Quantitative Economics 7 (2016), 671–725 1759-7331/20160671

Why medical innovation is valuable: Health, human capital, and the labor market

Nicholas W. Papageorge
Department of Economics, Johns Hopkins University

I develop a dynamic 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, they choose labor supply and medicine in light of potential side effects in an effort to jointly manage two forms of human capital: their health and their work experience. I use the framework to examine the treatment and employment decisions of human immunodeficiency virus (HIV) positive men before and after a medical breakthrough known as highly active anti-retroviral treatment. A novelty of this application is my use of data containing both objective health measures along with reports of physical ailments. This allows me to model each HIV drug along two dimensions of quality—effectiveness and side effects. Using the framework, I am able to identify the impact of side effects on demand and show that counterfactual innovations that reduce side effects can be very valuable. I also show that when no treatment dominates along both dimensions of drug quality, patients exhibit health-state-dependent cyclicality in their medical treatment decisions, 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.

Keywords. Innovation, health, human capital, labor, structural models, HIV/AIDS.

JEL classification. I10, J24, O31.

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 dominant in the literature for assessing the value of improvements

Nicholas W. Papageorge: papageorge@jhu.edu

This paper was originally a chapter in my doctoral thesis completed at Washington University in St. Louis. Special thanks go 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 Eraslan, Amy Finkelstein, Stephanie Heger, Glenn MacDonald, Robert Moffitt, Harry Paarsch, Carl Sanders, and Seth Sanders along with seminar participants at Duke, NYU, Penn, UNC, and the Johns Hopkins University Bloomberg School of Public Health. Finally, I thank the co-editor along with two anonymous referees. All errors are my own.

DOI: 10.3982/QE459

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 trade-off 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, Drummond, Fryback, Gold, and Revicki (2009)).

The framework centers around estimation of a dynamic model where forward-looking agents make repeated medical treatment and labor supply decisions until retirement. The model captures the following two key trade-offs. 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 forgo treatment to avoid painful or uncomfortable side effects, thus allowing their health capital to depreciate. The second trade-off is the one faced by agents choosing whether or not to work. By working, agents earn income and accumulate human capital, but also forgo valuable leisure time. The structure of the model captures various ways in which these two trade-offs 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 capturing the trade-offs 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 a 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 factors that influence the quality of life, like consumption and leisure. Exploiting the characteristics 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]

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 generally meaningful 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, 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 reason I can model a second dimension of drug quality in such detail is that I not only observe individual treatment choices and objective health, but also have

[1] In the model, out-of-pocket costs are also included.

[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 drugs or molecules (see, e.g., Arcidiacono, Ellickson, Landry, and Ridley (2013)). This approach 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.

data on the same individuals' reports of physical ailments, including fatigue, diarrhea, headaches, fever, and drenching sweats. For individuals in poor health, these ailments are symptoms of disease. For individuals in better health, but who are on medication, these ailments are side effects. Observing and incorporating physical ailments into a model of treatment choice stands in contrast to many studies where models of pharmaceutical demand are matched to data where side effects are not observed (Crawford and Shum (2005), Chan and Hamilton (2006), Chintagunta, Jiang, and Jin (2009), Ching (2010), Chan, Narasimhan, and Xie (2013)).[4] Therefore, observing and incorporating reports of physical ailments leads to a rich model of demand for two distinct drug characteristics, which rationalizes observed and potentially puzzling variation in treatment choices—not only across individuals, but also for the same individuals across time.[5]

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: being infected with human immunodeficiency virus (HIV; henceforth 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.[6] Further, according to the Centers for Disease Control and Prevention, nearly 50% of adults in the United States suffer from a chronic condition, about one-quarter of whom experience significant limitations in daily activities like working.[7]

Several features of HIV and the acquired immune deficiency syndrome (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 trade-off 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.

[4]Several of these studies model side effects as a residual used to explain why patients do not always choose the health-maximizing treatment. An exception is Fernandez (2013), who explains clinical trial participation in part through observed side effects. I should note that my study omits learning, which is a focus of the aforementioned studies.

[5]By rationalizing strong variation in demand among similarly healthy individuals, the framework offers a compelling explanation for why multiple drugs of similar average effectiveness (often known as *me-too* drugs) can coexist within a single market.

[6]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 forgo medication for limited periods of time, though this can accelerate disease progression (Kerbrat et al. (2011)).

[7]For this point, see http://www.cdc.gov/chronicdisease/overview/. In principle, the structure of the model means it could be applied to illnesses that are not chronic, but where 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 trade-off between working and getting better more quickly. More generally, agents do not need to be ill at all to face a trade-off 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.

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. Helpful in identifying model preference parameters is that I observe treatment and employment choices for the same individuals both before and after an unanticipated medical breakthrough known as highly active antiretroviral treatment (HAART).[8] 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.[9]

Turning to results, I find that from the perspective of an HIV+ patient, a dynamically optimal treatment plan is not consistent with full compliance or 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)).[10] 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*.[11]

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 on 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

[8]There is no vaccine or cure for HIV or AIDS, but HAART is the current standard treatment.

[9]Duggan and Evans (2008) also use HAART introduction to study the effects 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 (2016).

[10]Treatment cycling is dynamically optimal from the perspective of the patient, but may be suboptimal from the perspective of society, especially in the context of illnesses where treatment choices have externalities, which is the case with HIV since medication use can lower infectiousness, which benefits the HIV negative (HIV−) sexual partners of HIV+ men. I return to this point in Section 5, where I present results, and again in the Conclusion, where I discuss the policy relevance of the main findings.

[11]Even if treatment cycling deteriorates health, it is not incongruent with an optimal treatment plan since 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 would further underscore the dynamic optimality of treatment cycling (Stebbing and Dalgleish (2009)).

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 than it was 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 parameter estimates, treatment cycling, and the value of pharmaceutical innovation; Section 6 discusses policy experiments highlighting how drug innovation interacts with employment; and Section 7 concludes. Replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/459/code_and_data.zip.

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+ and HIV− (the latter referring to men who are not infected with HIV) homosexual and bisexual men conducted at four sites: Baltimore, Chicago, Pittsburgh, and Los Angeles.[12] Visits occur semiannually. Accordingly, for the remainder of the analysis, one time period is set to be 6 months long. At each visit, information is collected on medical treatment choices, employment decisions, labor market outcomes, and health status, including CD4 (cluster of differentiation 4) count via blood sample, which provides an objective measure of immune system health. Subjects also report physical ailments, which include fatigue, diarrhea, headache, 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.

In constructing the subsample used in analysis, I use data on HIV+ men between 1990 and 2003. Beginning in 1990, drugs with some effectiveness at combating HIV emerge and MACS treatment data collection becomes more consistent across time. To

[12]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, 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). The MACS website is located at http://www.statepi.jhsph.edu/macs/macs.html.

limit the size of the state space, I drop 9 observations where individuals are over 65 years old and 60 observations where individuals are under 30. I also drop observations with missing data. The resulting analysis sample is an unbalanced panel of 8,300 observations: 743 HIV+ subjects over 26 visits and between the ages of 30 and 65.[13]

Subjects report all medications they have used since their previous interview. As there are dozens of medications (known as anti-retrovirals (ARVs)) used to fight HIV infection, I follow previous research (see, for example, Detels, Tarwater, Phair, Margolick, Riddler, and Muñoz (2001)) in creating four broad and mutually exclusive treatment categories: no treatment, mono-therapy, combo-therapy, and HAART.[14] 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.[15] 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*

Summary statistics are presented in Table 1 for the full analysis sample and then separately by HAART presence (pre-HAART versus post-HAART), by health status (high or low CD4 count), and by employment status (full time or not). In the full sample, average age is 33 at the start of the AIDS epidemic (1984) and 44 over the sample period. Healthy HIV− individuals exhibit an average CD4 count of 1,000 units per cubic millimeter of blood. The sample average of HIV+ men is about 450, though this number obfuscates important variation: the pre- and post-HAART averages are 393 and 500, 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 for "high CD4 count" that takes the value 0 when patient CD4 count is low enough to indicate AIDS (<250) and 1 otherwise. 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 being "free of ailments," which takes the value 0 if individuals report persistently experiencing one of the ailments fatigue, diarrhea, headaches, fever, or drenching sweats, and 1 otherwise; 59% of subjects report that they are free of such

[13]The full MACS data set, including pre-1990 observations and uninfected (HIV−) individuals, contains information on 5,622 subjects at 41 possible visits for a total of 98,886 subject-visits. For the sample period (1990–2003), the data set contains information on 769 HIV+ individuals, corresponding to 9,837 subject-visits.

[14]An individual 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).

[15]Employment histories are constructed using all available data, including observations when HIV+ individuals were observed HIV−, if applicable, along with pre-1990 observations and observations with up to two missed subsequent visits, in which case I assume that individuals engage in the same employment status as in the last observed period.

TABLE 1. Summary statistics.

		HAART Presence		Health Status		Employment Status	
	Full Sample (1)	Pre-HAART (2)	Post-HAART (3)	High CD4 (4)	Low CD4 (5)	Full Time (t) (6)	Not Full Time (t) (7)
Age	44.77	41.92	47.38	45.01	44.10	44.32	45.67
Age in 1984	32.71	33.16	32.30	32.51	33.29	32.40	33.32
(Std. dev.)	(6.87)	(6.31)	(6.30)	(6.87)	(6.83)	(6.67)	(7.17)
CD4 count	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 prob.	0.04	0.07	0.02	<0.01	0.14	0.02	0.09
Net income	37,407.80	38,354.40	36,539.06	38,863.08	33,377.69	42,884.86	26,579.41
Treat. costs	542.39	426.02	648.79	505.02	645.10	478.24	668.60
Experience	12.56	11.77	13.28	12.68	12.21	12.89	11.91
Insurance							
None	0.05	0.07	0.03	0.05	0.05	0.04	0.08
Private	0.81	0.84	0.80	0.84	0.73	0.92	0.59
Public	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
Mono	0.20	0.32	0.09	0.17	0.29	0.20	0.19
Combo	0.17	0.23	0.12	0.14	0.25	0.17	0.17
HAART	0.33	–	0.62	0.35	0.26	0.30	0.39
Work ($t+1$)							
Full time	0.66	0.68	0.64	0.72	0.46	0.92	0.11
Not full	0.34	0.32	0.36	0.28	0.54	0.08	0.89
N	8,300	3,972	4,328	6,098	2,202	5,512	2,788
N ($t+1$)	7,954	3,694	4,260	6,057	1,897	5,403	2,551

Note: The pre-HAART era contains observations from 1991 until mid 1995 (9 half-year periods). The post-HAART era contains observations from 1996 until 2003 (17 half-year periods). Each entry represents the mean over individuals and time periods for the period or group in question. Entries for high CD4, no ailments, insurance, treatment, and work are proportions. Age and experience are in years. Treatment costs and net income are in 2003 dollars per year. Death probability is per half-year period. The age range for the whole sample and for all subgroups is 30–64.5 except for full-time work, where the maximum age is 64.

ailments. Finally, death probability is about 4% per half-year 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 CD4 count, AIDS level CD4, and survival). Further, better health correlates with fewer ailments (see columns 4 and 5 of Table 1, which compare high and low CD4 count individuals). Despite improved average health after HAART is introduced, the same proportion of individuals reports suffering physical ailments (59%) in the pre- and post-HAART eras. This parity may appear puzzling given changes in the market for HIV drugs and of health over time. It arises since HAART both mitigates and exacerbates physical ail-

TABLE 2. Logistic regressions: physical ailments.

	Ailments Reported for Period t					
	[Fatigue] (1)	[Diarrhea] (2)	[Headache] (3)	[Fever] (4)	[Sweating] (5)	[One or More] (6)
High CD4	−1.41***	−1.11***	−0.73***	−1.43***	−1.13***	−1.41***
× Mono	0.57***	0.74***	0.25*	0.35**	0.48***	0.58***
× Combo	0.57***	0.49***	0.31**	0.5***	0.42***	0.57***
× HAART	0.46***	0.95***	0.17	0.37**	0.15	0.66***
Low CD4	–	–	–	–	–	–
× Mono	−0.08	0.2	−0.03	−0.28	−0.16	−0.07
× Combo	0.25*	0.1	0.04	0.02	−0.05	0.1
× HAART	−0.21	0.45***	−0.08	−0.38*	−0.24	−0.07
Time trend	0.003	−0.006	−0.03***	−0.02***	−0.002	−0.006
Constant	−0.36**	−0.98***	−1.26***	−1.11***	−1.36***	0.53***
Observations	7,954	7,954	7,954	7,954	7,954	7,954

Note: Coefficients from logistic regressions of binary variables indicating each physical ailment separately on treatment choice interacted with current period health. The final column regresses a binary variable for having one of the five ailments.

ments. It combats symptom-causing illness, but simultaneously causes side effects and these two effects cancel each other out.

In support of the view that observed ailments can be symptoms of illness or side effects of treatment, I next study each ailment separately. In particular, I regress a binary variable indicating a report of each individual ailment onto treatment choices interacted with health and a time trend. Estimated coefficients are presented in Table 2.[16] First, each ailment probability is lower for individuals with higher CD4 counts, which suggests that ailments are appropriately viewed as symptoms. Second, HAART has mixed effects on ailments for individuals in good health relative to other medications. For example, in comparison to mono- and combo-therapy, diarrhea is strongly predicted by HAART use when healthy, whereas headaches are not. These differences can be used as an informal way to test whether ailments remaining constant over time truly reflects that symptoms are replaced by side effects after HAART is introduced. Indeed, in the post-HAART era there is a 30% drop in the likelihood of headaches being reported, but a 16% increase that diarrhea is reported. In other words, though the likelihood of no ailments being reported remains nearly constant in the pre- versus the post-HAART eras, underlying this parity are two countervailing dynamics. In the post-HAART era, the types of ailments reported shift away from those associated with illness (symptoms) toward those associated with strong medications (side effects). The model developed in the following section is designed to capture how these countervailing dynamics influence treatment decisions.

[16]I use logistic regressions for this exercise. The table also includes coefficients from regressing a variable indicating if at least one physical ailment is reported onto the same explanatory variables.

Turning next to the interactions between health and the labor market, Table 1 shows that average reported income is about $37,400 (in 2003 dollars per year).[17] Non-wage income averages about $26,500, which is lower than the average amount reported by workers (about $43,000), but may seem high at first glance. It reflects the fact that HIV/AIDS is considered a disability, which opens up the possibility of Social Security disability payments and private pensions, which would presumably increase with pre-disability wage income.[18] Despite high non-wage income, it is clear that individuals 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 subjects 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. 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.[19]

2.2 *Treatment choices and employment decisions*

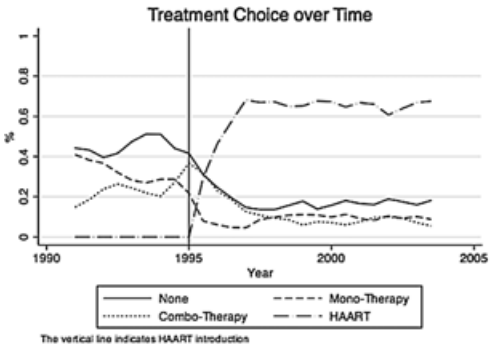
Next, I study factors affecting decisions. Again referring to summary statistics in Table 1, notice that a plurality of sample subjects (45%) eschewed all medical treatments in the pre-HAART era, but a majority of subjects (62%) use HAART after it is introduced. This shift is depicted in Figure 1(a), which plots treatment choices over the sample period. Not only do subjects substitute HAART for other treatments, but those who refrained from using earlier, less effective treatments appear to switch onto HAART after 1996.

A puzzling feature that emerges in Figure 1(a) is that not everyone who is HIV+ in any given period uses HAART. One possibility is that after HAART is introduced, individuals 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 individuals who do not use HAART (about 38% after 1997). A second possibility that is consistent with observed usage patterns over time is that a subgroup of individuals never goes onto HAART. This is certainly true of individuals who leave the sample prior to HAART introduction (mostly due to death). However, consider Figure 1(b), which plots whether

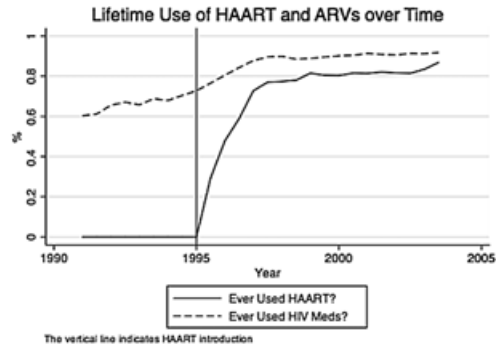
[17] Income is a categorical variable reported in increments of $10,000, where the highest income category is $50,000 or more. To convert per-period income into dollars, I take the midpoint of each category and then divide it by 2. The highest income category is set to $27,500 per half-year, though reduced-form 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.

[18] In support of this possibility, auxiliary regressions show that, among HIV+ men who were also observed when HIV−, non-wage income is positively correlated to their wage income in periods before they were HIV−.

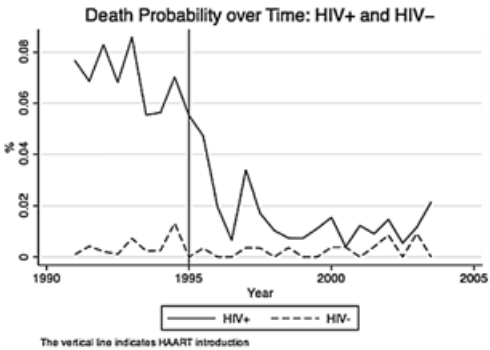
[19] Insurance options are no insurance, private only, or some mixture of public and private, which includes public only. The third option accounts for how many employed individuals may have private insurance from work, but also have access to public insurance that pays for HIV medications.



(a)

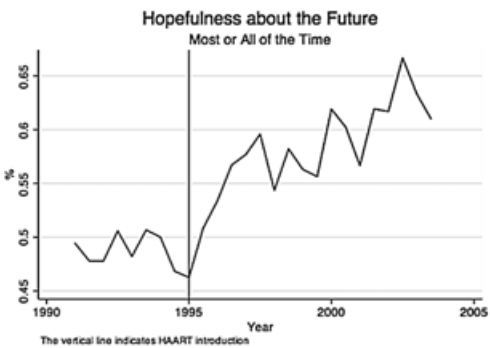


(b)

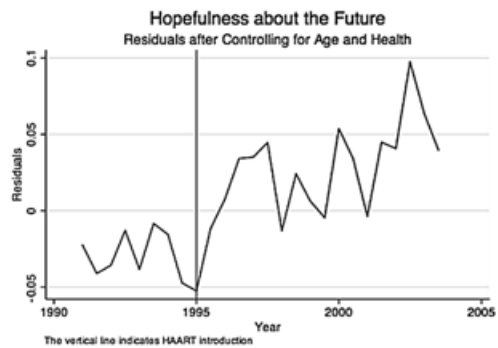


(c)

(d)



(e)



(f)

FIGURE 1. Summary trends over time. (a) Average treatment choice. (b) Average lifetime HAART and HIV medications use. (c) Probability of nonsurvival until period $t+1$ given survival until t (HIV− and HIV+). (d) Average full-time employment, observed and extrapolated from the pre-HAART trend. (e) and (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).

TABLE 3. Transition matrix: treatments.

		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

Note: Period-by-period HIV treatment choice transitions from periods t to $t+1$.

an individual observed in any given period has previously used HAART or some form of HIV medication. By the end of the sample period, nearly 90% of observed individuals have used HAART at least once in the past. Nonetheless, there is no period in which 90% of observed individuals are on HAART at the same time. This discrepancy is crucial as it means that being on HAART is not an absorbing state. In other words, some individuals used HAART at least once in the past, but have switched off.

To shed light on the dynamics shown in Figure 1(a) and (b), Table 3 presents a transition matrix for individual treatment choices. First, treatment choices are highly persistent, though more strongly so for individuals in better health. Second, a small proportion of both healthy and unhealthy individuals go off of HAART in any given period. About 13% of AIDS-level patients and about 6% of healthier individuals go off of HAART. Among individuals who are off of treatment, healthier individuals are more likely than unhealthy individuals to remain off. About 87% of healthier individuals remain off. In comparison, about half of sick individuals start taking some type of medication and 35% go 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.

Table 3 shows that some individuals go off of HAART. Doing so lowers the likelihood of recovering (if unhealthy) and raises the likelihood of transitioning to AIDS-level CD4 (if healthy). Table 3 also shows that sick individuals are highly likely to go onto HAART, where they face a higher probability of getting healthy. Taken together, these dynam-

TABLE 4. Evidence of cyclicality in HIV medication usage.

	Number of Individuals
Total in sample	743
Never observed on medication	95
Never observed on HAART	428
Observed on medication at least once	648
Switch off medication at least once	196
Switch back onto medication at least once	122
Observed on HAART at least once	315
Switch off HAART at least once	131
Switch back onto HAART at least once	86

Note: Of the 743 individuals in the sample, 648 are observed on HIV medications, including mono-therapy, combo-therapy, or HAART at least once. Of these, 196 (30% of the 648) are observed going off of all medications at least once and 122 (19%) are observed going back onto an HIV medication at least once. Of the 743 individuals, 315 individuals are observed on HAART. Of these, 131 (42% of the 315) are observed going off of HAART and onto a less effective medication (or no medication at all) and 86 (27%) are observed going back onto HAART.

ics suggest the possibility of health-state-dependent cyclicality in treatment decisions, whereby some healthier agents go off of HAART, tending to remain off until their health declines, at which point they go back on. To explore the possibility of health-state-dependent cyclicality in treatment choices more explicitly, I next ask how many individuals are observed switching among medications. Results are presented in Table 4. According to the table, of the 743 individuals in the sample, 648 use an HIV medication at least once. Of these, 196 (30% of 648) go off of medication at least once and 122 (19%) go off and then switch back onto medication at least once. These patterns are not driven solely by cycling among low-effectiveness medications. Focusing on HAART, 315 of the individuals in the sample are observed using HAART at least once.[20] Of these individuals, 131 (42% of 315) are observed going off of HAART at least once and 86 (27%) are observed going off and then back on at least once.[21] Taken together, these patterns suggest that many individuals who use HAART eventually go off of HAART and, if so, face a higher probability of getting sick. If sick, they are likely to go back onto HAART. In other words, the data exhibit health-state-dependent cyclicality in HAART usage. A key contribution of the framework developed in this paper is to rationalize this cyclicality.

Given the post-HAART drop in death rates (see Figure 1(c)), the next question is why an HIV+ individual 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

[20]Of the 428 individuals who are never observed using HAART, 341 are not observed after HAART is introduced. Of these, 278 die prior to HAART introduction and 63 attrit. Of the 87 individuals who are observed after HAART introduction, but never observed using HAART, 64 attrit and 23 die prior to ever using HAART.

[21]Ten percent of individuals who ever take HAART are observed going off of HAART at least twice.

by improving CD4 count, it does so at the immediate cost of inducing physical ailments via side effects. Hence, individuals may avoid going onto HAART so as 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.[22] 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.[23] This exercise suggests that a counterfactual world without HAART may have witnessed lower employment among HIV+ men. Table 1 also shows that individuals reporting physical ailments are less likely to engage in full-time work.

To explore employment patterns further, consider Table 5, which presents transitions into and out of the labor force for the pre- and post-HAART eras and for individuals with high and low CD4 counts. First, employment is very persistent across groups. Yet, healthier individuals are more likely to enter and less likely to leave the labor market in comparison to individuals with low CD4 counts. Moreover, for a given health state, persistence in both employment states is higher after HAART. Post-HAART persistence in employment may reflect better health, whereas post-HAART persistence out of full-time work likely reflects retirement (as average sample age is considerably higher after HAART is introduced).

TABLE 5. Transition matrix: employment.

		Time $t+1$	
		Not Full Time	Full Time
		Pre-HAART & Low CD4	
Time t	Not full time	0.92	0.08
	Full time	0.19	0.81
		Pre-HAART & High CD4	
Time t	Not full time	0.80	0.20
	Full time	0.07	0.93
		Post-HAART & Low CD4	
Time t	Not full time	0.95	0.05
	Full time	0.10	0.90
		Post-HAART & High CD4	
Time t	Not full time	0.89	0.11
	Full time	0.05	0.95

Note: Period-by-period employment transitions (full time or not full time) from periods t to $t+1$.

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

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

TABLE 6. Logistic regression of employment decisions.

	Employment Choice $t+1$ Conditional on Treatment Choices				
	(1)	(2)	(3)	(4) Pre-HAART	(5) 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***
Experience 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	7,954	7,954	7,954	3,694	4,260

Note: Dichotomous employment choices (full time or not full time) at period $t+1$ conditional on treatment choices in periods t and $t+1$. Health and not having physical ailments separately predict employment.

Still, it is not clear if physical ailments have an independent effect on employment (which would suggest side effects) or merely capture the effect of poor health (which would suggest symptoms). In other words, if ailments that discourage work simply reflect poor health, then it would be difficult to argue that individuals have an incentive to go off of effective medication to avoid working while suffering side effects. To examine the relationship between work and ailments, I present coefficient estimates from static logistic regressions where the dependent variable is a dichotomous employment choice (Table 6). Estimates indicate that, independently of CD4 count, physical ailments reduce labor force participation. Therefore, the data suggest that side effects discourage work. Further, a health-dependent cyclicality in employment emerges: individuals 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 low CD4 count, physical ailments in the form of symptoms, and death. In what follows, I specify a theoretical model to rationalize these patterns.

3. MODEL

In each period, agents enjoy flow utility, which is a function of current choices and state variables. Given that individuals in the sample are observed every 6 months, each period in the model is also 6 months long. 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.[24] 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 represents 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 $F_{it} = f \in \{0, 1\}$, where 1 signifies being free of ailments and 0 signifies suffering ailments. Flow utility is given by

$$\begin{aligned} U\big(C_{it}, F_{it}, d_{it}, d_{i,t-1}^M\big) = \sum_{f=0}^{1} \mathbf{1}\{F_{it} = f\} \\ \times \Big[u\big(C_{it}, f, \gamma(f)\big) + \theta_{1,f}^U + \big(\theta_{2,f}^U \times \mathbf{1}\big\{d_{it}^L = 1\big\}\big) \\ + \theta_{3,f}^U \times \mathbf{1}\big\{d_{i,t-1}^M = 0\big\} \times \mathbf{1}\big\{d_{it}^M \neq 0\big\} \\ + \theta_{4,f}^U \times \mathbf{1}\big\{d_{i,t-1}^M \neq d_{it}^M\big\} \times \mathbf{1}\big\{d_{i,t-1}^M \neq 0\big\} \\ + \theta_{5,f}^U \times \mathbf{1}\big\{d_{i,t-1}^M \neq 0\big\} \times \mathbf{1}\big\{d_{it}^M = 0\big\} + \varepsilon_{it}(d_{it}) \Big]. \end{aligned} \tag{2}$$

The first term on the right-hand side of equation (2) is a sum over each ailment status $F_{it} = f \in \{0, 1\}$ along with an indicator function. This term is multiplied with the remainder of the terms so that flow utility is health-state-dependent through the impact of health on ailments. 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 f and $u(\cdot)$ is a constant relative risk aversion (CRRA) utility function with parameter $\gamma(f)$ so that

$$u\big(C_{it}, f, \gamma(f)\big) = \frac{1}{1 - \gamma(f)} C_{it}^{1-\gamma(f)}. \tag{3}$$

[24] After retirement, evolution of utility is a function of survival probability, which is a function of the final treatment choice made upon retiring and health upon retirement. Variation in post-retirement utility is solely due to random shocks.

The second term $\theta^U_{1,f}$ represents the direct utility level effect of suffering ailments (when $f = 0$) or being free of ailments (when $f = 1$), where $\theta^U_{1,1}$, capturing utility when free of ailments, is normalized to zero. The third term on the second line is the ailment-specific utility cost of full-time work $\theta^U_{2,f}$. 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 the third to fifth lines of equation (2). The terms $\theta^U_{3,f}$, $\theta^U_{4,f}$, and $\theta^U_{5,f}$ capture the utility costs of going onto, switching, and going off of treatment, respectively, and are interacted with ailments. Finally, $\varepsilon(d_{it})$ is a choice-specific utility shifter, which captures factors that affect agent choices, but that are not observable to the econometrician. In particular, $\varepsilon_{it} : D_t \to \mathbb{R}$ and I use $\varepsilon(d_{it})$ to denote the utility shifter associated with choice d_{it}. Finally, $\varepsilon_{it}(d_{it})$ are extreme value Type I distributed.[25]

Switching costs capture factors—beyond preferences over ailments and long-term health—that affect agent treatment decisions, including the social benefits of HIV medications, concerns about drug resistance, and doctors' orders. Effective HIV treatments lower viral loads (the amount of virus in a patient's blood), which renders patients less infectious to HIV− sex partners. In other words, there is a positive externality associated with the use of medications and so it is possible that agents feel pressured to use medication to lower their infectiousness due to altruism or social norms. Further, medical decisions are likely influenced by a doctor or other medical professional even though the agent in the model is assumed to act alone. Both altruism (due to positive externalities of staying on drugs) and doctors' orders (perhaps reflecting fears of resistance) are captured by switching costs.[26] Finally, note that switching costs are generic, that is, 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, including generic switching costs, 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, in which case 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.

[25]This assumption, along with conditional independence of states and outcomes, which will be formally stated in Section 3.2, follows Rust (1987).

[26]The externality from using HIV medications raises the concern that switching costs likely vary by illness and so switching costs estimated in the context of HIV may differ from those in other contexts. This possibility underscores how the framework in this paper is possibly applicable to other scenarios, but that parameters would need to be estimated for each medical context.

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

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

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

$$H_{i,t-1} \in \{0,1\}, \quad \text{high (non-AIDS) CD4 count at } t,$$
$$A_{i,t-1} \in \{30, 30.5, 31, \ldots, 65\}, \quad \text{age at } t,$$
$$E_{i,t-1} \in \{10, 20, \ldots, 50\}, \quad \text{half-years of full-time experience at } t,$$
$$v_{t-1} \in \{1, \ldots, 15\}, \quad \text{period } t \text{ dummy}.$$

Recall from Section 2 that HIV infection leads to a low CD4 count, which means that the patient's immune system is compromised.[27] Since I observe agents every 6 months, work experience is measured in half-years (and one period of work means agents gain a half-year of work experience). Finally, S_{it} also includes the unobserved, choice-specific utility shifters ($\varepsilon(d_{it})$'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 $\mathcal{Y}_{it} = [F_{it}, C_{it}]$.[28] I assume conditional independence of $\mathcal{Y}_{it}$, that is, outcomes are independent of realizations of unobservable flow utility shifters. Formally,

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

Ailments F_{it} evolve according to

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

[27] 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 has crossed the threshold below which AIDS is likely (250). One argument against this simplification is that someone with a CD4 count of 1,000 is much less likely to reach AIDS-level CD4 count in the following period in comparison to an individual with a CD4 count of 500, but the binary variable puts these two into the same category. However, agents who have been HIV+ long enough to need medication rarely exhibit CD4 counts of 1,000 and most "high CD4" count individuals in the sample are between 300 and 600 and therefore face a more similar probability of getting AIDS. 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 model implications.

[28] 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.

where $X_{it}^F \equiv [H_{i,t-1}, v_{i,t-1}, H_{i,t-1} \times d_{it}^M]$ 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}).[29] 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 from missed work. 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

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

where $X_{it}^I \equiv [(E_{i,t-1}, E_{i,t-1}^2, A_{i,t-1}, H_{i,t-1}, v_{i,t-1}) \times d_{it}^L]$, $\varepsilon_{it}^I \sim N(0, \sigma_I^2)$, and θ^I denotes a vector of parameters governing the income process.[30] Note that state variables affecting the income process are interacted with period t employment decisions. This reflects the fact 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 by state variables and labor supply decisions.[31] Formally,

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

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

Underlying health, as measured by CD4 count, is affected by treatments. In capturing this effect, it is important to understand that effective medications like HAART raise CD4 count for unhealthy agents and maintain CD4 count for healthy agents. Given this

[29] Agents in the model cannot save. I discuss the implications of this assumption in Section 5.

[30] 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.

[31] Insurance could instead be modeled as a choice. However, the 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.

dual role, if one simply regresses CD4 count or changes to CD4 count onto medication use, the more effective drugs may appear to move CD4 the least. Therefore, I have developed a two-step procedure to appropriately capture the impact of drugs on health. First, ΔH_{it} indicates whether an agent's CD4 increased or remained unchanged (versus decreased) between periods t and $t+1$. The change ΔH_{it} evolves according to

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

where $X_{it}^{\Delta H}\equiv[H_{i,t-1},v_{i,t-1},d_{it}^{M}\times H_{i,t-1}]$. In other words, both treatments and period t CD4 count determine if CD4 count increases or not. Then period t CD4 count and the predicted 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

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

where $X_{it}^{H}\equiv[\Delta H_{it}\times H_{i,t-1}]$. In this manner, a drug that maintains CD4 count of a high CD4 count agent is not improperly categorized as a low-effectiveness drug.[32]

Out-of-pocket treatment costs are modeled as

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

where $X_{it}^{P}\equiv[H_{it}\times F_{it},I_{it},N_{it}\times d_{it}^{M},v_{it}]$, $\varepsilon_{it}^{P}\sim N(0,\sigma_{P}^{2})$, and θ^{P} is a vector of parameters.[33] Given the processes specified above, expected consumption is formally defined as

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

Note that both income and treatment costs are assumed to be nonnegative.[34]

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

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

[32]In computing standard errors, the fact that the CD4 count process is estimated in two steps where the second step includes a predicted value from the first step, is fully accounted for. See Appendix B.3 for a discussion of standard errors computation.

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

[34]The insurance and out-of-pocket treatment costs processes could be simplified without changing qualitative model predictions. The reason is that treatment costs are small compared to income and do not vary much. However, to allay concerns that cycling off effective treatments is due to treatment costs, I have chosen to model these processes to exploit all variation in state variables.

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

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

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, $E_{it} = E_{i,t-1} + 0.5 \times \mathbf{1}[d_{it}^L = 1]$. 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

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

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

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

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

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 (subsets of θ^U from equation (2)) to vary by type: the cost of ailments ($\theta_{1,f}^U$ with $f = 0$), the utility cost of work with and without ailments ($\theta_{2,f}^U$, where $f = 0$ and $f = 1$, respectively), and the marginal utility of consumption for each ailment status $\gamma(f)$. The utility of being free of ailments ($\theta_{1,f}^U$ with $f = 1$) is normalized to zero for both preference types. 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 from equations (6) and (10), respectively) and in parameters governing health transitions θ^H from equation (11).[35]

The decision to permit unobserved heterogeneity is driven by the data. First, there is high observed persistence in labor supply choices within individuals over time, which is consistent with permanent unobserved heterogeneity in distaste for work. Moreover,

[35]Conversely, 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.

medical research suggests that unobserved factors, including genetic variations, can imply permanent differences in reactions to HAART (see, for example, Scherer (2010)). Perhaps most importantly, I have also estimated models that do not permit unobserved heterogeneity in either utility or health outcomes. Key model predictions, including health-state-dependent cyclicality in treatment decisions, are also generated by the simpler model.[36] However, the simpler model is unable to match some important moments in the data. For example, a model that permits no unobserved heterogeneity in the impact of HAART on health outcomes predicts that all agents go onto HAART immediately upon reaching AIDS-level CD4 counts. This finding is broadly consistent with the data in the sense that low health predicts a relatively high likelihood of going onto treatments including HAART (refer to Table 3), yet the probability of cycling back onto HAART when in low health is not 1. Therefore, the data suggest that there is variation in agent health behaviors arising from factors that are not observable.

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 $K = 4$.[37] 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

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

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

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

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

4. Estimation

The vector of parameters ψ is estimated using a nested procedure.[38] At the "inner" step and given a proposed 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

[36] I return to this point when I present results and describe precisely which components of the model (identified from data moments) generate cyclicality in treatment decisions.

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

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

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 nonretired agents differs and each will be described in turn. The value of retirement is a constant stream of flow utility supposing that agents no longer work, given by

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

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 $\mathrm{P}[B_{it} = 1|\cdot]$.[39] Therefore, total retirement value for a given treatment choice and set of state variables is equal to an infinite sum, given by[40]

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

Let us now turn attention to nonretired 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 Bellman principle, I can define the value function for periods before retirement as

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

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

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

[39]The discount factor β is set to $\sqrt{0.95}$ per half-year.

[40]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}.

where $\bar{V}(\cdot)$ is the expectation of the value function taken over the distribution of $\varepsilon(d_{it})$ 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.[41] 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 $\varepsilon_{it}(d)$, but its distribution implies the 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 $\varepsilon_{it}(d_{it})$)

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

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

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

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

Portions of equation (28) can be extracted from the summation over k in cases where equation parameters are constrained to be equal across types, for example, in the equations describing the survival, 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

[41] 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.

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 (e.g., Stata). 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 computational burden and the time it takes to estimate the model.[42] Standard errors are computed taking the variance of first-stage estimates into account. In particular, I construct the full likelihood for the purposes of computing standard errors. Details are given in Appendix B.3.[43]

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, including those for income, insurance, out-of-pocket payments, physical ailments, and health. Parameters in the flow utility function are identified through observed state-dependent choice probabilities. Ailment-dependent utility parameters are identified through differences choices across ailment status for a given set of state variables. The quasi-experimental nature of HAART introduction, which implies that the same decision-makers are observed making choices over time and facing unanticipated variation in the features of available products, helps to identify utility parameters.

The CRRA coefficient γ, which measures the curvature in the consumption utility function, is identified by differences in how agents choose both treatments and employment at different consumption levels. Employment decisions imply large changes to consumption and treatment choices, which induce variability in medical expenditures, imply small changes to consumption. Insofar as choice probabilities vary for different consumption levels, these choices trace out the marginal utility of consumption. It would therefore be difficult to credibly identify γ if only few individuals—or only individuals with one level of consumption—made decisions leading to changes in consumption. To allay this concern, I first note that of the 743 individuals in the sample, fully 350 (47%) are observed switching from full-time to not full-time work at least once and 191 (26%) are observed switching into the labor market at least once. Moreover, in Appendix B.2, I plot income distributions for the entire sample of 8,300 observations and then separately for individuals when they are observed switching into or out of the labor market and then again for individuals when they switch treatments (Figure B.1). In both plots, the solid line is the distribution of income for the full sample. In comparison,

[42]The model could also be estimated in a single step, which would take longer, but would yield identical parameter estimates due to additive separability of the log likelihood function.

[43]Estimation in two stages also means that processes that are estimated in a first stage will not suffer from scaling issues. This is important with the income process, for example, where the scale of the parameters is up to 20,000 times the scale of parameters in other processes. If all parameters had been estimated jointly, it would have been necessary to scale income process parameters to avoid numerical problems in estimation.

average income for individuals switching out of the labor market is noticeably lower. It is even lower for individuals switching into the labor market since their current consumption consists of non-wage income (panel (a) of Figure B.1). Still, it is clear that labor supply changes occur across the distribution of income. Moreover, the income distribution for treatment switchers is very similar to the income distribution for the full sample (panel (b) of Figure B.1). Together, the figures provide evidence that decisions that trigger changes to consumption occur across the income distribution.

To identify unobserved heterogeneity in preferences, transitions, and outcome processes, I rely on the fact that the data set I use is a panel so that the same individuals are observed making decisions over time. Unobserved heterogeneity in preferences is identified by observing one set of individuals repeatedly making decisions over time that are consistently different from decisions of another set of individuals even though both sets of individuals face similar choice-specific distributions over outcomes. Similarly, unobserved heterogeneity in the health and ailments processes is identified from observably similar individuals making the same decisions over time, but facing different outcomes. For example, if two individuals consistently use HAART over time, but one of them exhibits better health, I can use this variation to identify persistent, unobservable differences in the health transition function.[44]

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 data set to investigate how latent types relate to variables not included in the model state space. Tables containing parameter estimates are provided in Appendix B.1.

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, that is, there are 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 A.1).

Preference parameter estimates (found in Table A.2) 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;

[44]Identification also requires normalizations. As mentioned in Section 3.1, the flow utility of no ailments (θ^U_{11}) is set to 0 for both latent types. If the agent has no ailments and neither works nor switches treatments, then flow utility is equal to consumption utility, computed as income minus out-of-pocket treatment costs, the latter partly determined by the insurance provision process.

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. Further, 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 and switching treatment imply a gain in utility. This may capture that agents with ailments find it easier to go off of treatment, perhaps since there is less pressure to stay on from their doctor or because fears of resistance or altruistic reasons to stay on are less compelling when agents face debilitating side effects.

Estimates of the CRRA utility parameters are in line with estimates discussed in Chetty (2006), who considers a variety of studies that find γ estimates between 0.15 and 1.78, with a mean of 0.71. My estimates range between 0.77 and 0.80, which means that agents are risk averse, but are slightly less risk averse than log utility would imply. Preference Type I indicates a small increase in the marginal utility of consumption with ailments. For preference Type II, there is no change. Positive state dependence arises if the HIV+ men I observe tend to consume goods that they value more when they feel worse. This finding is different from findings in Finkelstein, Luttmer, and Notowidigdo (2013), who also study health and the marginal utility of consumption. However, our studies are difficult to compare. For one, Finkelstein, Luttmer, and Notowidigdo (2013) use stated well-being to proxy for utility. Further, the samples we study may exhibit different consumption patterns when in relatively good health, which is possible since they study the elderly while I consider working-age HIV+ men.[45]

Moving on to health transitions and outcomes, the model reveals unobserved heterogeneity in drug effectiveness and side effects. For both latent health types, HAART is the most effective treatment in terms of increasing CD4 count (see Table A.3). 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 A.1, 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 (see Table A.4). The only coefficient that significantly varies by health type is the one governing the interaction between a high CD4 count and the predicted probability of a CD4 count increase.

[45]Further, ailments are not illness per se, but only reflect how an agent is feeling due to symptoms or side effects. In contrast, Finkelstein, Luttmer, and Notowidigdo (2013) study the marginal utility of consumption when individuals report chronic illness in comparison to when they do not. Strictly speaking, the individuals in my sample are chronically ill throughout the sample period.

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 A.5 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, that is, 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 A.6), a high CD4 count drastically reduces death probability.[46] 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 A.7, A.8, and A.9, respectively. Recall that income is modeled to be a function of high CD4, experience, experience-squared, age, and a time trend along with current employment fully interacted with these variables. Income increases with human capital (as measured by experience), 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 and disability payments are fairly high.

Health insurance (public, private, or no insurance, the last 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 and a lower probability of public insurance. Finally, treatment costs are a function of treatment choice, health status, and insurance. Estimates indicate that 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.[47]

[46]All parameters of remaining processes (including of the survival process) are estimated in a separate first stage using standard statistical software and then taken as given in the second-stage estimation of utility parameters and other structural parameters.

[47]Results from a model fit exercise are found in Table A.10. 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 percent-

Here, it should be noted that the income process could be simplified without affecting key qualitative results. The most important income patterns are the difference between wage and non-wage income along with controls for age and experience to capture human capital accumulation and separately identify it from age. Together, these factors capture incentives to work. In Appendix B.1, I consider alternative specifications for the income process. Table B.1 shows that working is more lucrative than not working and that income declines with age but rises with experience in all specifications. However, the table also shows that further interactions provide richer patterns on the returns to age, experience, and health for wage versus non-wage income. For example, the returns to experience are positive for workers and nonworkers, but stronger for workers. Therefore, the income process used in the structural model includes these interactions. Likewise, the insurance and out-of-pocket costs processes could be simplified (or, indeed, dropped altogether) without affecting main qualitative results. However, this would mean that the model would be unable to effectively rule out that treatment switching is driven by treatment costs, which it is not. Further, it would preclude policy simulations that affect the cost of treatments.

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.[48] 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.[49] 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 A.11. 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, for example, those requiring inflexible schedules or physical labor. To explore this possibility, I compute average 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

age 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.

[48]I construct likelihood contributions for each individual conditional on the individual being in one of the four possible latent types. The likelihood contribution of individual i conditional on type k parameters is

$$L_i(\theta) = \prod_{t=1}^{T_i} \mathrm{P}\big(d_{it}|\mathcal{X}_{it};\theta^k\big) \times \prod_{t=1}^{T_i} g_Y\big(\mathcal{Y}_{it}|\mathcal{X}_{it},d_{it};\theta^{XYk}\big) \times \prod_{t=1}^{T_i-1} g_X\big(\mathcal{X}_{i,t+1}|\mathcal{X}_{it},d_{it};\theta^{XYk}\big), \tag{29}$$

which is equation (28) computed separately by type. For each type and individual, this number is divided by each individual's actual likelihood contribution (which is the contribution calculated using equation (28)) given estimated unconditional type probabilities. The resulting posterior ratios measure how likely a given individual, given his behavior and outcomes, belongs to each of the four type combinations.

[49]Available data do not offer more specific occupational information. Moreover, occupation data were collected only in 1984, so health-induced occupation change is impossible to measure.

have a professional specialty or work in a craft industry, both of which may permit more flexible schedules.[50]

5. The value of pharmaceutical innovation

In this section, I use the estimated model to place a value on pharmaceutical innovations, including HAART. Next, I examine how, from the individual's perspective, 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. Finally, and 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.

5.1 *The value of HAART*

This section converts the value of HAART into a measure of willingness to accept payment in 2003 dollars (henceforth WTAP). To begin, for each set of state variables, I compute expected lifetime utility of being in state $\mathcal{X}_{it}$ at the first period after HAART is introduced. This occurs in three steps. First, for each set of state variables in the first period after HAART introduction, I compute a choice-specific net-of-error value, which is denoted $\tilde{V}(\mathcal{X}_{it}, d_{it})$ and is defined in equation (27). Second, I use the structure of the logistic errors to compute choice probabilities given the choice-specific net-of-error value. Third, I compute expected lifetime value as the weighted average of choice-specific value, where weights are the computed choice probabilities. The resulting quantity is denoted $\tilde{V}(\mathcal{X}_{it}[\text{HAART} = 1])$, where $[\text{HAART} = 1]$ signifies that I am simulating expected lifetime utility under the factual scenario where HAART has been invented. Next, I repeat the procedure under the counterfactual scenario that HAART is not invented. Instead, a third drug is introduced with the same attributes as combo-therapy.[51] Denote expected lifetime utility under this counterfactual regime $\tilde{V}(\mathcal{X}_{it}[\text{HAART} = 0])$, where $[\text{HAART} = 0]$ signifies that HAART has not been invented.

Once I have computed $\tilde{V}(\mathcal{X}_{it}[\text{HAART} = 1])$ and $\tilde{V}(\mathcal{X}_{it}[\text{HAART} = 0])$ for each set of state variables $\mathcal{X}_{it}$, I convert the difference into 2003 dollars. To do this, I start by adding income under the non-HAART scenario in the form of 2003 dollars at each possible fu-

[50]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.

[51]Agents therefore effectively face three types of drugs (no medication, mono-therapy, or combo-therapy), but face four choice-specific shocks. Valuing HAART in comparison to this scenario controls for the fact that the addition of HAART to the choice set gives agents an additional draw from the error distribution, which raises expected lifetime utility without improving technology and could therefore artificially inflate the value of HAART.

ture age and state (including post-retirement) and then recompute value. I continue this procedure until an agent's expected lifetime utility with the additional income is equal to his expected lifetime utility where HAART is invented, but without the additional income. This procedure essentially computes a compensating variation in the form of an annuity. Once I have added sufficient income to make the agent indifferent, I compute the present discounted value of the annuity, where I use expected years of life simulated with estimated model parameters.

There are many different ways to compute the value of a counterfactual medical innovation. One could simply ask a sample of individuals what they would be willing to pay for a drug with a given set of characteristics. Another common approach multiplies average expected gains to life-years gained by some amount of money. The approach I follow differs in that I explicitly rely on revealed preferences identified from observed agent decisions. This approach is designed to capture the fact that HAART extended life, but did so at the cost of harsh side effects. Agents revealed their ambivalence by not always using HAART, specifically, by not using HAART religiously when in relatively good health states. Hence, and in line with the broad idea behind computing the value of a "quality-adjusted life-year," extended years of life were certainly very valuable, but this value was attenuated by the side effects that came along with HAART use. Further, the computation of the present value of an annuity (weighted by expected survival in the pre-HAART era) may lead to conservative estimates of the value of the innovation relative to other approaches. I use this method because the structural model does not allow savings. Therefore, decreasing marginal returns to consumption would make the value of HAART excessively high were I to offer agents a lump sum in exchange for HAART not being made available.

WTAP for each type combination is presented in Figure 2, which graphs the present 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: it is 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

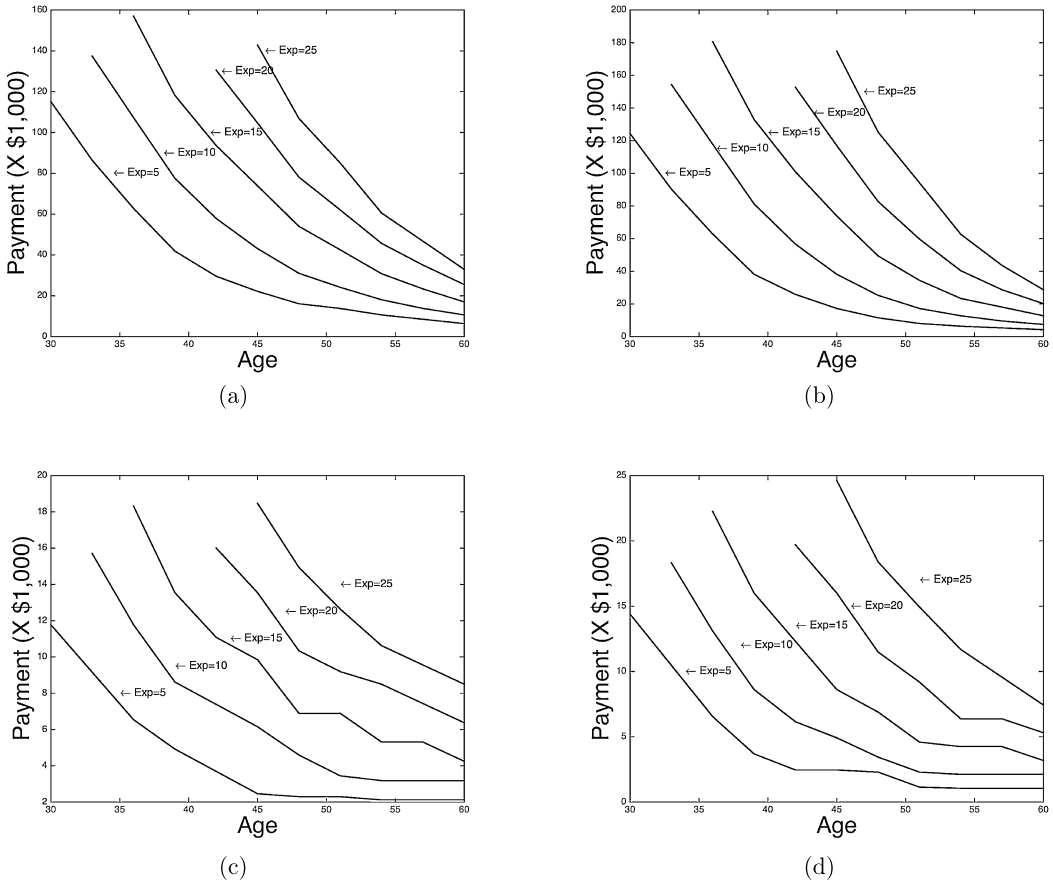


FIGURE 2. Heterogeneity in the value of HAART: the value of HAART for preference Type I and health Type I (panel (a)), preference Type II and health Type I (panel (b)), preference Type I and health Type II (panel (c)), and preference Type II and health Type II (panel (d)). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

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 (effectiveness and side effects), agents optimally choose to cycle among available treatments. Optimal treatment cycling generated by the model confirms evidence from the raw data

that 27% of individuals in the sample who ever use HAART are observed going off of HAART and also observed going back on (see Table 4). According to the model, health-state-dependent cyclicality in treatment decisions means that a dynamically optimal treatment path is a nonstationary closed loop driven by three factors: (i) persistence in underlying health, (ii) a nonconvexity in discrete treatment choices, and (iii) health-state-dependent flow utility captured by the estimated disutility of ailments induced by symptoms or side effects. Cycling may not occur in other medical contexts where one of these components is not present or, indeed, where switching costs are higher. 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

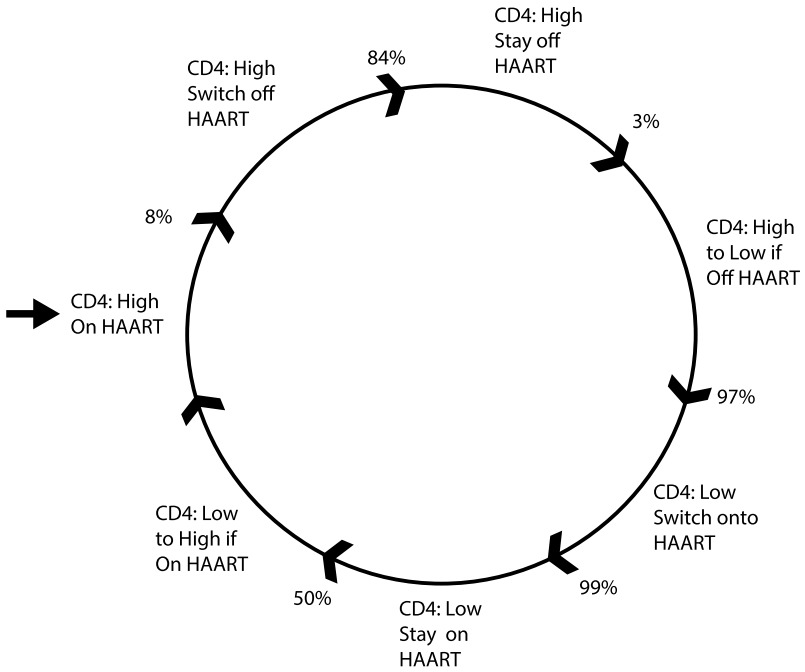


FIGURE 3. Optimal treatment cycling: probabilities along the cycle are simulated using model parameters. Beginning with the rightward pointing arrow at the left, preference Type I and health Type I agents with a high CD4 count cycle off of HAART with 8% probability in each period. Once off of HAART, 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. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

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 noncompliance 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)).[52] Optimal treatment cycling is similar in that current de-

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

cisions reflect previous treatment outcomes, though it is driven by patient decision-making.

It is crucial to note that cyclicality is rational from the individual's perspective, but may not be socially optimal. In the context of HIV, going off of medication can increase infectiousness, which has a negative impact on HIV− sexual partners. In a related matter, the patient is not the only person deciding on medical treatment, a choice that is usually made under a doctor's care. Indeed, doctors may encourage individuals to take effective medications despite uncomfortable side effects precisely because of the externality. In other words, doctors could act as imperfect agents in trying to pressure patients into taking drugs that have side effects. This would be consistent with findings in the medical literature that medical doctors, despite recommending 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)).

To capture the role of doctor's orders or altruism in light of positive externalities (along with the possibility that doctors encourage patients to stay on medications because of the positive externalities), the model includes switching costs. To assess the importance of switching costs, in supplemental analyses, I set switching costs to zero and then simulate behavior. I find that, absent switching costs, agents are more likely to switch off of HAART when in good health. This means that switching costs appear to capture pressures (beyond preferences over longevity, consumption, and ailments) that keep agents from going off of effective medicine. It also means that switching costs do not drive state-dependent cyclicality in treatment decisions, but rather attenuate it.

Finally, I note that optimal treatment cycling is not an artifact of other modeling choices either, including unobserved heterogeneity in preferences or health processes. In supplementary analyses, I estimate a simpler version of the structural model that does not permit unobserved heterogeneity. I find that cyclicality in treatment choices remains since the key factors driving it also remain: persistence in health across time, nonconvexities in the set of available treatment options, and a distaste for side effects, where the latter is identified by observing individuals going off of drugs with harsher side effects when they are in better health. As discussed earlier in Section 3.3, the problem with the simpler model is that it fails to replicate key data moments, vastly overpredicting cycling behavior, including the probability of going onto HAART when in poor health. Therefore, permitting heterogeneity captures unobserved factors driving variance in individual-specific match value with HAART, but does not drive key patterns that are potentially applicable to other medical contexts, such as health-state-dependent treatment cycling.[53]

[53]Another modeling assumption is that agents do not save. This assumption is made since I lack data on asset accumulation. Here, I offer brief remarks on the impact of omitting savings on key findings. The model predicts health-state-dependent cyclicality in treatment choices, which is partly explained by a distaste for work while suffering side effects. If agents save, then the drop in consumption they face from not working is smaller than the amount used to identify model parameters. If so, then the negative interaction between work and ailments is underestimated, which means that the incentive to go off of drugs to return to work

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 and II and health Type I) and Figure 5 (for preference Types I and 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.[54] 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 equivalent to HAART, but entails fewer side effects, generates high value.[55]

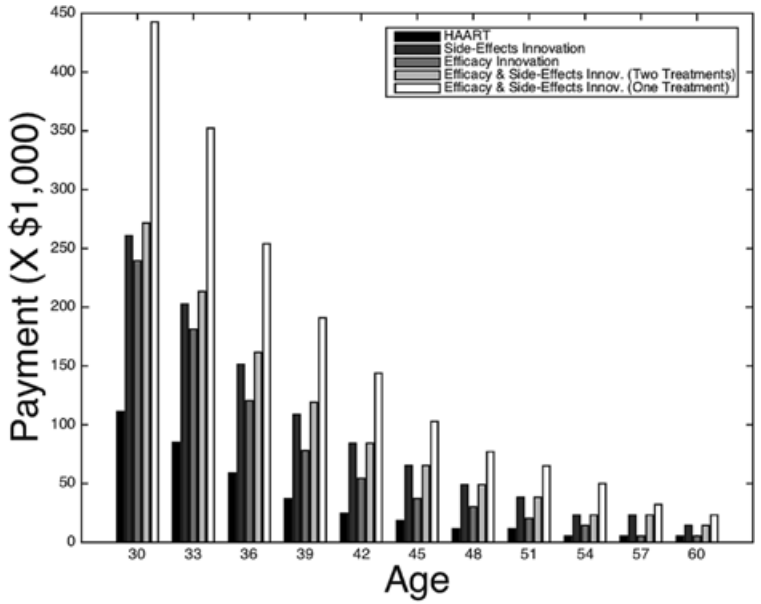
Suppose that instead of a reduction in side effects, HAART is improved along the effectiveness dimension. In particular, low-CD4 agents who use HAART face a 32% probability of non-AIDS CD4 in the following period. Under the counterfactual improvement, this probability is tripled. For each age, the third grey bar in Figures 4 and 5 depicts 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 upward of \$1,100,000 for a 30-year-old with 5 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

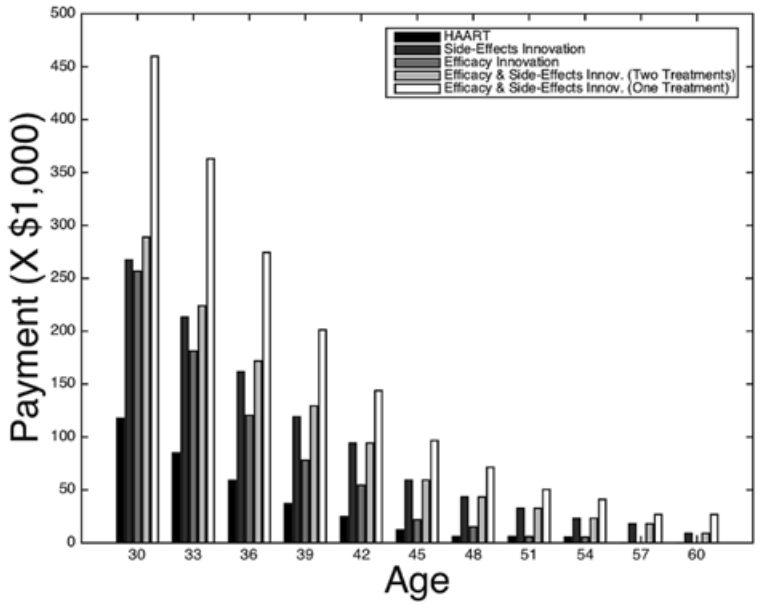
could be stronger for individuals with low savings. Also, there is a strong disutility of side effects that is independent of work decisions, which means that agent cycling across treatments is not driven solely by the desire to go back to work. Both points mean that omission of savings does not drive the model prediction of optimal treatment cycling. Finally, WTAP is computed using an annuity instead of using a lump sum payment to avoid overestimating value due to agents in the model being unable to save.

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

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

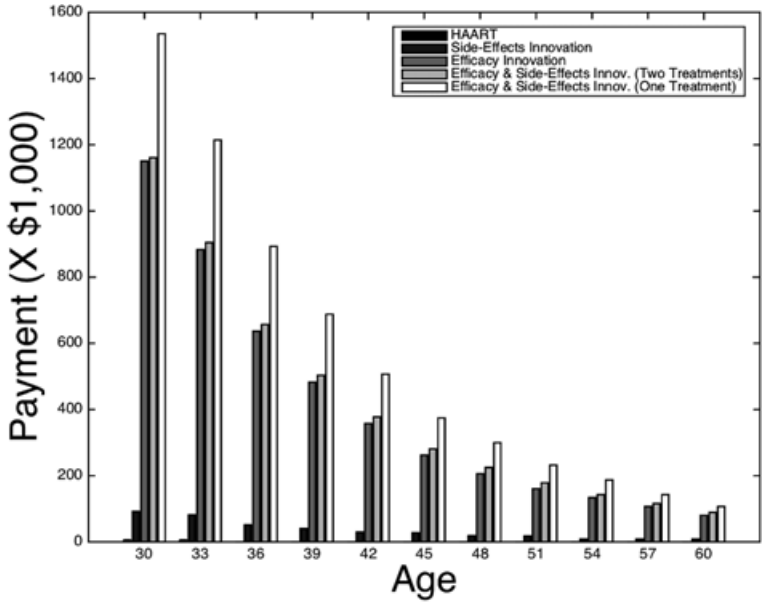


(a)

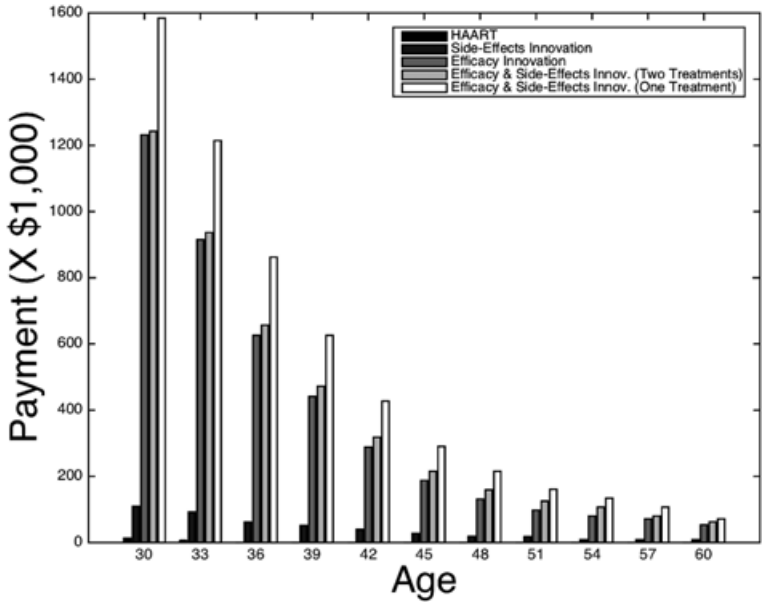


(b)

FIGURE 4. Heterogeneity in the value of pharmaceutical innovation: the value of counterfactual innovations for preference Type I and health Type I (panel (a)) and preference Type II and health Type I (panel (b)). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.



(a)



(b)

FIGURE 5. Heterogeneity in the value of pharmaceutical innovation: the value of counterfactual innovations for preference Type I and health Type II (panel (a)) and preference Type I and health Type I (panel (b)). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

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, that is, 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 effectiveness) occur simultaneously in separate treatments, so that two new drugs are introduced. Return to Figures 4 and 5 and consider the fourth bar for each age. As compared to the effectiveness innovation, the two innovations create little additional value. This finding is striking: a side effects innovation is valuable absent an effectiveness innovation, but creates little value given an effectiveness 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.[56] Nonetheless, combining an effectiveness and side effects innovation into a single drug does imply additional value since it permits agents to live without side effects, but to avoid risks associated with cycling. Such an innovation approximates 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.[57]

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.[58] I compare three of

[56]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.

[57]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.

[58]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.

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 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 the potential mechanism that public investments in biomedical research are partially offset by increased tax receipts due to increased labor supply.[59]

[59]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

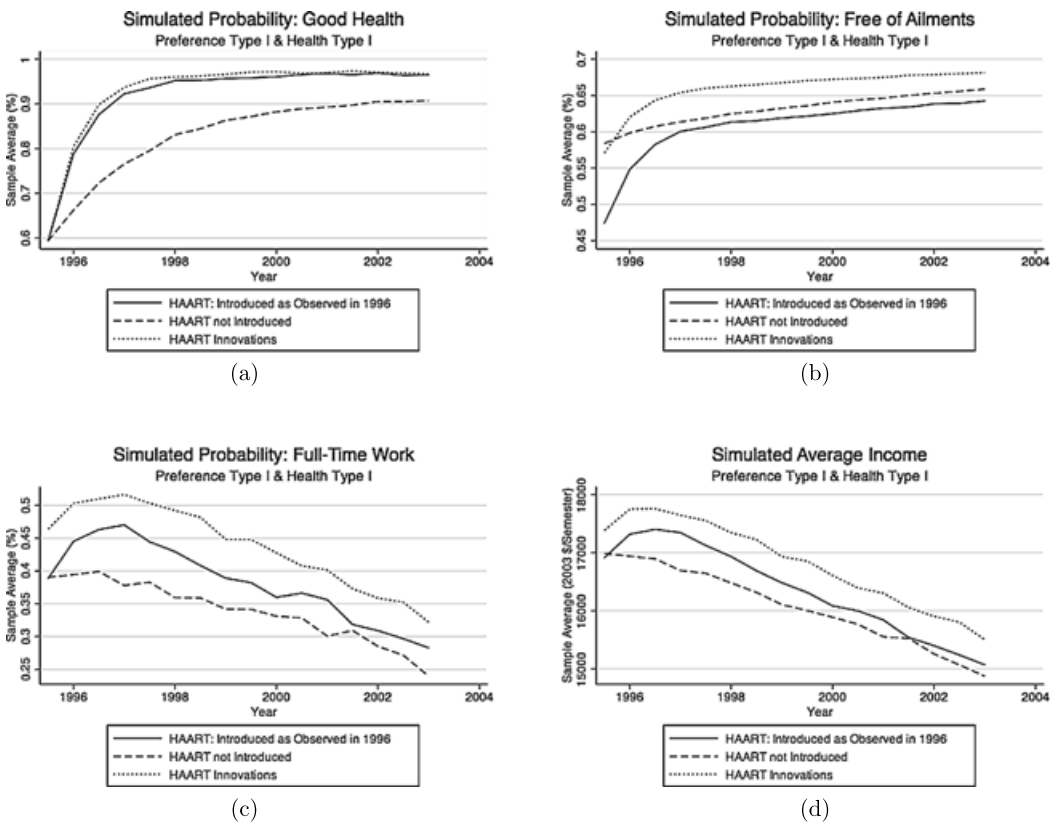


FIGURE 6. Counterfactual policy simulations—treatment innovations: three treatment environments are explored: (i) HAART is introduced as observed, (ii) HAART is not introduced, and (iii) instead of HAART two treatments are introduced, one with high effectiveness with HAART-level side effects, the other with HAART effectiveness and no side effects. For each simulated environment and for preference Type I and health Type I, panel (a) shows the average probability of high CD4 count over time; panel (b) shows average probability of not suffering from physical ailments; panel (c) shows average probability of working full time; panel (d) shows average net income in \$2003/year. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

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.[60] The goal of the following experiment is to as-

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 I agents will often avoid employment.

[60]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,

certain 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 by 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 Figure 7(a) and (b)) so as to improve their ailment status. When non-wage income declines

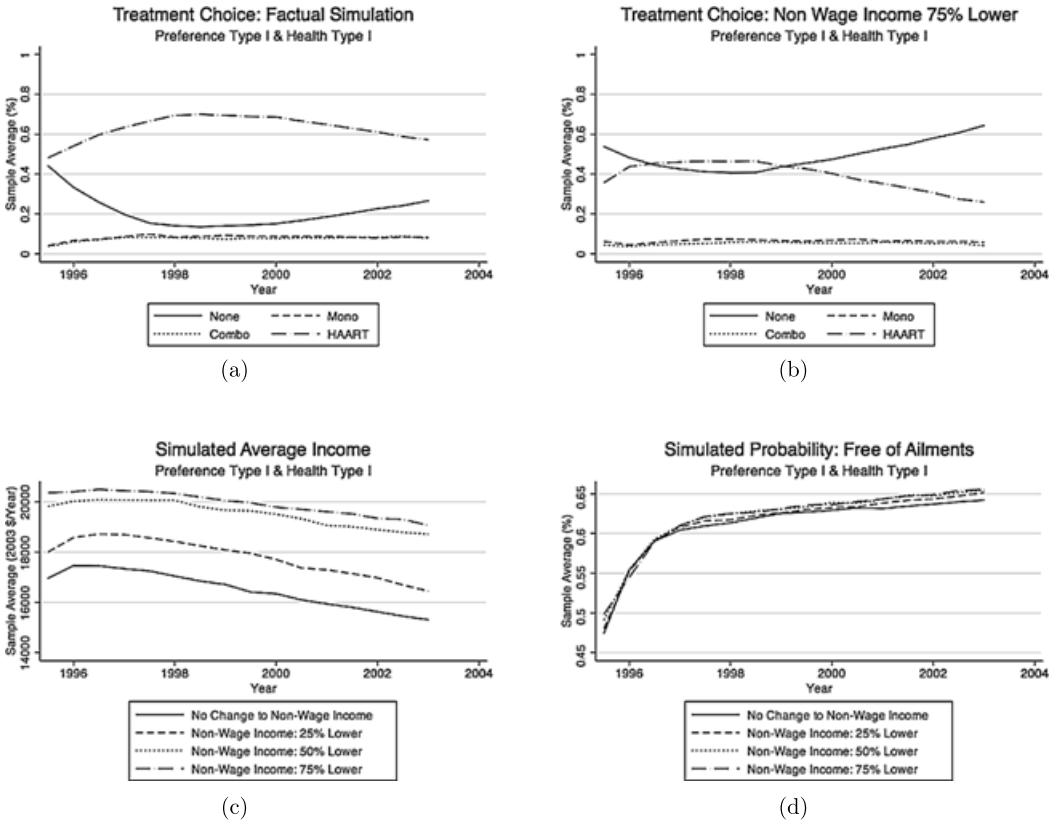


FIGURE 7. Counterfactual policy simulations—a decline in non-wage income: non-wage income is simulated to decline 25%, 50%, and 75%. For preference Type I and health Type I, panel (a) depicts treatment choices over time for no decline in non-wage income and panel (b) depicts treatment choices under a 75% drop in non-wage income. For each simulated environment, panel (c) shows average net income in $2003/year and panel (d) shows the probability of not suffering from physical ailments. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

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/.

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%.[61] 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. In other words, average health deteriorates marginally, but not enough to lower survival rates appreciably. 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 (in 2003 dollars) per year for treatment. However, the actual cost paid by insurance (both public and private) is much higher. A year of HAART therapy costs about $12,000, combo-therapy costs $8,000, and mono-therapy costs $6,000. 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 Figure 8(a) and (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 framework to value medical innovation that emphasizes the quality of life and highlights links between health, human capital, and the labor market.

[61] 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.

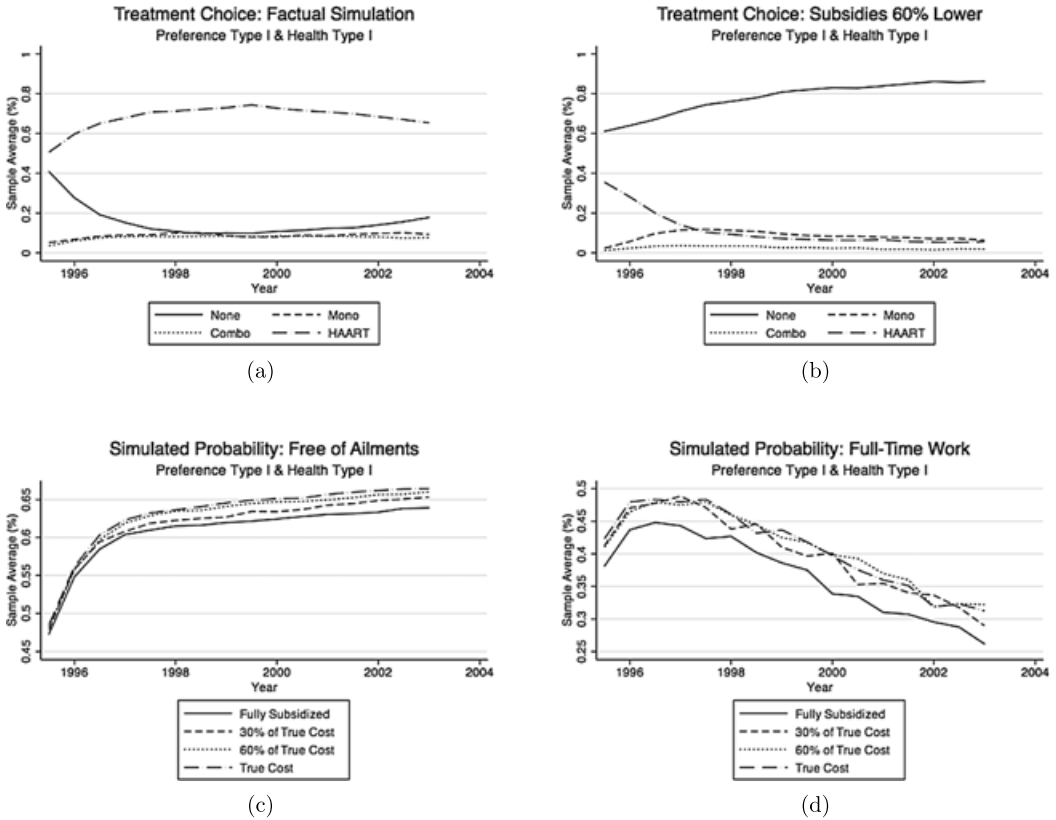


FIGURE 8. Counterfactual policy simulations—an increase in treatment costs: out-of-pocket treatments costs are simulated to increase to 20%, 40%, and 60% of actual treatment costs. For preference Type I and health Type I, panel (a) depicts treatment choices over time for no change in costs and panel (b) depicts treatment choices under a 60% change in costs. For each simulated environment, panel (c) depicts the probability of not suffering from physical ailments and panel (d) depicts simulated labor choice. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

A novelty of this project is my incorporation of ailments data to explicitly model medical treatments as having two dimensions of drug quality: effectiveness and side effects. Doing so, I am able to show that individual treatment choices are not consistent with strict longevity or health maximization. Rather, when no treatment dominates along all dimensions of drug quality, forward-looking agents optimally choose to cycle among available options.

Optimal treatment cycling reveals complex relationships between health and employment. 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 find-

ing 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.

Health-state-dependent cyclicality in treatment choices may be individually rational, but not necessarily socially optimal, especially when medication use exerts a positive externality on society. In the case of HIV, HAART renders users effectively noninfectious, which benefits individuals who are not infected, but who are at risk of becoming infected. Rather than undermining my findings, this externality has important implications for public health. In particular, if a policy-maker aims to harness the externality that HIV medication use entails, my findings suggest that the drugs to develop must not only be effective, but must also have mild side effects. Otherwise, agents will devise dynamically optimal plans to switch onto and off of drugs, thereby putting their sexual partners at greater risk of infection.

Future research could apply this framework 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. However, there are some caveats since switching costs in other medical contexts are surely different. Therefore, one could apply the framework in this paper, but would need to reestimate the model for each context. Future research could also extend the characteristics approach to other dimensions of 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 the quality of life.

Appendix A: Structural parameter estimates and model fit

Table A.1. Structural parameter estimates: type probabilities.

	Latent Type (Preferences)		
	Type I	Type II	$\sum$
Latent type (transitions and outcomes)			
Type I	0.148	0.262	0.410
Type II	0.345	0.245	0.590
$\sum$	0.493	0.507	

Note: Estimated unconditional latent type probabilities (π^K). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

TABLE A.2. Structural parameter estimates: utility.

	Type I		Type II	
	Coefficient	Error	Coefficient	Error
No ailments	–	–	–	–
× CRRA	0.81	0.02	0.80	0.02
× Labor disutility	−2.35	0.62	−2.59	0.64
× Begin treatment	13.60	4.57	–	–
× Change treatment	−6.18	0.39	–	–
× End treatment	−12.52	3.41	–	–
Ailments	−42.55	4.21	−53.70	3.93
× CRRA	0.77	0.02	0.80	0.02
× Labor disutility	−11.59	0.91	−5.26	0.64
× Begin treatment	−42.73	5.86	–	–
× Change treatment	4.32	0.55	–	–
× End treatment	30.07	3.30	–	–

Note: Estimated utility parameters (θ^U). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

TABLE A.3. Parameter estimates: CD4 count increase.

Variables	Coefficient	Error
High CD4	−0.40	0.10
× Mono	0.42	0.06
× Combo	0.50	0.07
× HAART [Type I]	0.77	0.11
× HAART [Type II]	0.71	0.11
Low CD4	–	–
× Mono	0.06	0.03
× Combo	0.09	0.03
× HAART [Type I]	2.20	0.43
× HAART [Type II]	0.17	0.06
Time trend	0.02	0.00
Constant	−0.70	0.12

Note: Estimated coefficients for the process governing period-by-period CD4 count increases ($\theta^{\Delta H}$). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

TABLE A.4. Structural parameter estimates: high CD4 at $t+1$.

	Type I		Type II	
Variables	Coefficient	Error	Coefficient	Error
High CD4	6.16	0.64	5.26	0.73
Predicted increase × Low CD4	4.33	1.02	3.89	1.46
Predicted increase × High CD4	2.93	0.38	0.66	0.18
Constant	−3.82	0.63	−3.71	0.71

Note: Estimated coefficients for the process governing one-period-ahead CD4 count (θ^H), where explanatory variables include the predicted likelihood of a CD4 increase as a function of current-period drug choices. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

TABLE A.5. Structural parameter estimates: ailments.

Variables	Coefficient	Error
High CD4	0.94	0.06
× Mono	−0.19	0.02
× Combo	−0.20	0.02
× HAART [Type I]	−0.20	0.03
× HAART [Type II]	−0.19	0.03
Low CD4	–	–
× Mono	0.26	0.05
× Combo	0.24	0.05
× HAART [Type I]	−1.04	0.24
× HAART [Type II]	0.25	0.05
Time trend	0.01	0.00
Constant	−0.45	0.06

Note: Estimated coefficients for the process governing whether the agents suffers ailments in the current period (θ^F). Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

TABLE A.6. Structural parameter estimates: mortality.

Variables	Coefficient	Error
High CD4	−6.16	0.82
Age × High CD4	0.07	0.02
Age	−0.01	0.01
Constant	−1.37	0.24

Note: Estimated coefficients of the mortality process (θ^B).

TABLE A.7. Structural parameter estimates: income.

Variables	Coefficient	Error
High CD4	1,508.47	132.33
Experience	664.04	15.19
Experience^2	−4.79	0.19
Age	−633.16	22.24
Time trend	165.88	10.08
Full-time employment	28,273.17	910.59
× High CD4	−871.93	186.41
× Experience	248.74	21.44
× Experience^2	1.58	0.32
× Age	−674.88	28.40
× Time trend	81.69	10.66
Constant	26,988.61	699.02
σ_I^2	6,673.70	31.61

Note: Estimated coefficients of the income process (θ^I).

TABLE A.8. Structural parameter estimates: insurance.

Variables	Coefficient	Error
Private insurance only		
Full-time employment	1.17	0.08
High CD4	−0.50	0.09
Experience	0.16	0.01
Experience^2	−0.00	0.00
Age	−0.30	0.07
Age^2	0.00	0.00
Time trend	0.07	0.02
Time trend (post-HAART)	0.02	0.03
Constant	6.97	1.36
Public only or public & private		
Full-time employment	−0.91	0.10
High CD4	−1.04	0.10
Experience	−0.03	0.02
Experience^2	−0.00	0.00
Age	0.16	0.08
Age^2	−0.00	0.00
Time trend	0.10	0.02
Time trend (post-HAART)	−0.05	0.04
Constant	−5.42	1.71

Note: Estimated coefficients of the insurance provision process (θ^N).

TABLE A.9. Structural parameter estimates: out-of-pocket treatment costs.

Variables	Coefficient	Error
High CD4	−105.21	27.38
High CD4 with ailments	169.93	15.33
Low CD4 with ailments	144.58	33.16
Income	44.58	3.41
Mono	323.58	90.91
Combo	167.65	68.17
HAART	−55.25	85.91
Private	−76.94	45.89
× Mono	6.54	89.53
× Combo	−59.83	95.16
× HAART	−248.97	133.18
Public	122.50	72.64
× Mono	−150.50	116.75
× Combo	389.09	86.86
× HAART	126.93	118.67
Time trend	13.60	1.30
Constant	−511.52	53.72
σ_p^2	682.76	0.90

Note: Estimated coefficients of the out-of-pocket payment process (θ^P).

TABLE A.10. 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	0.49	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	0.69	0.27	0.28	0.19	0.20	0.17	0.17	0.37	0.36
Exp ≤ 10	0.63	0.64	0.33	0.33	0.21	0.21	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	0.23	0.19	0.19	0.23	0.23
Pre-HAART	0.68	0.66	0.45	0.45	0.32	0.32	0.23	0.23	–	–
× Low CD4	0.50	0.50	0.25	0.26	0.40	0.40	0.35	0.33	–	–
× High CD4	0.76	0.74	0.54	0.54	0.29	0.28	0.18	0.18	–	–
Post-HAART	0.64	0.66	0.18	0.17	0.09	0.10	0.12	0.12	0.62	0.61
× Low CD4	0.41	0.48	0.12	0.09	0.11	0.13	0.11	0.12	0.66	0.67
× High CD4	0.69	0.70	0.19	0.19	0.08	0.09	0.12	0.12	0.61	0.60

Note: 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 A.11. Posterior type probabilities.

	Latent Type			
	Pref. Type		Health Type	
	Type I	Type II	Type I	Type II
Full sample	0.48	0.52	0.42	0.58
College				
No	0.54	0.46	0.40	0.60
Yes	0.44	0.56	0.42	0.57
Occupation				
Professional specialty	0.44	0.56	0.42	0.57
Admin. or clerical	0.50	0.50	0.40	0.60
Waiter	0.58	0.42	0.38	0.62
Craft	0.26	0.74	0.62	0.38
Mining	0.54	0.46	0.40	0.60
Transportation	0.59	0.41	0.45	0.55

Note: 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. Recall that for latent preference Type I, the disutility of work is sensitive to ailments. For latent health Type I, HAART is more effective, but entails harsher side effects.

APPENDIX B: SUPPLEMENTARY ANALYSIS AND NOTES

This section contains supplementary analysis and notes on the computation of standard errors.

B.1 *Alternative income process specifications*

The income process used in the structural model contains a number of interactions. The most important income patterns in generating model predictions are that income from full-time work is larger compared to income from not working full time, that better health leads to higher income, and that additional work raises future income through work experience. Together, these patterns generate the model predictions that (i) lost income due to not working full time can induce agents to eschew medications with side effects that make work difficult and that (ii) one incentive to remain healthy is higher income. In this appendix, I show that the key patterns are found in simpler specifications of the income process. Column 1 in Table B.1 reports coefficients from a Tobit regression relating income to a dummy variable for full-time work. Column 2 adds health. Column 3 adds a second-order polynomial in experience (measured by half-years of work experience). Column 4 adds age and a time trend. Column 5 fully interacts previous variables with the full-time-work indicator. Notice that in all specifications full-time work, health, and experience lead to more income. However, the specification in column 5

TABLE B.1. Tobit regressions: semiannual income.

	Income at $t+1$				
	(1)	(2)	(3)	(4)	(5)
Constant	13,430.71***	12,875.36***	9,156.18***	34,710.38***	26,988.61***
Full time $(t+1)$	8,193.38***	8,005.30***	7762.55***	5,930.47***	28,273.17***
High CD4	–	892.52***	790.49***	1,113.97***	1,508.47***
Experience	–	–	260.39***	712.61***	664.04***
Experience^2	–	–	−3.35***	−3.91***	−4.79***
Age	–	–	–	−835.81***	−633.16***
Time since combo	–	–	–	161.25***	165.88***
Full time $(t+1) \times$ High CD4	–	–	–	–	−871.93**
Full time $(t+1) \times$ Experience	–	–	–	–	248.74***
Full time $(t+1) \times$ Experience^2	–	–	–	–	1.58*
Full time $(t+1) \times$ Age	–	–	–	–	−674.88***
Full time $(t+1) \times$ Time trend	–	–	–	–	81.69**
Observations	7,954	7,954	7,954	7,954	7,954

Note: Additional income regressions including varying sets of explanatory variables.

shows that there is significant variation in the returns to health, age, and experience for full-time workers versus others. Therefore, the specification used in the structural model includes all interactions.

B.2 *Identification of the CRRA parameter* γ

In Section 4.1, I argue that identification of the curvature of the consumption utility function γ relies on variation in consumption generated by employment and medical treatment choices. It would be worrisome if this variation occurred only for limited income levels. In this appendix, I show that transitions into and out of the labor market and changes to medical treatment occur across the income distribution in the analysis sample. In Figure B.1, I plot a smoothed income distribution (using the command `kdensity` in the program Stata). The solid line is the unconditional smoothed density in both plots. In panel (a), I plot the density for individuals switching out of full-time work, into full-time work, or either out of or into full-time work. Notice that the density shifts to the left for switchers versus non-switchers, but that transitions occur across the income distribution. Similarly, panel (b) shows that the income distribution conditional on treatment changes is very similar to the unconditional distribution. This means that γ is identified from consumption changes rooted in employment and treatment choices that occur across the income distribution in the sample.

B.3 *Computation of standard errors*

I compute standard errors by constructing the Hessian of the likelihood function using the outer product measure. To compute the outer product measure, I calculate two-sided numerical derivatives of the likelihood function for each estimated parameter. In

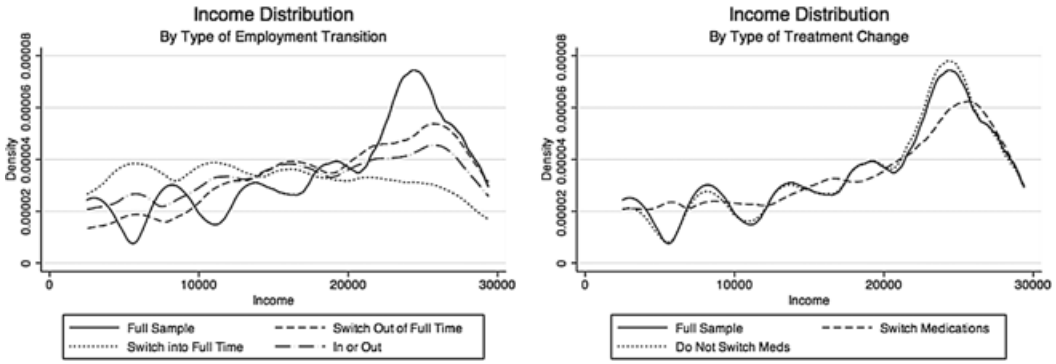


FIGURE B.1. Income distributions for the full sample and for individuals who transition into or out of work (panel (a)) or who switch medical treatments (panel (b)).

each direction, the derivative is calculated by perturbing each parameter, then solving for equilibrium choice probabilities, and then computing the likelihood. Although the estimation algorithm proceeds in two stages, where first-stage estimates are taken as fixed in solving for second-stage parameters, I compute errors using the full likelihood function that is conditioned on the full set of parameters as if they had been estimated jointly. In this way, I take full account of how first-stage estimates (taken as fixed in the second-stage of estimation) are measured with error. This method of calculating numerical derivatives therefore takes full account of how estimated first-stage parameters affect not only optimal responses given predicted outcomes, but also affect the processes governing the outcomes.

REFERENCES

Aguirregabiria, V. and P. Mira (2010), "Dynamic discrete choice structural models: A survey." *Journal of Econometrics*, 156 (1), 38–67. [692]

Angell, M. (2000), "The pharmaceutical industry—To whom is it accountable?" *The New England Journal of Medicine*, 342 (25), 1902–1904. [706]

Arcidiacono, P., P. B. Ellickson, P. Landry, and D. B. Ridley (2013), "Pharmaceutical followers." *International Journal of Industrial Organization*, 31 (5), 538–553. [673]

Becker, G. S. (1985), "Human capital, effort, and the sexual division of labor." *Journal of Labor Economics*, 3 (1), 33–58. [672]

Becker, G. S. (2007), "Health as human capital: Synthesis and extensions." *Oxford Economic Papers*, 59 (3), 379–410. [672]

Cawley, J. (2004), "The impact of obesity on wages." *Journal of Human Resources*, 39 (2), 451–474. [672]

Chan, T., C. Narasimhan, and Y. Xie (2013), "Treatment effectiveness and side effects: A model of physician learning." *Management Science*, 59 (6), 1309–1325. [674]

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. [674]

Chan, T. Y., B. H. Hamilton, and N. W. Papageorge (2016), "Health, risky behaviour and the value of medical innovation for infectious disease." *Review of Economic Studies*, 83 (4), 1465–1510. [675]

Chetty, R. (2006), "A new method of estimating risk aversion." *American Economic Review*, 96, 1821–1834. [697]

Ching, A. T. (2010), "Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration." *International Journal of Industrial Organization*, 28 (6), 619–638. [674]

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. [674]

Conti, G., J. Heckman, and S. Urzua (2010), "The education–health gradient." *American Economic Review*, 100 (2), 234–238. [672]

Crawford, G. S. and M. Shum (2005), "Uncertainty and learning in pharmaceutical demand." *Econometrica*, 73, 1137–1173. [674]

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. [672]

Currie, J. and B. C. Madrian (1999), "Health, health insurance and the labor market." In *Handbook of Labor Economics*, Vol. 3, Chapter 50, 3309–3416, Elsevier, Amsterdam. [672]

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), 347–355. [677]

Duggan, M. G. and W. N. Evans (2008), "Estimating the impact of medical innovation: A case study of HIV antiretroviral treatments." *Forum for Health Economics & Policy*, 11 (2), 1–39. [675]

Eckstein, Z. and K. I. Wolpin (1989), "Dynamic labour force participation of married women and endogenous work experience." *Review of Economic Studies*, 56 (3), 375–390. [672]

El-Sadr, W. M., J. D. Lundgren, J. D. Neaton, F. Gordin, D. Abrams, R. C. Arduino, A. Babiker, W. Burman, N. Clumeck, C. J. Cohen et al. (2006), "CD4+ count-guided interruption of antiretroviral treatment." *The New England Journal of Medicine*, 355 (22), 2283–2296. [675]

Fernandez, J. M. (2013), "An empirical model of learning under ambiguity: The case of clinical trials." *International Economic Review*, 54 (2), 549–573. [674]

Finkelstein, A., E. F. P. Luttmer, and M. J. Notowidigdo (2013), "What good is wealth without health? The effect of health on the marginal utility of consumption." *Journal of the European Economic Association*, 11 (S1), 221–258. [697]

Garthwaite, C. (2012), "The economic benefits of medical innovations: The case of Cox-2 inhibitors." *American Economic Journal: Applied Economics*, 4 (3), 116–137. [672]

Goldman, D. P. and Y. Bao (2004), "Effective HIV treatment and the employment of HIV+ adults." *Health Services Research*, 39 (6, part 1), 1691–1712. [684]

Grossman, M. (1972), "On the concept of health capital and the demand for health." *Journal of Political Economy*, 80 (2), 223–255. [671]

Hall, R. E. and C. I. Jones (2007), "The value of life and the rise in health spending." *Quarterly Journal of Economics*, 122 (1), 39–72. [672]

Heckman, J. J. and F. Cunha (2007), "The technology of skill formation." *American Economic Review*, 97 (2), 31–47. [672]

Hotz, V. J. and R. A. Miller (1993), "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies*, 60 (3), 497–529. [692]

Judd, K. L. (1998), *Numerical Methods in Economics*. MIT Press, Cambridge, MA. [694]

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. [674]

Lakdawalla, D., N. Sood, and D. Goldman (2006), "HIV breakthroughs and risky sexual behavior." *Quarterly Journal of Economics*, 121 (3), 1063–1102. [675]

Lancaster, K. J. (1966), "A new approach to consumer theory." *Journal of Political Economy*, 74 (2), 132–157. [673]

Lipscomb, J., M. Drummond, D. Fryback, M. Gold, and D. Revicki (2009), "Retaining, and enhancing, the QALY." *Value in Health*, 12, S18–S26. [672]

Meredith, P. A. (1996), "Therapeutic implications of drug holidays." *European Heart Journal*, 17 (Suppl. A), 21–24. [704]

Mincer, J. and S. Polachek (1974), "Family investments in human capital: Earnings of women." *Journal of Political Economy*, 82 (2), 76–108. [672]

Murphy, K. M. and R. H. Topel, eds. (2003), *Measuring the Gains From Medical Research: An Economic Approach*. University of Chicago Press, Chicago, IL. [672]

Murphy, K. M. and R. H. Topel (2006), "The value of health and longevity." *Journal of Political Economy*, 114 (5), 871–904. [672]

Murphy, S. A. (2005), "An experimental design for the development of adaptive treatment strategies." *Statistics in Medicine*, 24 (10), 1455–1481. [704]

Petrin, A. (2004), "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, 110 (4), 705–729. [673]

Philipson, T. J. and R. A. Posner (1993), *Private Choices and Public Health: The AIDS Epidemic in an Economic Perspective*. Harvard University Press, Cambridge, MA. [675]

Rosen, S. (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy*, 82 (1), 34–55. [673]

Rust, J. (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica*, 55 (5), 999–1033. [687, 692]

Sabate, E. (2003), *Adherence to Long-Term Therapies: Evidence for Action*. World Health Organization, Geneva. Available at www.who.int/chp/knowledge/publications/adherence_introduction.pdf. Accessed 4 Oct 2011. [704]

Scherer, R. M. (2010), "Pharmaceutical innovation." In *Handbook of the Economics of Innovation* (B. H. Hall and N. Rosenberg, eds.), Chapter 12, 539–571, Elsevier, Amsterdam. [692]

Stebbing, J. and A. Dalgleish (2009), "Not so SMART?" *Lancet Infectious Diseases*, 9 (5), 268–269. [675]

Stigler, G. J. (1945), "The cost of subsistence." *Journal of Farm Economics*, 27 (2), 303–314. [673]

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. [705]

Co-editor Petra E. Todd handled this manuscript.

Submitted May, 2014. Final version accepted September, 2015.