$. In other words, both treatments and period t CD4 count determine if CD4 count increases or not. Then, period t CD4 count and the direction of change ΔH_{it} determine whether CD4 is above or below AIDS levels in $t+1$. In particular, for parameters θ^{H}, the CD4 count process is modeled as:

$$\mathrm{P}\left[H_{it}=1|X_{it}^{H},d_{it}^{M};\theta^{H}\right]=\frac{\exp(X_{it}^{H}\theta^{H})}{1+\exp(X_{it}^{H}\theta^{H})} \tag{11}$$

where $X_{it}^{H}\equiv[\Delta H_{it}\times H_{i,t-1}]$.

Out-of-pocket treatment costs are modeled as

$$p_{it} \;=\; X_{it}^{P}\theta^{P}+\epsilon_{it}^{P}, \tag{12}$$

where $X_{it}^{P}\equiv\left[H_{it}\times F_{it},I_{it},N_{it}\times d_{it}^{M},v_{it}\right]$, $\epsilon_{it}^{P}\sim N\left(0,\sigma_{P}^{2}\right)$ and θ^{P} is a vector of parameters.[26] Given the processes specified above, expected consumption is formally defined as:

$$\mathrm{E}\left[C_{it}|\mathcal{X}_{it},d_{it}\right]=\mathrm{E}\left[I_{it}|I_{it}\geq 0,\mathcal{X}_{it},d_{it}\right]-\mathrm{E}[p_{it}\left|p_{it}\geq 0,\mathcal{X}_{it},d_{it}\right]. \tag{13}$$

Note that both income and treatment costs are assumed to be non-negative.

[25] Insurance could also be modeled as a choice. However, MACS includes no data on insurance options. Also, insurance provision is highly persistent in the data and largely dependent on employment, so I model insurance provision as a process that agents indirectly control through their labor supply decisions.

[26] Note that the costs process includes I_{it} to account for the possibility that treatments are subsidized according to income.

Until now, I have described the stochastic processes governing each component of flow utility. The model is dynamic in the sense that, in making his current decision, the agent must also evaluate how his choices and current state affect the distribution over future states. Formally, define the state-to-state distribution function for current (observable) state $\mathcal{X}_{it}$, current choice d_{it} and period $t+1$ (observable) state $\mathcal{X}_{i,t+1}$ as

$$G_X(\mathcal{X}_{i,t+1}|\mathcal{X}_{it}, d_{it}). \tag{14}$$

I further assume that the distribution over future states is independent of current unobservable state variables $\epsilon(d_{it})$ conditional on current observable state variables and choices. Formally,

$$\mathrm{E}[\mathcal{X}_{i,t+1}|\mathcal{X}_{it}, d_{it}, \epsilon_{it}] = \mathrm{E}[\mathcal{X}_{i,t+1}|\mathcal{X}_{it}, d_{it}]. \tag{15}$$

Furthermore, note that $H_{i,t-1} \in \mathcal{X}_{it}$ evolves according to equation (11).

Full-time work experience at t, $E_{i,t-1}$ increases by 0.5 for each period of full-time employment. Formally, $E_{it} = E_{i,t-1} + 0.5 \times \mathbf{1}\left[d_{it}^L = 1\right]$. Next, age at t $A_{i,t-1}$ and the time dummy $v_{i,t-1}$ evolve deterministically. Specifically, $A_{it} = A_{i,t-1} + 0.5$ and $v_{it} = v_{i,t-1} + 1$.

Finally, the probability of dying between periods t and $t+1$ is denoted

$$\mathrm{P}\left[B_{it} = 1|X_{it}^B; \theta^B\right] = \frac{\exp(X_{it}^B \theta^B)}{1 + \exp(X_{it}^B \theta^B)} \tag{16}$$

where $X_{it}^B = [H_{i,t-1}, A_{i,t-1}, H_{i,t-1} \times A_{i,t-1}]$, B_{it} is an indicator function for death and θ^B is a vector of parameters that govern death probability. Current period decisions do not affect the probability of dying; upon entering the period and learning his state variable realizations, the agent either continues on to enjoy period t flow utility or dies, in which case he receives flow utility 0 forever.

3.3 Parameters and Unobserved Heterogeneity

Flow utility parameters from equation (2) are collected into a vector denoted θ^U. Parameters governing processes and transition probabilities are denoted θ^{XY} so that

$$\theta^{XY} \equiv \left[\theta^F, \theta^I, \theta^N, \theta^H, \theta^P, \theta^B\right]. \tag{17}$$

Collect these parameters into a vector θ so that $\theta \equiv \left[\theta^U, \theta^{XY}\right]$.

Unobserved heterogeneity is introduced into a subset of utility parameters via latent types, of which there is a finite number K^U. I allow the following preference parameters

to vary by type: the utility cost of work, the cost of ailments, the interaction between the two along with the marginal utility of consumption for each ailment status. This modeling decision arises from high observed persistence in labor supply choices within individuals over time, which is consistent with heterogeneity in distaste for work. Parameters governing health parameters can also vary by unobserved type, the number of which is denoted K^{XY}. I permit unobserved heterogeneity in: the effectiveness and side effects profiles of HAART (subsets of θ^D and θ^F) and in parameters governing health transitions θ^H. This is motivated by research suggesting that unobserved factors, including genetic variations, can imply different reactions to HAART (see, for example, Scherer (2010)).[27] The joint distribution of latent preference types and latent health types is also freely estimated, which means that the total number of unobserved latent classes is $K \equiv K^U \times K^{XY}$. For the remainder of this study, I set $K^U = 2$ and $K^{XY} = 2$ so that $K = 4$.[28] Let θ^k denote latent class-k parameters, where $k \in \{1, \ldots, K\}$. Denote agent i's parameters as θ_i. Type probabilities are given by:

$$\pi_k \equiv \mathrm{P}[\theta_i = \theta^k], \tag{18}$$

where

$$\sum_{k=1}^{K} \pi_k = 1. \tag{19}$$

The subject knows his type k, but the econometrician does not, which means that the distribution over types must be integrated out and the π_k's jointly estimated. Finally, collect all parameters to be estimated into a vector ψ, where

$$\psi = \left[\theta^1, \ldots, \theta^K, \pi_1, \ldots, \pi_K\right]. \tag{20}$$

This concludes the specification of the theoretical model. The following section describes how ψ is estimated.

[27] In principle, all parameters could vary by latent type. I have experimented with a variety of specifications permitting unobserved heterogeneity in parameters governing both health and labor market processes, but cannot reject that other parameters do not vary by type.

[28] Experimentation with larger numbers of suggests this is a good number as the search algorithm places very small probability on a third preference or transition type.

4 Estimation

The vector of parameters ψ is estimated using a nested procedure.[29] At the "inner" step and given a proposed a set of parameters (denoted $\psi^{(g)}$), the dynamic programming problem is solved via backward induction for each set of observed state variables $\mathcal{X}_{it}$. This yields a set of transitions and choice probabilities, which maximize utility. At the "outer" step, the algorithm searches for parameters that maximize a likelihood function computed from the data.

The structure of the value functions for retired and non-retired agents differs and each will be described in turn. The value of retirement is an infinite stream of flow utility supposing that agents no longer work, given by

$$\tilde{U}(C_{it}, F_{it}, d_{it}^M, d_{it}^L = 0|\mathcal{X}_{it})) \tag{21}$$

where $\tilde{U}(\cdot)$ is flow utility as defined in equation (2) with the utility-shifter netted out. Agents receive this flow utility at all post-retirement ages, though in each period weighted by the discount factor β and the probability of dying conditional on state variables at retirement $\mathrm{P}[B_{it} = 1|\cdot]$.[30] Therefore, total retirement value for a given treatment choice and set of state variables is equal to an infinite sum, given by.[31]

$$\begin{aligned} V^R(A_{i,t-1} = 65, S_{it}) &= \\ &\left[\frac{\mathrm{P}[B_{it}|\cdot]}{1-\beta\ \mathrm{P}[B_{it}|\cdot]} \times \tilde{U}\left(C_{it}, F_{it}, d_{it}^M, d_{it}^L = 0|\mathcal{X}_{it}\right)\right] + \epsilon(d_{it}). \end{aligned} \tag{22}$$

Let us now turn attention to non-retired agents. In every period t, they choose $d_{it} \in D_t$ to maximize

$$\mathrm{E}\left[\sum_{j=0}^{T_i-1} \beta^j U(C_{i,t+j}, F_{i,t+j}, d_{i,t+j}|\mathcal{X}_{it}) + \beta^{T_i} V^R(A_{i,t-1} = 65, S_{it})\right] \tag{23}$$

where $T_i \equiv (65 - A_{it}) \times 2$ represents the number of periods until retirement. Using the

[29] I employ estimation methods developed by Rust (1987) and Hotz and Miller (1993) and surveyed in Aguirregabiria and Mira (2010).

[30] The discount factor β is set to $\sqrt{.95}$ per semester.

[31] This structure assumes thats agents remain in the same health state and make the same treatment choice in each period after they retire. This is a reduced-form way to capture that good health is valuable at retirement. Further, allowing V^R to be a function of both S_{it} and A_{it} is a slight abuse of notation since A_{it} is an element in the vector S_{it}. Strictly speaking, S_{it} in this case refers to the vector of observable state variables without A_{it}.

Bellman principle, we can define the value function for periods before retirement as follows:

$$V(S_{it}) = \max_{d_{it} \in D_t} \left\{ \mathrm{E}[U(C_{it}, F_{it}, d_{it})] + \beta \int V(S_{i,t+1}) \mathrm{d}G_X(\mathcal{X}_{i,t+1} | \mathcal{X}_{it}, d_{it}) \right\} \tag{24}$$

where $G_X(\mathcal{X}_{i,t+1}|\mathcal{X}_{it}, d)$ is defined in equation (14). Choice-specific value functions can be written as:

$$v(S_{it}, d_{it}) \equiv \mathrm{E}[U(C_{it}, F_{it}, d_{it})] + \beta \sum_{\mathcal{X}_{i,t+1}} \bar{V}(\mathcal{X}_{i,t+1}) g_X(\mathcal{X}_{i,t+1} | d, \mathcal{X}_{it}), \tag{25}$$

where $\bar{V}(\cdot)$ is the expectation of the value function taken over the distribution of $\epsilon(d_{it})$ and $g_X(\cdot)$ is the transition density of $\mathcal{X}_{it}$ corresponding to transition distribution function $G_X(\cdot)$. Notice that $\bar{V}(\cdot)$ takes the form of an expected maximization since the agent does not know future realizations of ϵ_{it}.

Given this setup, I obtain choice probabilities for each set of observable variables via backward induction.[32] For example, suppose that agent i enters period t at age 64.5, so that $A_{i,t-1} = 64.5$. Then, each choice will imply a probability distribution over $\mathcal{X}_{i,t+1}$, from which I compute expected retirement value. Given state-specific retirement value, I compute choice-specific value functions for each state at age 64.5. Once I have obtained choice and state-specific value functions for age 64.5, I can compute choice and state specific value functions for age 64 using equation (25) and so on until age 30. I do not observe $\epsilon_{it}(d)$, but its distribution implies the following choice probabilities:

$$\mathrm{P}(d_{it} | \mathcal{X}_{it}) = \frac{\exp\{\tilde{V}(\mathcal{X}_{it}, d_{it})\}}{\sum_{d'_{it} \in D_t} \exp\{\tilde{V}(\mathcal{X}_{it}, d'_{it})\}} \tag{26}$$

where $\tilde{V}(\cdot)$ is the net-of-error choice specific value function (i.e., equation (25) minus $\epsilon_{it}(d_{it})$):

$$\tilde{V}(\mathcal{X}_{it}, d_{it}) = \mathrm{E}[U(C_{it}, F_{it}, d_{it})] + \beta \sum_{\mathcal{X}_{i,t+1}} \bar{V}(\mathcal{X}_{i,t+1}) g_X(\mathcal{X}_{i,t+1} | d, \mathcal{X}_{it}). \tag{27}$$

Finally, in the preceding derivations, I have omitted notation identifying type-specific parameters. For each set of suggested parameters $\psi^{(g)}$, the estimation routine includes solving the dynamic programming problem to obtain choice probabilities for each set of type-specific

[32] Experience (E_{it}) is measured at five grid points, but estimation requires evaluating value functions between these grid points. For example, if an agent with 10 periods of experience decides to work in period t, his period $t+1$ experience will be 11. I use linear-spline interpolation (see Judd (1998)) to compute value functions for state variable values that lie between grid points.

parameters θ^k. The likelihood contribution of individual i is therefore:

$$\begin{aligned} L_i(\theta) &= \textstyle\sum_{k=1}^{K} \pi_k [\Pi_{t=1}^{T_i} \mathrm{P}(d_{it}|\mathcal{X}_{it};\theta^k) \times \Pi_{t=1}^{T_i} g_Y(\mathcal{Y}_{it}|\mathcal{X}_{it},d_{it};\theta^{XYk}) \\ &\times \Pi_{t=1}^{T_i-1} g_X(\mathcal{X}_{i,t+1}|\mathcal{X}_{it},d_{it};\theta^{XYk})], \end{aligned} \tag{28}$$

where g_Y denotes the density function derived from processes governing F_{it} and C_{it} and θ^{XYk} denotes type-specific θ^{XY}.[33]

4.1 Identification

This section discusses how moments in the data identify estimated model parameters. In the data, each period t choice and state combination implies a probability distribution over period $t+1$ states and these moments identify parameters governing state-to-state transitions and outcomes. Parameters in the flow utility function are identified through observed state-dependent choice probabilities. Here, I exploit the quasi-experimental nature of HAART introduction, which implies that the same decision-makers are observed making choices over time, facing unanticipated variation in the features of available products.

The CRRA coefficient γ, which measures the curvature in the utility function, is identified by differences in how agents choose both treatments and employment at different consumption levels. Employment decisions imply large changes in consumption and treatment choices, which induce variability in medical expenditures, imply small changes in consumption. Insofar as choice probabilities for given state variables change at different rates for different consumption levels, these choices trace out the marginal utility of consumption. State-dependent utility parameters are identified through differences in γ across health-status. Finally, parameters describing the distribution of latent types and type-specific parameters are identified through repeated observed choices of the same subject over time and given different values of state variables.

[33] Portions of equation (28) can be extracted from the summation over k in cases where equation parameters are constrained to be equal across types, e.g., in the equations describing the income, insurance and out-of-pocket treatment cost processes. The log likelihood function then consists of additively-separable components that can be separately maximized and parameters outside of the sum over types can be estimated in a separate first step, which does not involve solving the dynamic programming problem. This first step requires estimation of a set of Tobit, logistic and multinomial logistic regressions, all of which can be accomplished with standard statistical software. In the second step, I only search for remaining parameters along with probability masses π_k. This decreases the number of iterations, which greatly reduces computation burden. Standard errors are computed taking the variance of first-stage estimates into account.

4.2 Parameter Estimates

This section presents estimates of type probabilities (π_k), preference parameters (θ^U) and parameters governing outcomes and transitions (θ^{XY}). I then compute a posterior type probability for each individual in the dataset to investigate of how latent types relate to variables not included in the model state space.

Recall that there are two preference types and two health types. As will be discussed below, preference Type I agents suffer an additional utility cost of working when suffering from physical ailments. For health Type I agents, HAART is relatively more effective and causes more side effects. For each individual, preference type can be correlated with health type, i.e., there four possible type-combinations and the probability of each is freely estimated. About half the population is estimated to be preference Type I and 40% of agents correspond to health Type I. Among preference Type I agents, about 30% correspond to health Type I; among preference Type II agents, about half correspond to health Type I (see Table 5).

Preference parameter estimates (found in Table 6) reveal that both preference Type I and II agents experience a utility cost of ailments that far outweighs the utility cost of work. The key difference between preference types lies in the utility cost of working while suffering ailments. For Type I agents, this cost is about twice the analogous value for Type II agents. This difference has far-reaching consequences for agent behavior; given their preferences, Type I agents are more likely to avoid employment while suffering symptoms or side effects. These agents can essentially attenuate the utility cost of ailments by choosing not to work. Both preference types indicate an increase in the marginal utility of consumption with ailments, meaning that, on average, agents tend to consume goods that they value more when they are sicker. Finally, switching costs vary by ailment status. For agents free of ailments, it is costly to switch or end treatment, but there is a utility gain implied by beginning treatment. For agents with ailments, beginning treatment is costly, but ending treatment carries a benefit after controlling for the impact of this choice on other components of utility, including health.

Moving on to health transitions and outcomes, the model reveals unobserved heterogeneity in drug effectiveness and side effects (see Table 7). For both latent health types, HAART is the most effective treatment in terms of increasing CD4 count. Differences between health Types I and II emerge when considering agents with low CD4 counts. For Type I agents, HAART is vastly superior to previously available treatments. For Type II agents, HAART is a more limited improvement over combo-therapy. For all latent types, mono-therapy and combo-therapy, though less effective than HAART, are more effective than no treatment.

These differences in HAART effectiveness will imply different valuations of HAART between different health types. Going back to Table 5, estimated correlations between health and preference types imply that individuals for whom HAART is both highly effective and harsh are less likely to suffer a high additional utility cost of working with physical ailments.

Predicted values from this regression of the probability of a CD4 count increase are included as regressor in the model explaining period $t+1$ CD4 count. Results indicate that for both health types, high CD4 at t along with a higher predicted probability of a CD4 increase independently predict high period $t+1$ CD4 count.[34] The only coefficient that significantly varies by health type is that governing the interaction between a high CD4 count and an increase in the predicted probability of a CD4 count increase. This coefficient essentially measures the effectiveness of medication on agents who are already in relatively good health and is estimated to be much higher for Type I agents. This means that health Type I agents face a stronger incentive to remain on HAART while in good health. As will become apparent in the following section, differences in health transition probabilities imply differences in the valuation of counterfactual medical innovations specified with the same effectiveness. Type I agents attain a higher CD4 count with higher probability even with less effective medications. They therefore have less to gain from an effective medication and value it accordingly.

Estimates of parameters governing the side effects process are found in Table 8 and show that an absence of ailments is associated with higher CD4, which reflects that agents in better health are less likely to suffer ailments, i.e., symptoms (recall that $F_{it}=1$ indicates the absence of ailments during period t). Treatments also cause ailments via side effects and it generally holds that more effective treatments like HAART imply the harshest side effects, with the effect being stronger for health Type I agents. Regarding survival (Table 9), a high CD4 count drastically reduces death probability. Since higher age can signal good health among HIV+ subjects, I interact age with high CD4. The positive estimated coefficient indicates that HIV+ subjects with high CD4 counts face higher death probability as they age.

Estimates of parameters governing the income, insurance and treatment cost processes are found in Tables 10, 11 and 12, respectively. Income is modeled to be a function of: high CD4, experience, experience-squared, age, a time-trend along with current employment fully interacted with these variables. Income increases with human capital (as measured by expe-

[34]Recall from the previous section that this method of obtaining transition probabilities among health states is designed to capture two salient features of treatment technology without increasing the size of the state space, namely, treatment effectiveness at increasing CD4 count and whether this increase brings agents to non-AIDS immune system health.

rience), but at a decreasing rate. Age independently predicts a lower wage and good health is associated with higher income. The positive relationship between experience and income for non-full-time workers is consistent with increased non-wage income (e.g: disability payments) given a longer work history. For full-time workers, and with the exception of health, these effects are more pronounced. The effect of health on income is weaker for employed workers, which likely reflects that AIDS counts as a disability so that unemployment benefits are high for sick agents

Health insurance (public, private or no insurance, the latter being the base category) is modeled as a function of CD4 count, age, labor supply and experience. Low CD4 and higher age predict a higher probability of public insurance, which may again reflect that AIDS is considered a disability and that medicare eligibility is age-dependent. Also, full-time employment predicts private insurance provision, but predicts a lower probability of public insurance. Finally, treatment costs are a function of treatment choice, health status and insurance. Estimates indicate the following: HAART is more expensive than other treatments; costs increase over time; healthier subjects spend less on their medications; and both higher income and private insurance are associated with higher treatment costs.[35]

To gain further insight into the labor market heterogeneity captured by modeling latent types, I compute average 'posterior' type probabilities for a given set of labor market characteristics.[36] Next, I average over individuals for a given set of observable labor market characteristics, including education, race, and occupation category reported at the baseline interview.[37] This exercise permits an analysis of the correlation between unobserved latent type and labor market factors not included in the model state space. Results are presented in Table 14. As an example, low education agents (less than a college degree) belong to preference Type I with an average probability of 54% (versus 48% for the entire sample and 45% for college educated agents). Recall that preference Type I agents are more reactive to ailment status in choosing whether or not to work. It is likely that less educated agents tend to work in occupations in which feeling ill makes work especially difficult, e.g., those requiring inflexible schedules or physical labor. To explore this possibility, I compute average

[35]Results from a model fit exercise are found in Table 13. Taking current states as given, agent choices are simulated and then compared with state-dependent choices found in the data. The model successfully matches dynamics found in the data, though employment probability is overestimated by about 7 percentage points given low CD4 counts in the post-HAART era. This occurs given a low number of agents with AIDS-level CD4 after HAART is introduced.

[36]Specifically, for each preference and health type combination, I construct likelihood contributions for each individual. Then, I divide by each individual's actual likelihood contribution, given estimated unconditional type probabilities. The resulting 'posterior' ratios measure how likely a given individual, given his behavior and outcomes, belongs to each type combination.

[37]Available data does not offer more specific occupational information. Moreover, occupation data is asked only in 1984, so health-induced occupation change is impossible to measure.

type probabilities by occupation. Indeed, individuals in the service industry (e.g. waiters) or who work in extractive industries (e.g. mining) are more likely to be preference Type I versus individuals who have a professional specialty or work in a craft industry, both of which may permit more flexible schedules.[38]

5 The Value of Pharmaceutical Innovation

In this section, I use the estimated model to place a value on pharmaceutical innovations, including HAART. Contrary to previous studies that estimate drug demand using drug or molecule dummy variables, I exploit the characteristics approach to evaluate counterfactual treatments, each defined as a bundle of attributes. Next, I examine how a dynamically optimal treatment policy exhibits cycles. Specifically, sicker agents choose effective treatments despite harsh side effects and switch to less effective drugs with fewer side effects once their health improves.

5.1 The Value of HAART

This section converts the value of HAART into a measure of willingness-to-accept-payment in dollars (henceforth: WTAP). To compute WTAP for HAART, I compare computed value function values in the first post-HAART period to analogous values under the counterfactual scenario in which HAART is never introduced. Instead, a counterfactual treatment technology is introduced with the same attributes as combo-therapy. Next, I compute what per-period payment (similar to an annuity) under the counterfactual scenario is required to make agents indifferent to HAART introduction.[39] Finally, I compute the present discounted value of this annuity using expected years of life, which is also simulated with estimated model parameters.

Results for each type combination are presented in Figure 2, which graphs the present

[38]The same exercise is performed for both latent health types, but there are few noticeable differences, which means that unobserved factors determining drug effectiveness and side effects (e.g. genetic differences influencing biological responses to medications) are independent of occupation, education or race. The exception is craft industries, for which the likelihood of HAART being highly effective is large. If agents in craft industries are more likely to be self-employed and therefore have greater freedom to enter and exit the labor market, they would cycle on and off HAART more aggressively. The model would explain this with a higher drug match value among agents in this occupational category. Dropping craft workers would not appreciably affect results since they comprise 1% of individuals in the data set.

[39]Valuations of both factual and counterfactual treatment introductions capture present discounted value of utility in all future periods, i.e., gains in years-of-life weighted by utility over time. Further, valuation via an annuity essentially imposes savings onto individuals. Therefore, permitting savings behavior is not expected to appreciably affect these results.

value of future per-period payments for Type I agents of different ages and levels of human capital. Two key findings emerge. First, HAART has a high potential value: worth $180,000 for a 38-year-old with 15 years of work experience. Second, there is striking heterogeneity in the value of HAART. In opposition to standard critiques, this heterogeneity in value across individuals suggests why 'me-too' drugs can create value: a 'me-too' drug that, on average, is therapeutically similar to existing options may be welfare-enhancing for some subsets of agents (if not others), distinguished by observed and unobserved factors affecting demand. Indeed, Figure 2 shows that older agents value HAART less since their life horizon is shorter, implying fewer years during which they benefit from HAART. This effect is compounded for younger agents since health gains made earlier in life persist over time. Further, agents with higher human capital value HAART more since each life-year gained entails higher consumption. For example, a 45-year-old with high human capital values HAART at over $160,000, whereas a lower human capital agent values it at about $20,000.

Latent types also exhibit vastly different valuations of HAART. Health Type II agents value HAART at less than $30,000, which reflects the low probability that HAART improves their CD4 count in comparison to their Type I counterparts. Regarding latent preferences: Type I agents value HAART slightly less than preference Type II agents. This difference reflects how preference Type I agents can essentially attenuate the utility cost of suffering ailments by not working. Hence a treatment that ultimately improves their ailment status by lowering symptoms yields less value. Moreover preference Type I agents are more likely to exit the work force due to side effects, which slows their accumulation of human capital and lowers their expected future income were they to stay alive and on HAART. This effect is reflected in a lower valuation of a treatment that will keep them alive—but poorer—in comparison to their preference Type II counterparts.

5.2 Optimal Treatment Cycling

When no available treatment dominates along both dimensions of quality (efficacy and side effects), agents optimally choose to cycle among available treatments. An optimal treatment path is therefore a non-stationary closed loop driven by three factors: (1) persistence in underlying health; (2) a non-convexity in discrete treatment choices and (3) health-state dependent flow utility captured by the estimated disutility of ailments induced by symptoms or side effects. To fix ideas, note that cycling may not occur in other medical contexts where one of these components is not present. For example, in the case of diabetes, health deterioration is immediate absent treatment with insulin. Therefore, cycling off of insulin to enjoy periods free of side effects would be a short-lived endeavor and likely not part of an

optimal dynamic plan.

Details of cycling behavior indicate that agents with AIDS-level CD4 count are likely to switch to the more effective treatment (in this case HAART), akin to a phase of investment in health 'stock'. While on HAART, agents face a higher probability of health improvements. Once their health improves, some agents switch back to less effective treatment with fewer side effects (including the no-treatment option). At this point on the cycle, agents essentially exploit previous investments in their health stock, trading a higher probability of diminished future health for several periods with fewer side effects. During these periods, agents are more likely to engage in full-time employment. Treatment cycling rationalizes systematic avoidance of HAART as part of an optimal dynamic plan. Moreover, and as will become evident in the following section, optimal treatment cycling is the key mechanism through which counterfactual environments affect agent choices and outcomes. Agents respond to counterfactual environments primarily through shifts in the frequency of going off (and staying off) of HAART in good health and going back onto HAART in poor health.

Consider Figure 3, which illustrates the anatomy of optimal treatment cycling for agents with preference Type I and health Type I. These are agents who face a high utility cost of working while suffering physical ailments and for whom HAART is vastly more effective, but with harsher side effects, versus other available treatments. Behavior is simulated in an environment where available treatments correspond to actual options in the factual post-HAART world: no treatment, mono-treatment, combo-treatment or HAART. In any given period when agents are healthy and on HAART, about 8% switch off of HAART. Of these, 84% remain off of HAART and face a 3% probability of AIDS level CD4 in each period. During periods in good health, agents are more likely to work when off HAART (44% versus 40%). When they become ill, these agents go onto HAART with 97% probability and remain on HAART with nearly 100% probability. They face a 50% chance of regaining non-AIDS CD4, at which point the cycle begins again.

Different latent types exhibit different cycling behavior. For example, agents for whom HAART is relatively less effective (health Type II agents) are less likely to go onto HAART once their health deteriorates (40% versus 84%). This difference reflects that these agents face a 16% probability of health improvements (versus 50% for health Type I agents). Comparing preference types, agents who face a disutility of working with ailments (preference Type I) are less likely to stay off HAART when in good health. This occurs because they can essentially attenuate the utility cost of ailments by exiting employment rather than by facing the health consequences of going off of HAART.

Treatment cycling is often considered to be a form of suboptimal non-compliance that

should be curbed (Sabate, 2003). Switching off of treatment is sometimes referred to as a 'drug holiday' and some medical literature points to individual-level dangers of engaging in such behavior (Meredith, 1996). In contrast, I find that a cyclical treatment pattern can be the result of optimal forward-looking behavior and refer to the phenomenon as optimal treatment cycling. Recent medical research on long-term, chronic illness suggests adapting current treatments to patient responses to previous treatment (Murphy, 2005).[40] Optimal treatment cycling is similar in that current decisions reflect previous treatment outcomes, though it is driven by patient decision-making. Cycling is also consistent with findings that medical doctors, despite advocating highly effective treatments for their patients, often opt for less effective drugs with fewer side effects when faced with similar medical conditions (Ubel, Angott, and Zikmund-Fisher, 2011).

5.3 The Value of Counterfactual Treatment Innovations

A key benefit of the characteristics approach to modeling treatment quality used in this study is the possibility to evaluate counterfactual treatment innovations. For example, suppose that once HAART is introduced, patients are faced with an improvement on HAART along one or both dimensions of drug-quality. One possibility is a version of HAART without side effects. Computing WTAP as in Section 5.1, I present valuations of such an innovation for low-human-capital patients of different ages in Figure 4 (for preference Types I-II and health Type I) and Figure 5 (for preference Types I-II and health Type II). Counterfactual innovations occur once HAART has already been introduced. In each figure and for each age, black bars depict the value of HAART introduction. Given HAART, the value of HAART without side effects is depicted by the difference between the black bars and the dark grey bars to the immediate right. A version of HAART without side effects has enormous potential value: between \$100,000 and \$125,000 (for a 30-year-old belonging to health Types II and I, respectively). Health Type I agents exhibit higher willingness-to-pay since HAART is a more effective drug for them. Consistent with previous results, older agents value the innovation less since they have fewer periods to enjoy it.[41] This valuation is especially striking since the innovation entails no improvement on underlying health or longevity. In this sense, a version of HAART without side effects could be seen as a 'me-too' innovation since, by design, it is therapeutically equivalent to an existing treatment. Contrary to arguments that 'me-too' innovations offer little benefit to consumers, I find that a treatment that is therapeutically

[40] Specifically, this line of research suggests designing medical trials involving multiple randomizations to better formulate decision rules for adaptive treatments.

[41] According to results that are not shown and are consistent with HAART valuations, high-experience agents value the innovation more highly than low-experience agents

equivalent to HAART, but entails fewer side effects, generates high value.[42]

Suppose that instead of a reduction in side effects, HAART is improved along the efficacy dimension. In particular, low-CD4 agents who use HAART face a 32% probability of non-AIDS CD4 in the following period. Under the counterfactual improvement, this probability is tripled. For each age, the third great bar in Figures 4-5 depict how agents with different latent types value this innovation. For health Type I agents, this value is about $275,000 (or about $100,000 above HAART). In contrast, health Type II agents would be willing to pay upwards of $1,100,000 for a 30-year-old with five years of accumulated work experience. The massive difference between health Types I and II is explained via differences in health probabilities: Type II agents are more likely to have a low CD4 count and so value an equally effective medical innovation much more highly.

Agents would be expected to place high value on a life-improving and life-saving technology. What is more surprising is that optimal treatment cycling underlies some portion of this value. In general, switching onto milder treatments is risky since the full treatment cycle includes periods where CD4 count is low and death probability is high. If a highly effective version of HAART exists, however, agents anticipate fewer periods of poor health once their health deteriorates. They respond by cycling more aggressively, i.e., by more frequently switching to low side effects treatments once their CD4 count is high. In other words, the value of an effective treatment includes the implied option value of optimally cycling off of it in periods of relatively good health.

Another key finding is that the value of counterfactual treatments depends on existing treatments. Suppose that the two aforementioned innovations (side effects and efficacy) occur simultaneously in separate treatments, so that two new drugs are introduced. Returning to Figures 4-5 and consider the fourth bar for each age. As compared to the efficacy innovation, the two innovations create little additional value. This finding is striking: a side effects innovation is valuable absent an efficacy innovation, but creates little value given an efficacy innovation. Again, the underlying mechanism is optimal treatment cycling: if a highly effective version of HAART already exists, agents can simply cycle off of treatment altogether (avoiding all side effects), retaining the option value of resuming treatment once their health deteriorates. A drug without side effects adds little additional value in such a scenario.[43] Nonetheless, combining an efficacy and side effects innovation into a single drug does imply additional value since it permits agents to live without side effects, but to avoid

[42]See, for example, Angell (2000) for a summary of popular arguments on why 'me-too' drug development should be curtailed.

[43]This does not necessarily imply that a private pharmaceutical firm would not profit from investing in marginal improvements on either dimension of drug quality since a high proportion of patients would presumably switch to the improved treatment despite the small implied value increase.

risks associated with cycling. Such an innovation starts to approximate a cure and its value reflects this: about $450,000 for a health Type I 30-year-old and $1,500,000 for a health Type II 30-year-old.

6 Medical Innovation and the Labor Market

The framework developed in this paper permits an explicit analysis of how pharmaceutical innovation creates value in part through its interaction with labor market choices and outcomes. In what follows, I provide results from three counterfactual policy simulations exploring treatments innovations, a reduction in non-wage income and higher out-of-pocket treatment costs. For illustrative purposes, I present results for preference Type I and health Type I agents, for whom both the effects of HAART and the interaction between health and employment are strong.[44]

6.1 Counterfactual Treatments

In the first policy simulation, I trace agent decisions along with health and labor market outcomes from the time of HAART introduction until the end of the sample period under regimes distinguished by available treatment technologies.[45] I compare three of the treatment scenarios outlined in the previous section. The first is the baseline (factual) regime where HAART is introduced in 1996. In the second, a treatment identical to combo-therapy is introduced at the time of HAART introduction. This scenario mimics a continuation of the pre-HAART world in the sense that a new treatment becomes available, but does not improve upon existing technology. In the third scenario, two counterfactual improvements upon HAART are simultaneously introduced: HAART with no side effects and a highly effective version of HAART with HAART-level side effects. This final scenario illustrates behavior when innovations occur separately along two dimensions of treatment quality. Under each policy, agent behavior is optimal in the sense that choices arise from solution of the dynamic programming problem given estimation preferences parameters. Results are depicted in Figure 6.

For health Type I agents, it is not surprising that HAART brought better average

[44]This choice of latent type is for illustrative purposes. Results for each latent type reflect estimated parameters. Health Type II agents exhibit a relatively weak response to HAART. For preference Type II agents, the labor market effects of health are less apparent.

[45]For each simulation, the distribution of observed state variables at the time of HAART introduction is taken as given, with the exception that all agents are modeled to have chosen "no treatment" in the period immediately preceding HAART introduction.

health (see Figure 6(a)). Perhaps more surprising is that counterfactual improvements upon HAART imply negligible health improvements. In this scenario, a high proportion of agents opt for the version of HAART without side effects. The availability of a highly effective version of HAART encourages this behavior: since they are forward-looking, they maintain the option of using the effective treatment—and quickly recuperating—should they fall ill in the future. The outcome is a lower probability of suffering ailment in comparison to the scenario where only HAART is available (Figure 6(b)). Also apparent in Figure 6(b) is that, on average, fewer agents suffer ailments in the scenario where combo-therapy is the best available technology. Under this regime, agents eschew medication altogether, which lowers average health, but also lowers the probability that they suffer ailments.

Health and physical ailments affect employment decisions, which are depicted in Figure 6(c). Absent HAART, a lower proportion of agents work since expected income at the time of the employment decision is lower, driving some agents out of the labor market. This effect is compounded by a shorter expected lifespan, which weakens the incentive to work to accumulate human capital. Recall, however, that preference Type I agents' employment disutility is sensitive to ailment status. Given improvements to HAART, which bring only small improvements to underlying health, agents work more since they are more likely to be free of physical ailments that increase the utility cost of work. In 1998, for example, employment is 45% given HAART and nearly 53% given improvements on HAART, a 15% increase. Given that preference Type I agents constitute about half of the population, this implies a 7.5% increase in employment among HIV+ men.

The connections between treatment innovations and employment highlight the importance of looking beyond underlying health to quantify the value of medical breakthroughs. Given counterfactual improvements to HAART, the average effect on health is negligible, but agents suffer fewer ailments and return to work. This not only increases their income (Figure 6(d), but also raises the income tax that they would pay, suggesting a mechanism through which public investments in biomedical research could be offset through taxation of the direct beneficiaries of medical innovation.[46]

[46]Preference Type I agents exhibit a fairly low probability of working full-time (between 25% and 55% versus 80% or more for preference Type II agents). This low probability arises, in part, from the timing of labor supply decisions: agents choose whether or not to work for a full period before they know their ailment status. A high enough probability of suffering ailments coupled with a high disutility of labor while suffering ailments, implies that preference Type II agents will often avoid employment.

6.2 A Decline in Non-Wage Income

The introduction of HAART occurred under very specific circumstances since HAART treats a condition that is legally considered a disability, giving patients access to disability payments should they exit the labor market. Therefore, income remains fairly high for agents who choose not to work.[47] The goal of the following experiment is to ascertain agent choices and outcomes in a counterfactual environment where non-wage income is lower. In the simulated environment, agents face reductions in non-wage income, operationalized via decreased parameters of the income process for agents not choosing full-time employment. In effect, non-wage income is reduced 25%, 50% and 75%.

Figure 7 shows that, facing lower non-wage income, health Type I and preference Type I agents engage in more pronounced optimal treatment cycling (compare Figures 7(a) and 7(b)) in order to improve their ailment status. When non-wage income declines to 25% of its original value, the probability that high-CD4-count agents switch off of highly effective treatment rises from 11% to 14% in any given period and the probability that healthy agents stay off of HAART rises from 84% to 93%.[48] Agents move into the labor market, which brings higher income (Figure 7(c)) and also leads to small improvements in ailments status (Figure 7(d)).

The estimated model implies that a subset of agents (latent preference Type I) face a higher utility cost of working with side effects. Faced with lower disability payments, these agents respond by more aggressively cycling off of effective treatments, thereby facing potential health deterioration and a lower probability of survival. Results from this policy simulation show that this possibility is not of great concern in the context studied here. However, the model suggests the possibility of unintended deleterious health consequences arising from lower disability payments, which may be of concern in other medical contexts.

6.3 Unsubsidized Treatment Costs

Recall that HIV+ agents pay on average about $500 per year for treatment. However, the actual costs, paid by insurance (both public and private) is much more. A year of HAART

[47] Under the Americans with Disabilities Act, people living with HIV/AIDS qualify for social security disability payments. These payments cover both symptoms of AIDS and side effects of treatment. Moreover, limited benefits can continue even if agents return to work, reflecting the cyclical nature of chronic disease. For more information, see http://ssa.gov/pubs/10019.html. For more information on government mandated payment calculations, also see: http://www.ssa.gov/policy/docs/statcomps/supplement/2011/.

[48] Preference Type II agents' response to low non-wage income is to slightly increase already high levels of employment. They do not, however, appreciably shift their treatment cycling behavior since they do not experience a utility cost of working with ailments.

therapy costs about $12,000. Combo-therapy costs $8,000 and mono-therapy $6000. What would happen to agent choices and outcomes if they were compelled to pay these unsubsidized costs? The following policy experiment addresses this question, simulating environments where agents would pay 20%, 40% or 60% of the full cost of treatment. Results are presented in Figure 8.

Again, more pronounced optimal treatment cycling is the key mechanism through which changes in the environment affect patient choices. Facing high costs, agents are more likely to switch off (and stay off) of HAART once their health improves (compare Figures 8(a) and 8(b), which depict use of HAART with full subsidies and 50% subsidies, respectively). As a result, agents experience lower average health, though survival probability remains largely unchanged. Agents do exhibit an improvement in their side effects status, shown in Figure 8(c), which encourages an increase in employment (Figure 8(d)). This finding underscores how the connection between health and labor affects medical treatment choices. Here, a decrease in treatment subsidies has an unintended benefit in the form of increased employment, consumption and, from a social perspective, income tax receipts.

7 Conclusion

This project develops a general framework to value medical innovation that includes various measures emphasizing the quality of life and highlights links between health, human capital and the labor market. Contrary to medical literature criticizing treatment non-compliance, I show that in the case of chronic illness, strict adherence to the most effective medication available is not part of an individually rational, dynamically optimal treatment plan. Rather, when no treatment dominates along all dimensions of drug quality, agents cycle among available options. I also find that similarly unhealthy patients exhibit substantial heterogeneity in the way they value a given drug, depending on their age, accumulated human capital along with unobserved factors affecting drug efficacy and side effects. Contrary to standard arguments that 'me-too' drugs imply few benefits to patients, these findings suggest two avenues through which they can create enormous value: by reducing side effects despite no improvements to average drug effectiveness and by generating welfare improvements for certain subsets of the patient population, distinguished by unobservable factors.

Optimal treatment cycling also reveals complex relationships between health and employment, which influence the effects of medical innovation and other policy changes. I show that agents facing unsubsidized drug costs quickly cycle off treatment when in relatively good health. This behavior can damage health, but also reduces ailments induced by side effects,

which encourages employment, thereby increasing income and accelerating the accumulation of human capital. This finding underscores the importance of looking beyond the length of life—to factors affecting the quality of life—to fully appreciate the value of pharmaceutical innovation.

The framework developed in this paper is applicable to other medical conditions, especially those that are chronic, that affect working-age adults and where treatment or prevention is costly, both financially and in terms of side effects. Examples include: diabetes, obesity and depression. Moreover, future research could extend the characteristics approach to other dimensions of treatment quality. For example, insulin pumps arguably increased the convenience of diabetes treatment. Perhaps less important in the face of life-threatening illness, convenience becomes more salient once treatments are effective, side effects are manageable and patients demand innovations that further improve their quality of life.

References

Aguirregabiria, V. and P. Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics* 156 (1):38–67.

Angell, M. 2000. "The Pharmaceutical Industry—To Whom Is It Accountable?" *New England Journal of Medicine* 342 (25):1902–1904.

Arcidiacono, P., P. Ellickson, P. Landry, and D.B. Ridley. 2012. "Pharmaceutical Followers." mimeo, Duke University.

Becker, G.S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3 (1):33–58.

———. 2007. "Health as Human Capital: Synthesis and Extensions." *Oxford Economic Papers* 59 (3):379–410.

Cawley, J. 2004. "The Impact of Obesity on Wages." *Journal of Human Resources* 39 (2):451.

Chan, Tat Y, Barton H Hamilton, and Nicholas W Papageorge. 2013. "Health and the Option Value of Medical Innovation." Mimeo, Johns Hopkins University.

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

Chintagunta, P.K., R. Jiang, and G.Z. Jin. 2009. "Information, Learning, and Drug Diffusion: The Case of Cox-2 Inhibitors." *Quantitative Marketing and Economics* 7 (4):399–443.

Conti, G., J. Heckman, and S. Urzua. 2010. "The Education-Health Gradient." *American Economic Review* 100 (2):234–38.

Crawford, G.S. and M. Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* :1137–1173.

Currie, J. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47 (1):87–122.

Currie, J. and B.C. Madrian. 1999. *Health, Health Insurance and the Labor Market*, *Handbook of Labour Economics*, vol. 3, chap. 50. Elsevier, 3309–3416.

Detels, R., P. Tarwater, J.P. Phair, J. Margolick, S.A. Riddler, and A. Muñoz. 2001. "Effectiveness of Potent Antiretroviral Therapies on the Incidence of Opportunistic Infections Before and after AIDS Diagnosis." *AIDS* 15 (3):347.

Duggan, M.G. and W.N. Evans. 2008. "Estimating the Impact of Medical Innovation: A Case Study of HIV Antiretroviral Treatments." NBER Working Paper.

Eckstein, Z. and K.I. Wolpin. 1989. "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience." *The Review of Economic Studies* 56 (3):375–390.

El-Sadr, WM, J.D. Lundgren, JD Neaton, F. Gordin, D. Abrams, RC Arduino, A. Babiker, W. Burman, N. Clumeck, CJ Cohen et al. 2006. "CD4+ Count-Guided Interruption of Antiretroviral Treatment." *New England Journal of Medicine* 355 (22):2283–2296.

Fernandez, J.M. 2008. "An Empirical Model of Learning under Ambiguity: The Case of Clinical Trials." Mimeo, University of Louisville.

Garthwaite, C. 2012. "The Economic Benefits of Medical Innovations: The Case of Cox-2 Inhibitors." *American Economic Journal: Applied Economics* 4 (3):116–137.

Goldman, D.P. and Y. Bao. 2004. "Effective HIV Treatment and the Employment of HIV+ Adults." *Health Services Research* 39 (6p1):1691–1712.

Grossman, M. 1972. "On the Concept of Health Capital and the Demand for Health." *The Journal of Political Economy* 80 (2):223–255.

Hall, Robert E and Charles I Jones. 2007. "The Value of Life and the Rise in Health Spending." *The Quarterly Journal of Economics* 122 (1):39–72.

Heckman, J.J. and F. Cunha. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

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

Judd, K.L. 1998. *Numerical Methods in Economics*. The MIT press.

Kerbrat, A, E Le Page, E Leray, T Anani, M Coustans, C Desormeaux, C Guiziou, P Kassiotis, F Lallement, D Laplaud et al. 2011. "Natalizumab and Drug Holiday in Clinical Practice: An Observational Study in Very Active Relapsing Remitting Multiple Sclerosis Patients." *Journal of the Neurological Sciences* 308 (1):98–102.

Lakdawalla, D., N. Sood, and D. Goldman. 2006. "HIV Breakthroughs and Risky Sexual Behavior." *The Quarterly Journal of Economics* 121 (3):1063–1102.

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

Lipscomb, J., M. Drummond, D. Fryback, M. Gold, and D. Revicki. 2009. "Retaining, and Enhancing, the QALY." *Value in Health* 12:S18–S26.

Meredith, PA. 1996. "Therapeutic Implications of Drug Holidays." *European Heart Journal* 17 (suppl A):21.

Mincer, J. and S. Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *The Journal of Political Economy* 82 (2):76–108.

Murphy, K.M. and R.H. Topel, editors. 2003. *Measuring the Gains from Medical Research: An Economic Approach*. University of Chicago Press.

Murphy, K.M. and R.H. Topel. 2006. "The Value of Health and Longevity." *The Journal of Political Economy* 114 (5):871–904.

Murphy, S.A. 2005. "An Experimental Design for the Development of Adaptive Treatment Strategies." *Statistics in Medicine* 24 (10):1455—1481.

Ostrow, DG, A. Monjan, J. Joseph, M. VanRaden, R. Fox, L. Kingsley, J. Dudley, and J. Phair. 1989. "HIV-Related Symptoms and Psychological Functioning in a Cohort of Homosexual Men." *American Journal of Psychiatry* 146 (6):737.

Petrin, A. 2004. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110 (4):705–729.

Philipson, T.J. and R.A. Posner. 1993. *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective*. Cambridge: Harvard University Press.

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

Rust, J. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5):999–1033.

Sabate, E. 2003. "Adherence to Long-Term Therapies: Evidence for Action." Tech. rep., World Health Organization, Available: www.who.int/chp/knowledge/publications/adherence_introduction.pdf. Accessed 4 Oct 2011.

Scherer, Robert M. 2010. "Pharmaceutical Innovation." In *Handbook of the Economics of Innovation*, edited by Bronwyn H. Hall and Nathan Rosenberg, chap. 12. Elsevier, 539–571.

Stebbing, J. and A. Dalgleish. 2009. "Not So SMART?" *The Lancet Infectious Diseases* 9 (5):268–269.

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

Ubel, P.A., A.M. Angott, and B.J. Zikmund-Fisher. 2011. "Physicians Recommend Different Treatments for Patients Than They Would Choose for Themselves." *Archives of Internal Medicine* 171 (7):630–634.

8 Tables and Figures

Table 1: Summary statistics

	Full Sample	Pre-HAART†	Post-HAART‡	High CD4	Low CD4	Full-Time	Not Full-Time
Age in 1984 (years)§	32.71	33.16	32.30	32.51	33.29	32.40	33.32
CD4 count (units/mm^3)	448.76	392.54	500.37	566.81	121.86	490.57	366.12
High CD4	0.73	0.65	0.81	1.00	0.00	0.81	0.59
No Ailments	0.59	0.59	0.59	0.65	0.41	0.68	0.41
Death Probability	0.04	0.07	0.02	<0.01	0.14	0.02	0.09
Income ($(2003)/year)	32,348.08	28,040.50	36,301.32	32,125.20	27,198.50	36,067.22	22,818.58
Out-of-Pocket Cost ($(2003)/year)	471.39	333.60	597.85	448.95	533.53	414.83	583.22
Experience (years)	12.56	11.78	13.28	12.68	12.21	12.89	11.91
No Insurance	0.05	0.07	0.03	0.05	0.05	0.04	0.08
Private Insurance	0.81	0.84	0.79	0.84	0.73	0.92	0.59
Public Insurance	0.14	0.09	0.18	0.11	0.22	0.04	0.33
Treatment: None	0.30	0.45	0.18	0.34	0.20	0.33	0.25
Treatment: Mono	0.20	0.32	0.09	0.17	0.29	0.20	0.19
Treatment: Combo	0.17	0.23	0.12	0.17	0.17	0.14	0.25
Treatment: HAART	0.33	.	0.62	0.35	0.26	0.30	0.39
Not full time	0.34	0.32	0.36	0.28	0.54	0.08	0.89
Full time	0.66	0.68	0.64	0.72	0.46	0.92	0.11
N [time t variables]	8,300	3,972	4,328	6,098	2,202	5,512	2,788
N [time $t+1$ variable]	7,954	3,694	4,260	6,057	1,897	5,403	2,551

† The pre-HAART era contains observations from 1991 until mid 1995 (9 periods).

‡ The post-HAART era contains observations from 1996 until 2003 (17 periods).

§ Each entry represents the mean over individuals and time periods for the period or group in question. Entries are proportions unless otherwise indicated.

Table 2: Transition Matrix

		Time $t+1$			
		None	Mono	Combo	HAART
		Pre-HAART & Low CD4			
Time t	None	0.67	0.23	0.10	.
	Mono	0.13	0.59	0.28	.
	Combo	0.06	0.29	0.64	.
		Pre-HAART & High CD4			
Time t	None	0.91	0.07	0.02	.
	Mono	0.07	0.74	0.19	.
	Combo	0.05	0.20	0.75	.
		Post-HAART & Low CD4			
Time t	None	0.53	0.05	0.07	0.35
	Mono	0.04	0.46	0.14	0.36
	Combo	0.04	0.06	0.37	0.53
	HAART	0.04	0.06	0.04	0.87
		Post-HAART & High CD4			
Time t	None	0.87	0.01	0.04	0.07
	Mono	0.03	0.68	0.07	0.21
	Combo	0.03	0.02	0.74	0.22
	HAART	0.02	0.03	0.01	0.94

HIV treatment choices.

Table 3: Logistic Regression of Employment Decisions

	Employment Choice $t+1$ Conditional on Treatment Choices				
	(1)	(2)	(3)	(4)	(5)
				Pre-HAART	Post-HAART
Full time (t)	4.26***	3.97***	3.97***	3.41***	4.47***
Experience	0.11***	0.17***	0.17***	0.2***	0.14***
Exper. Squared	-0.001***	-0.0005	-0.0005	-0.0005	-0.0002
Age	-0.11	-0.15	-0.15	-0.29	0.13
Age^2	0.0000414	-0.001	-0.001	-0.0007	-0.004*
High CD4	.	0.98***	0.95***	0.94***	0.83***
No Symptoms	.	0.56***	0.55***	0.61***	0.47***
Haart Available	.	1.02***	1.11***	.	.
Treatment: Mono (t+1)	.	.	-0.008	0.26	-0.24
Treatment: Combo (t+1)	.	.	-0.27**	0.2	-0.74**
Treatment: HAART (t+1)	.	.	-0.18	.	-0.64**
Treatment: Mono (t)	.	.	.	-0.23	-0.06
Treatment: Combo (t)	.	.	.	-0.41*	0.22
Treatment: HAART (t)	.	.	.	.	0.36
Observations	7954	7954	7954	3694	4260

Dichotomous employment choices (full time or not full time) at period $t+1$ conditional on treatment choices in periods t and $t+1$.

Table 4: Transition Matrix

		Time $t+1$	
		Not Full Time	Full Time
		Pre-HAART	
Time t	Not full time	0.86	0.14
	Full Time	0.10	0.90
		Post-HAART	
Time t	Not full time	0.91	0.09
	Full Time	0.06	0.94

Employment decisions (full time or not full time).

Table 5: Structural Parameter Estimates

	Latent Type (Preferences)		
	Type I	Type II	Σ
Latent Type:			
(Transitions and Outcomes)			
Type I	0.148	0.261	0.409
Type II	0.345	0.246	0.591
Σ	0.493	0.507	

Unconditional latent type probabilities.

Table 6: Structural Parameter Estimates

	Type I		Type II	
	Coefficient	Error	Coefficient	Error
No Ailments				
CRRA	0.81	0.10	0.80	0.02
Labor Disutility	-2.34	0.50	-2.59	0.48
Begin Treatment	13.56	3.95	.	.
Change Treatment	-6.17	0.37	.	.
End Treatment	-12.52	2.73	.	.
Ailments				
Constant	-42.27	3.80	-53.92	3.55
CRRA	0.77	0.09	0.74	0.01
Labor Disutility	-11.54	1.29	-5.22	1.16
Begin Treatment	-42.75	5.10	.	.
Change Treatment	4.33	0.53	.	.
End Treatment	30.08	2.76	.	.

Utility parameters.

Table 7: Structural Parameter Estimates

	Coefficient	Error
CD4 Increase ($\theta^{\Delta H}$)		
High CD4 at t	-0.40	0.09
High CD4 at t $\times$		
Mono-therapy	0.42	0.06
Combo-therapy	0.50	0.07
HAART [Type I]	0.77	0.10
HAART [Type II]	0.71	0.10
Low CD4 at t $\times$		
Mono-therapy	0.06	0.03
Combo-therapy	0.09	0.03
HAART [Type I]	2.21	0.41
HAART [Type II]	0.17	0.05
Time trend	0.02	0.00
Constant	-0.70	0.10
CD4 Count at $t+1$ (θ^{H})		
Type I $\times$		
High CD4 at t	6.16	0.66
Predicted CD4 increase (%) $\times$		
Low CD4	4.29	0.95
High CD4	2.94	0.08
Constant	-3.82	0.12
Type II $\times$		
High CD4 at t	5.26	0.67
Predicted CD4 increase (%)$\times$		
Low CD4	3.90	1.33
High CD4	0.66	0.15
Constant	-3.71	0.65

Drug effectiveness and CD4 count processes.

Table 8: Structural Parameter Estimates

	Coefficient	Error
Ailments (θ^F)		
High CD4 at t	0.94	0.05
High CD4 at t ×		
Mono-therapy	-0.19	0.02
Combo-therapy	-0.20	0.02
HAART [Type I]	0.01	0.00
HAART [Type II]	-0.19	0.03
Low CD4 at t ×		
Mono-therapy	0.27	0.04
Combo-therapy	0.24	0.04
HAART [Type I]	-1.04	0.22
HAART [Type II]	0.25	0.04
Time trend	-0.20	0.02
Constant	-0.45	0.06

Drug side effects process.

Table 9: Structural Parameter Estimates

	Coefficient	Error
Death (θ^B)		
High CD4	-6.16	1.16
Age × high CD4	0.07	0.02
Age	-0.01	0.01
Constant	-1.37	0.41

Survival process.

Table 10: Structural Parameter Estimates

	Coefficient	Error
Income (θ^I)		
High CD4 at t	1508.45	273.60
Experience	664.04	51.47
Experience-squared	-4.79	0.62
Age	-633.16	71.60
Time trend	165.88	32.39
Full-time employment	28273.11	3186.61
Full-time employment ×		
High CD4 at t	-871.93	371.21
Experience	248.74	74.78
Experience-squared	1.58	0.86
Age	-674.88	108.44
Time trend	81.70	40.98
Constant	26988.58	2154.88
σ_I^2	6673.68	53.06

Income process.

Table 11: Structural Parameter Estimates

	Coefficient	Error
Insurance (θ^N)		
Private Insurance:		
Full-time Employment	1.16	0.12
High CD4 at t	-0.50	0.14
Experience	0.16	0.03
Experience-squared	0.00	0.00
Age	-0.30	0.12
Age-squared	0.00	0.00
Time trend	0.07	0.02
Time trend (post-HAART)	0.02	0.03
Constant	6.95	2.49
Public Insurance:		
Full-time Employment	-0.88	0.14
High CD4 at t	-1.05	0.15
Experience	-0.03	0.03
Experience-squared	0.00	0.00
Age	0.17	0.14
Age-squared	0.00	0.00
Time trend	0.10	0.03
Time trend (post-HAART)	-0.05	0.04
Constant	-5.50	2.85

Insurance process.

Table 12: Structural Parameter Estimates

	Coefficient	Error
Out-of-Pockets Costs (θ^P)		
High CD4 at t	-131.52	28.58
High CD4 with Ailments	177.88	20.53
Low CD4 with Ailments	146.77	33.08
Income	44.50	5.56
Mono-therapy	502.94	105.26
Combo-therapy	140.47	117.70
HAART	-73.41	118.10
Private Insurance	-75.37	56.69
Public Insurance	84.66	79.05
Private Insurance ×		
Mono-therapy	-227.55	108.22
Combo-therapy	155.91	120.84
HAART	407.91	119.39
Public Insurance ×		
Mono-therapy	-508.47	129.12
Combo-therapy	-194.80	143.20
HAART	63.81	134.89
Time trend	14.49	1.46
Constant	-524.10	66.27
σ_P^2	720.49	6.60

Out-of-pocket treatment cost process.

Table 13: Model Fit

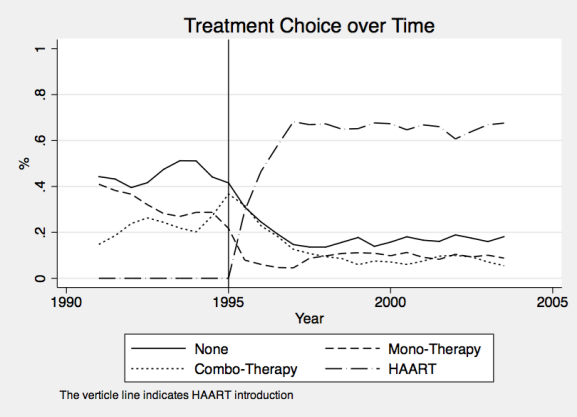
	Labor Employed		Drug Choice							
			None		Mono		Combo		HAART	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
Full Sample	0.66	0.66	0.30	0.30	0.20	0.20	0.17	0.17	0.33	0.33
Low CD4	0.46	0.46	0.20	0.20	0.29	0.29	0.25	0.25	0.26	0.26
High CD4	0.72	0.72	0.34	0.34	0.17	0.17	0.14	0.14	0.35	0.35
Exp > 10	0.70	0.68	0.27	0.28	0.19	0.19	0.17	0.17	0.37	0.36
Exp ≤ 10	0.63	0.64	0.33	0.33	0.21	0.20	0.17	0.17	0.29	0.30
Age > 45	0.63	0.62	0.23	0.24	0.17	0.17	0.15	0.15	0.45	0.44
Age ≤ 45	0.68	0.69	0.36	0.36	0.22	0.22	0.19	0.19	0.23	0.23
Pre-HAART	0.68	0.67	0.45	0.45	0.32	0.32	0.23	0.23	.	.
$\times$ Low CD4	0.50	0.50	0.25	0.26	0.40	0.40	0.35	0.34	.	.
$\times$ High CD4	0.76	0.74	0.54	0.54	0.29	0.27	0.18	0.18	.	.
Post-HAART	0.64	0.65	0.18	0.18	0.09	0.10	0.12	0.12	0.62	0.61
$\times$ Low CD4	0.41	0.48	0.12	0.09	0.11	0.13	0.11	0.12	0.66	0.67
$\times$ High CD4	0.69	0.70	0.19	0.19	0.08	0.09	0.12	0.12	0.61	0.60

Given different sets of state variables, choice probabilities are computed using model parameters and recorded in the columns labeled "Model". For comparison, analogous sample moments are recorded in the columns labeled "Data".

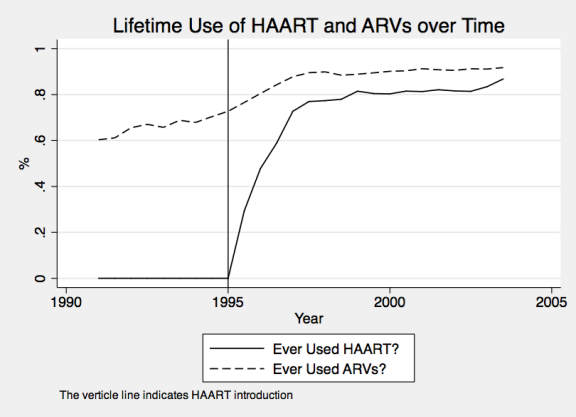
Table 14: Posterior Type Probabilities

		Latent Type			
		Pref. Type		Health Type	
		I	II	I	II
Full Sample		0.48	0.52	0.41	0.59
College	No	0.54	0.46	0.40	0.60
	Yes	0.45	0.55	0.42	0.58
Occupation:					
Professional specialty		0.44	0.56	0.42	0.58
Admin. or clerical		0.50	0.50	0.40	0.60
Waitor		0.59	0.41	0.38	0.62
Craft		0.26	0.74	0.61	0.39
Mining		0.55	0.45	0.40	0.60
Transportation		0.60	0.40	0.45	0.55

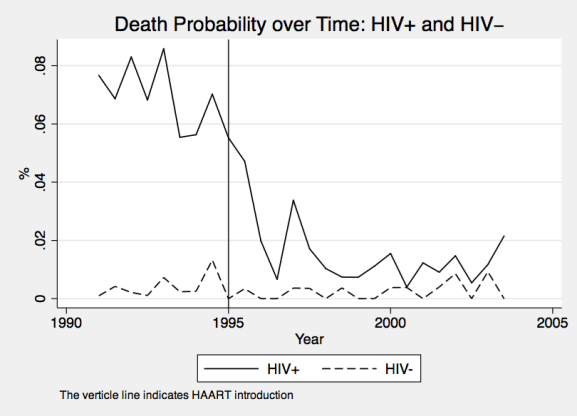
For each individual and for each latent type, a ratio is computed where the numerator is the likelihood contribution using estimated parameters for the given type and the denominator is the full likelihood contribution. The result is a number between 0 and 1 that provides a posterior probability that the individual belongs to each latent type. These ratios are averaged across groups of individuals distinguished by explanatory variables, like education, that are not included in the structural model. For example, the unconditional preference Type I probability is 0.48. The posterior indicates that college graduates are preference Type I with probability 0.45. Non-college graduates are preference Type I with probability 0.54.



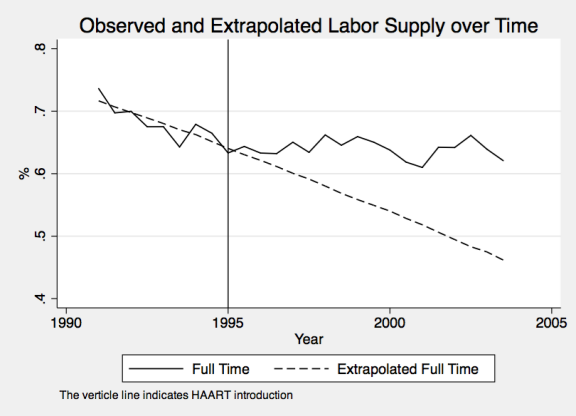
(a)



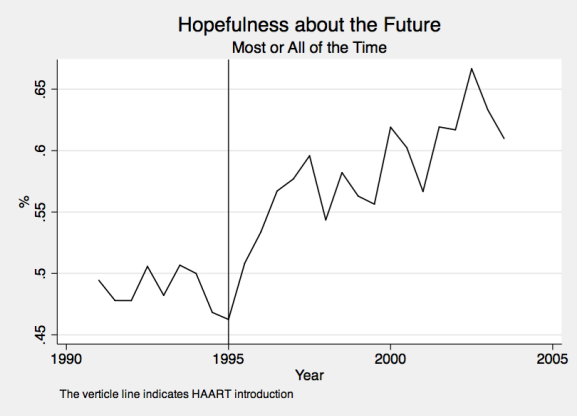
(b)



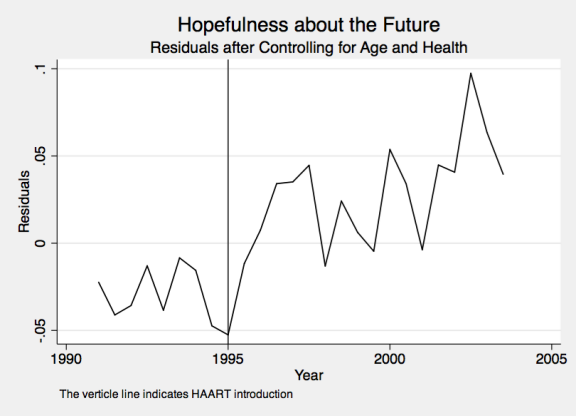
(c)



(d)



(e)



(f)

Figure 1: SUMMARY TRENDS OVER TIME. Panel 1(a): Average treatment choice. Panel 1(b): Average lifetime HAART and Antiretroviral (ARV) use. Panel 1(c): Probability of non-survival until period $t+1$ given survival until t (HIV− and HIV+). Panel 1(d): Average full-time employment: observed and extrapolated from the pre-HAART trend. Panels 1(e)-1(f): proportion of individuals reporting hopefulness about the future most or all of the time in the week prior to MACS interview (actual and residuals detrended for age and CD4-count, respectively).

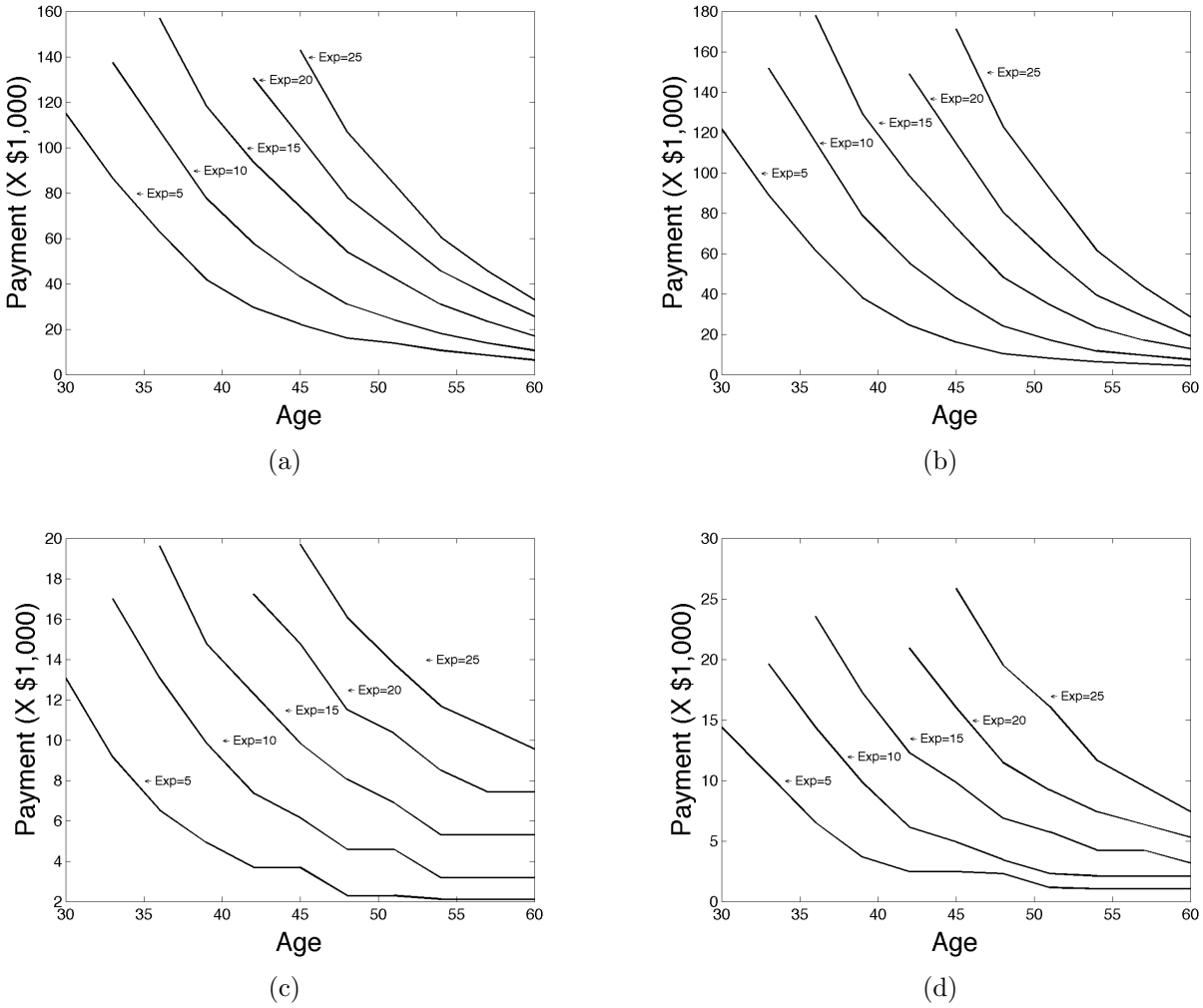


Figure 2: Heterogeneity in the value of HAART: Value of HAART for Preference Type I and Health Type I (Panel 2(a)), Preference Type II and Health Type I (Panel 2(b)), Preference Type I and Health Type II (Panel 2(c)) and Preference Type II and Health Type II (Panel 2(d)).

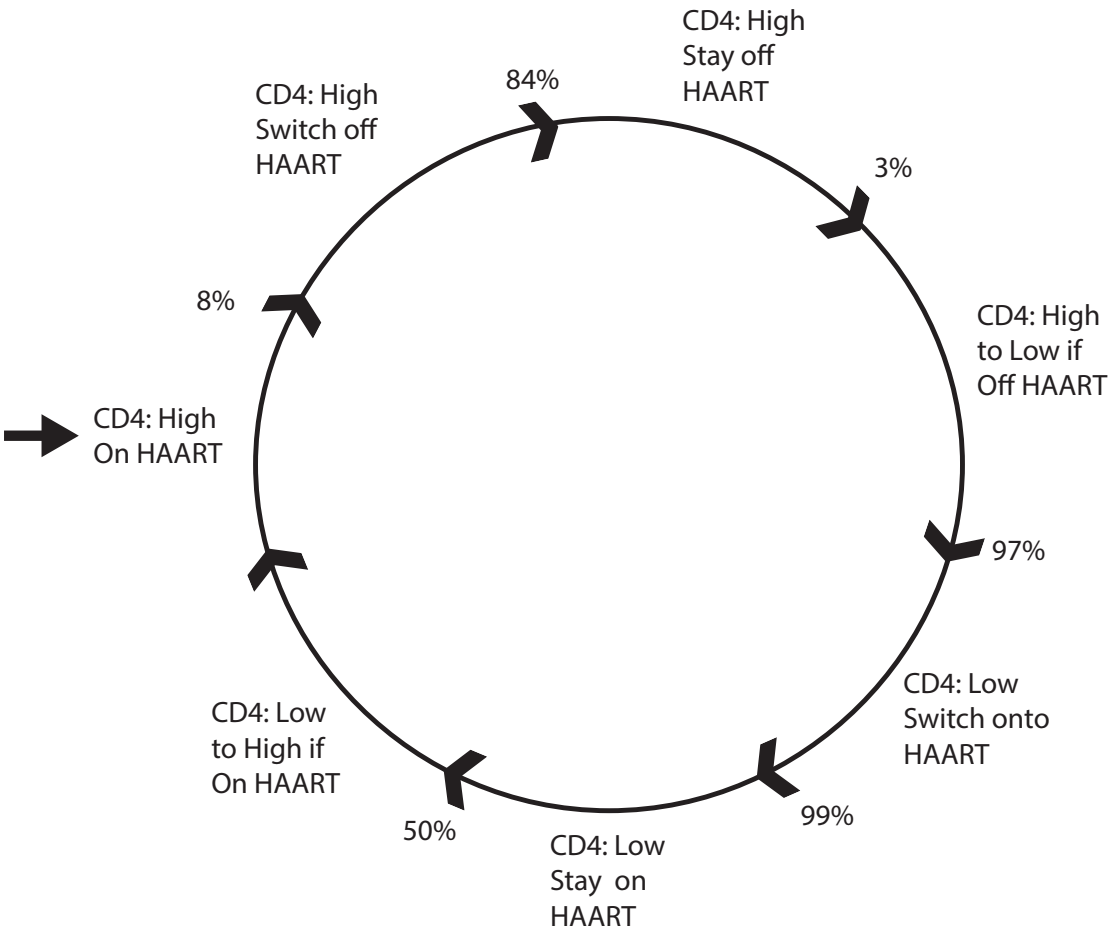
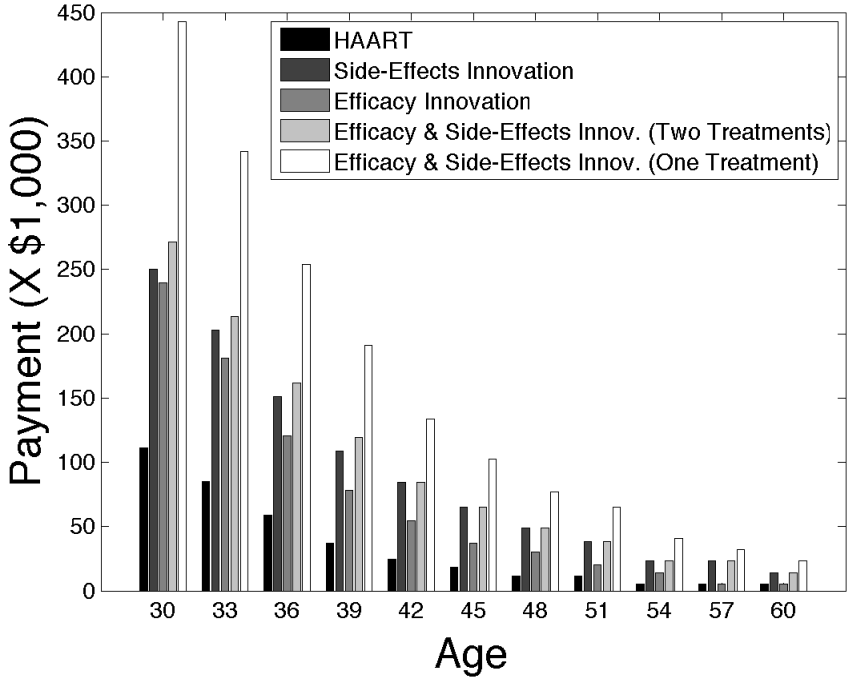
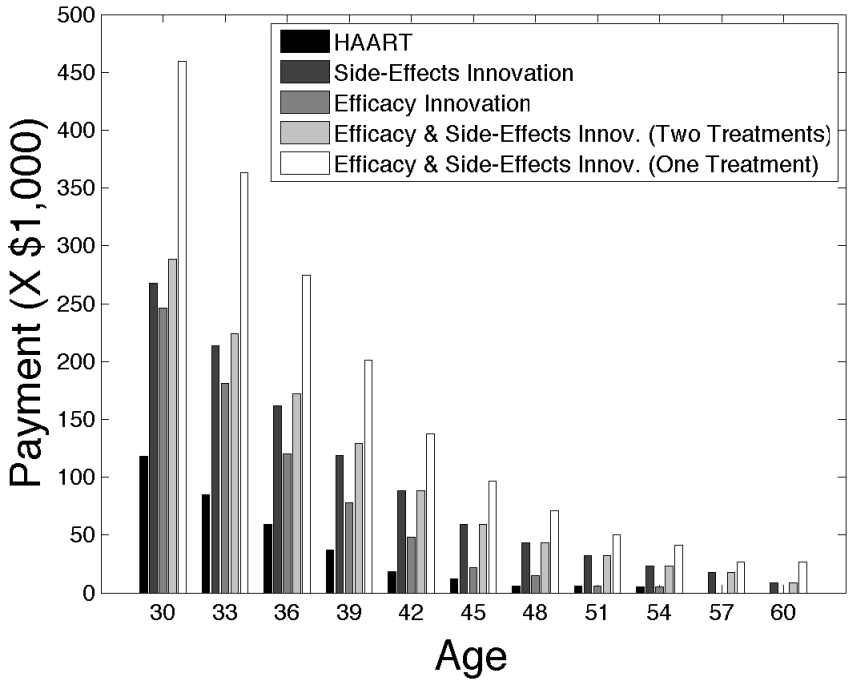


Figure 3: Optimal Treatment Cycling: Probabilities along the cycle are simulated using model parameters. Beginning with the rightward pointing arrow at the left, Preference Type I and Health Type I agents with a high CD4 count cycle off of HAART with 8% probability in each period. Once off of HAART, they remain off of HAART with probability 84% as long as their CD4 count is high. With 3% probability in each period, their health declines at which point, with 97% probability they go onto HAART, remaining there, given low CD4 count, with nearly 100% probability. With 50% probability in each period, they recuperate their health. Other latent types exhibit similar cycling behavior, with changes driven by HAART effectiveness. As HAART is not as effective for health Type II agents, given a low CD4 count, they switch onto HAART less quickly and are more likely to switch off of HAART even before attaining a high CD4 count.

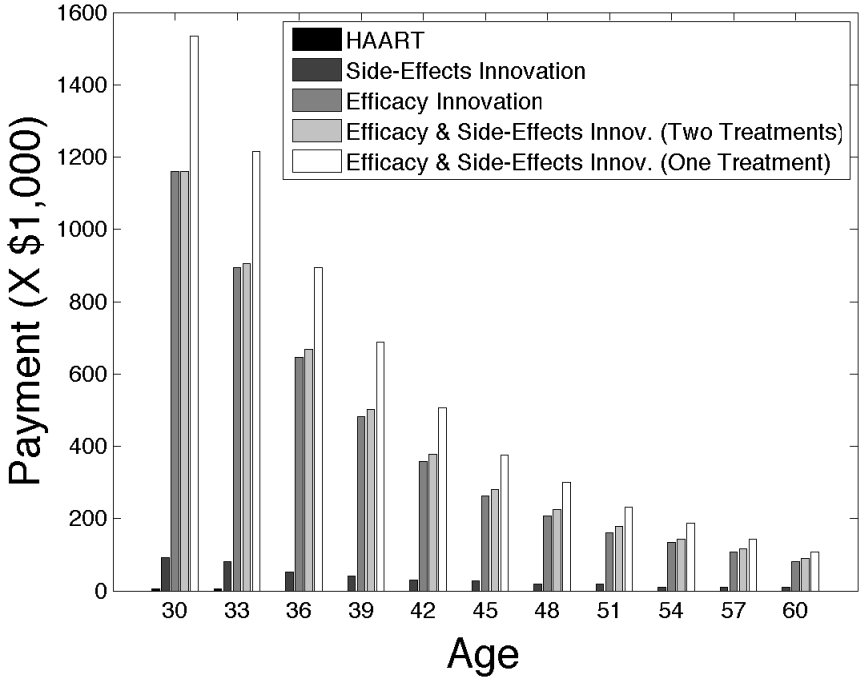


(a)

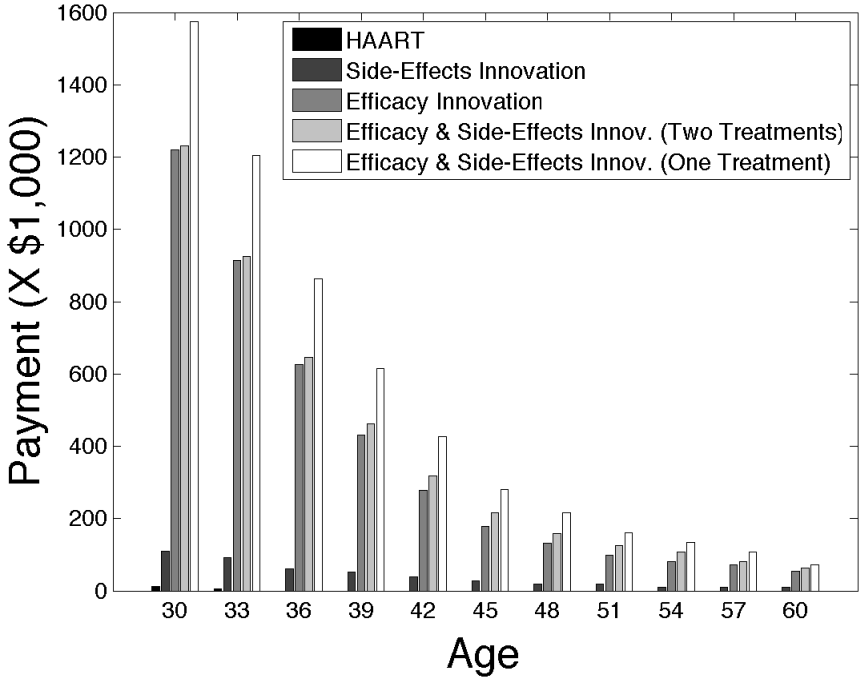


(b)

Figure 4: Heterogeneity in the value of Pharmaceutical Innovation: Value of Counterfactual Innovations for Preference Type I and Health Type I (Panel 4(a)) and Preference Type II and Health Type I (Panel 4(b)).

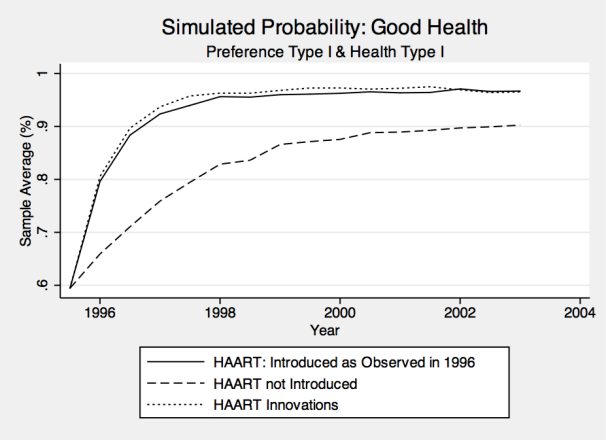


(a)

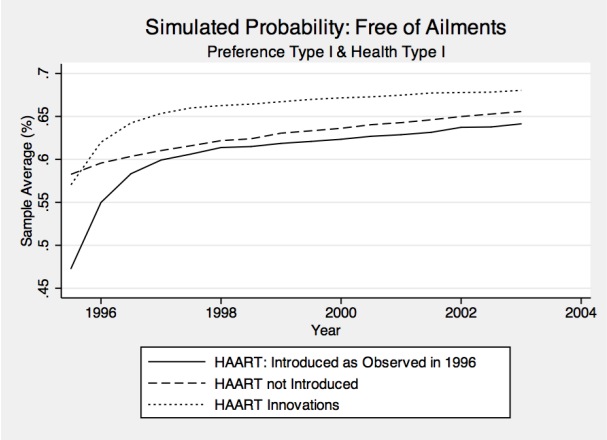


(b)

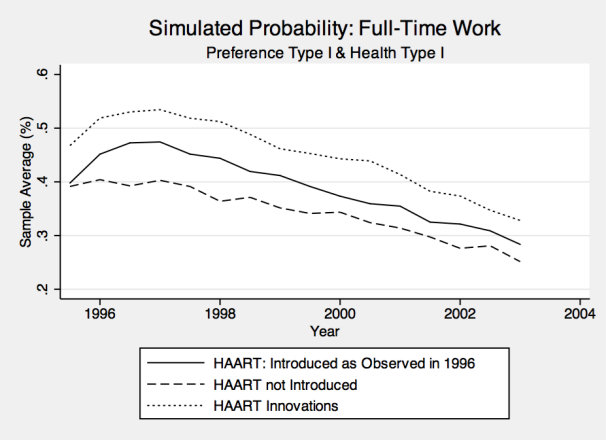
Figure 5: Heterogeneity in the value of Pharmaceutical Innovation: Value of Counterfactual Innovations for Preference Type I and Health Type II (Panel 5(a)) and Preference Type I and Health Type I (Panel 5(b)).



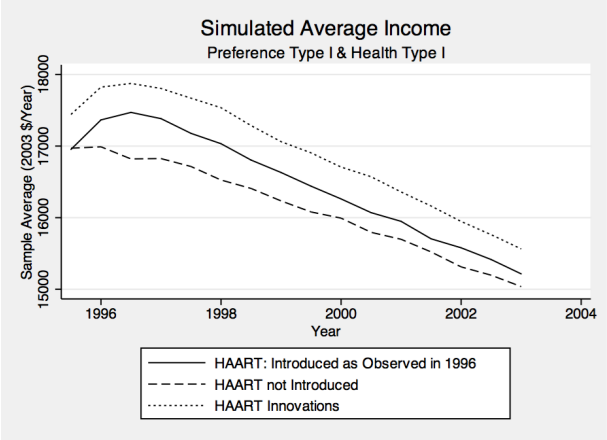
(a)



(b)

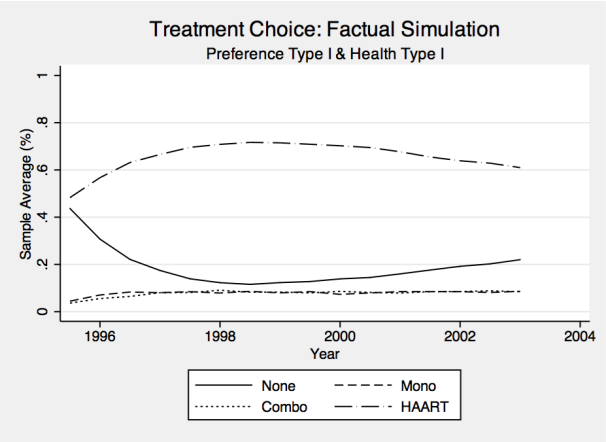


(c)

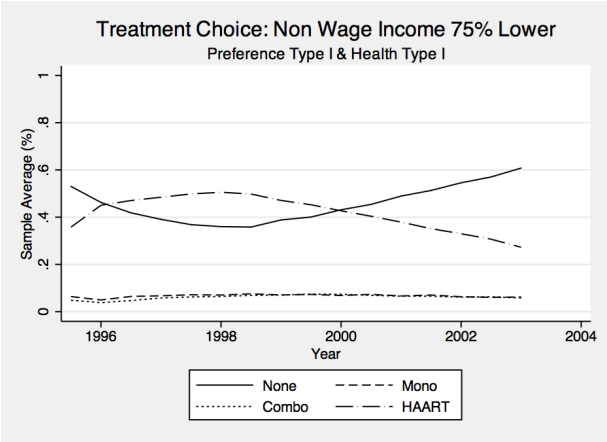


(d)

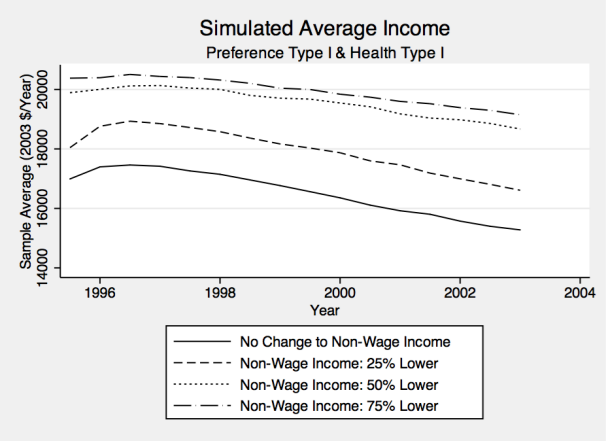
Figure 6: Counterfactual Policy Simulations - Treatment Innovations: Three treatment environments are explored (i) HAART is introduced as observed (ii) HAART is not introduced and (iii) instead of HAART two treatments are introduced, one with high effectiveness with HAART-level side effects, the other with HAART effectiveness and no side effects. For each simulated environment and for preference Type I and health Type I, Panel 6(a) shows the average probability of high CD4 count over time. Panel 6(b): Average probability of not suffering from physical ailments. Panel 6(c): Average probability of working full time. Panel 6(d): Average net income in $2003/year.



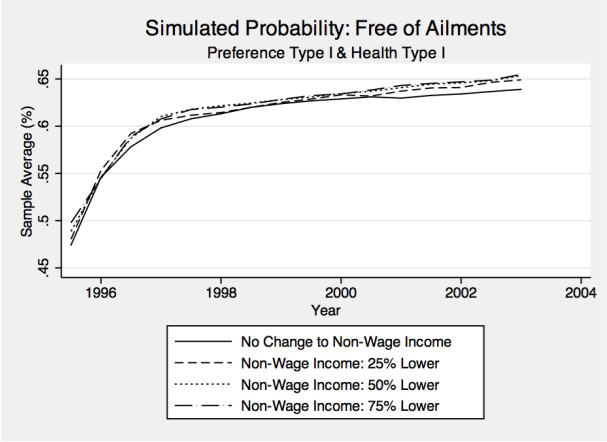
(a)



(b)

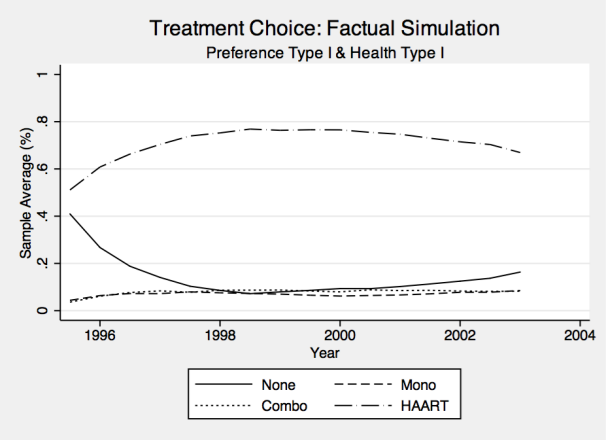


(c)

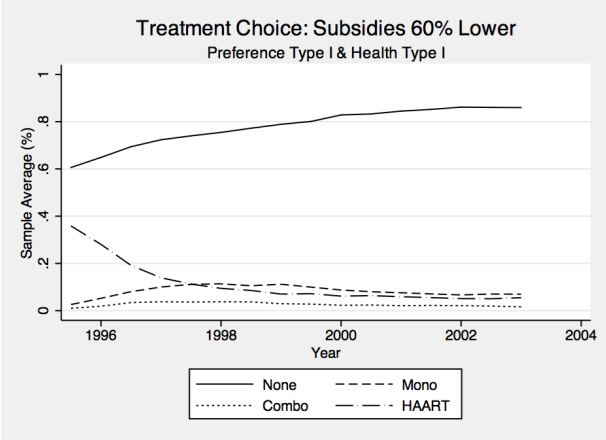


(d)

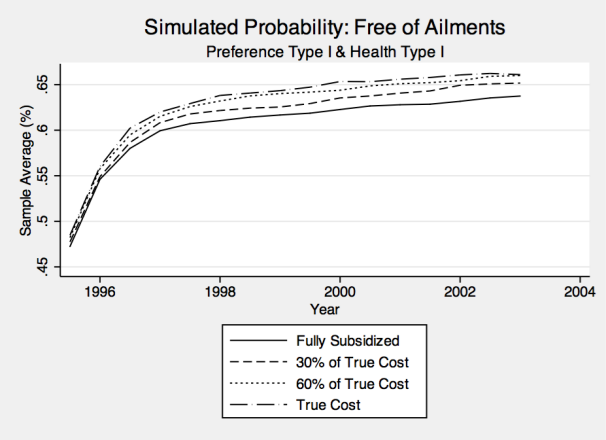
Figure 7: Counterfactual Policy Simulations - A Decline in Non-Wage Income: Non-wage income is simulated to decline 25%, 50% and 75%. For preference Type I and health Type I, Panel 7(a) depicts treatment choices over time for no decline in non-wage income and Panel 7(b) depicts treatment choices under a 75% drop in non-wage income. For each simulated environment, Panel 7(c) shows average net income in \$2003/year and Panel 7(d) shows the probability of not suffering from physical ailments.



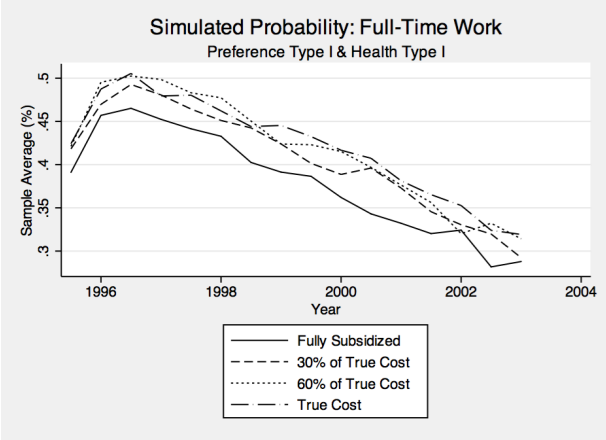
(a)



(b)



(c)



(d)

Figure 8: Counterfactual Policy Simulations - An Increase in Treatment Costs: Out-of-pocket treatments costs are simulated to increase to 20%, 40% and 60% of actual treatment costs. For preference Type I and health Type I, Panel 8(a) depicts treatment choices over time for no decline in non-wage income. Panel 8(b): Treatment choices under a 60% drop in non-wage income. For each simulated environment, Panel 8(c) depicts the probability of not suffering from physical ailments. Panel 8(d): simulated labor choice.