Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior

Michael Darden

George Washington University

I estimate a dynamic, stochastic model of smoking, expectations, and health that makes explicit channels through which an individual may learn about the health risks of smoking. Simulations of the structural model suggest that cardiovascular biomarker information at repeated health exams does not significantly alter smoking behavior because (*a*) signals of biomarker health are noisy within individuals, (*b*) the influence of biomarkers on major health outcomes is small, and (*c*) cigarette smoking is addictive. This paper also presents evidence of selection in smoking that, when not modeled, may cause an overstatement of the effect of smoking on expected longevity.

I. Introduction

The percentage of adults in the United States who smoke cigarettes has declined from 43 percent in 1965 to 19.3 percent in 2010. Figure 1 shows this decline across age groups and gender. The economic view of smoking is that individuals make decisions within an environment that reflects individual preferences but that is also subject to the costs of information

I am grateful to Donna Gilleskie, Helen Tauchen, Brian McManus, R. Vijay Krishna, Frank Sloan, Matthew Harris, Christopher Cronin, Manasigan Kanchanachitra, Jonathan Pritchett, Jay Shimshack, James Alm, James J. Heckman, and four anonymous referees for helpful comments, conversations, and guidance. The Framingham Offspring Study (FOS) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the FOS investigators. This manuscript was prepared using a limited access data set obtained from the NHLBI and does not necessarily reflect the opinions or views of the FOS or the NHLBI.

Electronically published August 24, 2017
[*Journal of Political Economy*, 2017, vol. 125, no. 5]

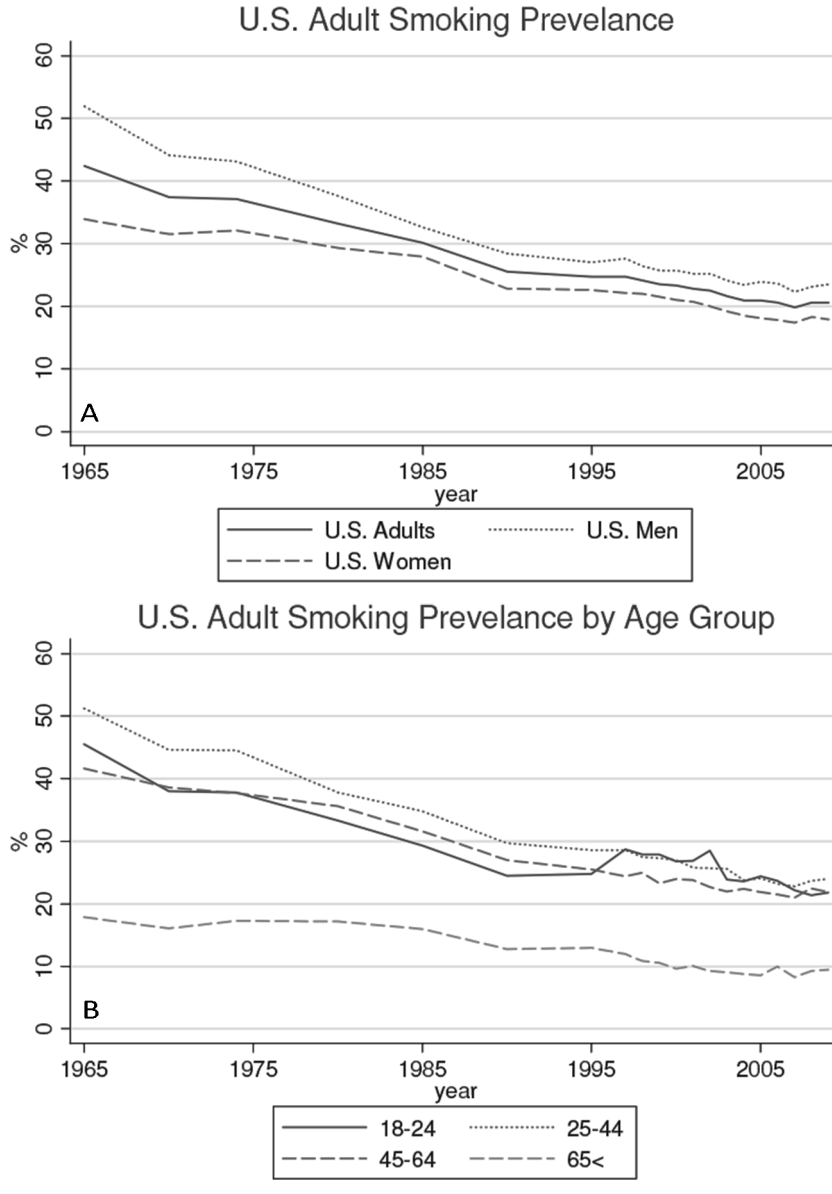

FIG. 1.—Smoking prevalence in the United States: percentage of US adults reported as current smokers by age and gender. Data are from the National Health Interview Survey and the American Lung Association.

acquisition.[1] The role of information may be an explanation then for the observed decline in cigarette smoking. An individual may change her smoking behavior if she rationally updates her beliefs regarding the extent to which cigarette smoking enters her health production function.[2] From a policy perspective, if smokers are unaware of the health hazards of smoking, then policies that promote awareness may induce smokers to quit. Luther L. Terry, surgeon general at the time of the landmark 1964 US surgeon general's report that linked smoking to lung cancer and certain birth defects, said that it "hit the country like a bombshell. It was front page news and a lead story on every radio and television station in the United States" (http://profiles.nlm.nih.gov/NN/Views/Exhibit/narrative/smoking.html).

However, if smokers have an "it won't happen to me" attitude, then general sources of health information may not induce smoking cessation. Indeed, Sloan, Smith, and Taylor (2003) suggest that heavy smokers require relevant and clear personalized messages to induce a change in their behavior. Most economic research on learning from personalized health messages frames major, chronic health shocks (e.g., myocardial infarction, cancer diagnoses) as informative events on the idiosyncratic effect of smoking on health (see, e.g., Smith et al. 2001; Khwaja, Sloan, and Chung 2006; Arcidiacono, Sieg, and Sloan 2007). Most papers also focus on individuals over age 50, when the major health implications of smoking begin to be observed. However, waiting for a major, later-life health shock to induce smokers to quit smoking may not result in life expectancy gains. Therefore, the first contribution of this paper is to study the role of personalized cardiovascular biomarker (e.g., blood pressure) information, conveyed by a physician, in the decision to smoke cigarettes. If learning of biomarker values that predict more serious conditions in the future can convince smokers to quit before the experience of major, later-life health shocks, then it might be worthwhile to institute policies to encourage obtaining biomarker information on a regular basis.

There exists a large medical and epidemiological literature on the effects of physicians using a variety of personalized biomarker signals to promote smoking cessation; however, the results of these studies have been sensitive to the specific signal of information and to the recall period.[3]

[1] In addition to the seminal rational addiction theory of Becker and Murphy (1988), other economic models of smoking capture regret and learning (Orphanides and Zervos 1995); time-inconsistent preferences (Gruber and Kőszegi 2001); cue-triggered mistakes (Bernheim and Rangel 2004); and compulsive consumption, temptation, and self-control (Gul and Pesendorfer 2001, 2007).

[2] Recent evidence from Hai and Heckman (2015) demonstrates the importance of treating health production as endogenous when modeling addictive behaviors, human capital investment, labor supply decisions, and savings behavior.

[3] See McClure (2001), Stead, Bergson, and Lancaster (2004), and Bize et al. (2009) for reviews of this literature.

For example, Parks et al. (2008) randomize whether a smoker is informed of her "lung age"—the age of an average healthy individual who would score similarly on pulmonary tests—at a general practitioner's appointment. They find that revealing lung age significantly increases the probability of smoking cessation. On the other hand, Hollands et al. (2012) randomize physician warnings on the elevated risk of Crohn's disease caused by cigarette smoking for a population of smokers with familial risk for the disease; they find no effect of those warnings on the probability of smoking cessation 6 months after randomization. Most studies measure outcomes as a snapshot of smoking behavior 6–12 months after randomization; none of the reviewed studies examine repeated, long-term exposure to cardiovascular biomarker information.[4]

To evaluate the role of repeated signals of cardiovascular biomarker information on smoking behavior, I examine changes in biomarkers and changes in smoking behavior from the 30-year longitudinal Framingham Heart Study (FHS)—Offspring Cohort. The data capture a wide variety of smoking and health transitions for a panel of roughly 5,000 individuals from 1971 to 2001. I report results from reduced-form smoking models in which current-period smoking is a function of recent changes in observed biomarker levels.[5] Both ordinary least squares and individual fixed-effects models of smoking behavior find potentially important effects of recent changes in biomarkers. For example, experiencing a recent single standard deviation increase in systolic blood pressure implies a 2 percent decrease in the probability of smoking. However, the results are generally mixed and insignificant, and increases in other biomarkers (e.g., total cholesterol) are found to slightly increase the probability of smoking.[6] Consistent with the literature cited earlier, the onset of cardiovascular disease or cancer (i.e., major health shocks) is found to significantly increase the probability of cessation, by 21 percent conditional on survival. Again, conditioning on survival may not serve policy interests.

If smokers are forward-looking with respect to health and they value future utility, then myopic empirical models of smoking responses to health changes will miss the inherently dynamic trade-off between current con-

[4] Smoking causes higher blood pressure and increases risk for hypertension because of increased arterial stiffness (Doonan et al. 2010). Smoking may also decrease high-density lipoprotein (HDL ["good" cholesterol]; Gepner et al. 2011).

[5] Most empirical studies of smoking demand, following from the rational addiction theory of Becker and Murphy (1988), estimate smoking models that include lag and lead prices—assuming perfect foresight with respect to prices—but abstract from health (see, e.g., Becker, Grossman, and Murphy 1994). Because the empirical literature suggests small price elasticities for middle-aged and older smokers and because of data limitations discussed below, I abstract from price effects in this paper.

[6] This paper abstracts from the important question of whether smoking and diet and exercise, e.g., are substitutes or complements in health production. See Kaestner, Darden, and Lakdawalla (2014) for evidence of substitutability between statin pharmaceutical use and smoking in the FHS—Offspring Cohort.

sumption and the potential lower future utility attributable to smoking-related health complications.[7] Furthermore, many standard reduced-form, dynamic models leave implicit agent expectations or assume that smokers have perfect foresight. Therefore, I instead develop a dynamic stochastic model of individual smoking decisions, health expectations, and health (both morbidity and mortality) over the life cycle, and I estimate the model with the same FHS data. This structural model identifies channels through which an individual may learn, via biomarker information, about the extent to which smoking is harming his or her own health. The model is consistent with that of Becker and Murphy (1988) in the sense that it allows for addiction mechanisms of reinforcement, tolerance, and withdrawal that may prevent responses to negative changes in health.[8] In addition to the structure and timing of this model, its identifying assumption is that the variation in the timing of FHS health exams influences smoking decisions, through the learning channel, but does not influence health. So, on the margin of two individuals with identical states but varied timing of exams, I identify the effect of the signal of health information.

To evaluate the roles of learning and information, I use the model and the estimated structural parameters to simulate smoking behavior and health and mortality outcomes under various counterfactual scenarios. While estimated expectations change with biomarker information, biomarkers are rather noisy signals of health information, and this information is not found to significantly change smoking behavior. A simulation in which the signal of information (i.e., the biomarker results) is clearer because of a smaller noise term suggests only a small decrease—not more that 2.5 percentage points—in smoking in later life. Furthermore, marginal changes in cardiovascular biomarkers are found to have modest effects on chronic health and mortality outcomes, and thus individuals may not view biomarker signals as significant evidence of the harm of smoking. Indeed, a simulation in which the perception of biomarkers as more influential in predicting morbidity and mortality outcomes suggests small declines in smoking. Finally, even in the event of clear, uniform signals of poor health, addiction may prevent individuals from quitting. I

[7] The economic literature on smoking responses to health changes typically has estimated myopic models of smoking behavior (see, e.g., Clark and Etilé 2002); however, several papers have found evidence of forward-looking behavior with respect to health (Viscusi 1990; Arcidiacono et al. 2007; Viscusi and Hakes 2008) and addiction (Adda and Lechene 2001; Gilleskie and Strumpf 2005).

[8] Becker and Murphy's (1988) "addictive stock of smoking capital" presents a latent variable problem. In estimating the structural model, this state variable is captured by a dynamic factor model with several endogenous smoking history measurements. The factor method identifies the latent state dynamics that reinforce current smoking, while also handling measurement error in the smoking history measurements. See Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010).

find that the reinforcement effect—the positive effect of past smoking on the current marginal utility of smoking—and the withdrawal effect—the reduction in utility following quitting—both dramatically encourage continued smoking in later life.

The simulation results present suggestive evidence that emphasizing cardiovascular biomarker data may not be an effective method of encouraging smoking cessation. Of course, the validity of the simulation results depends crucially on the assumptions of the model, and I cannot rule out the potential for biased simulation results stemming from model misspecification. Indeed, in the absence of data on subjective expectations, modeling an individual's expectations as potentially important in explaining observed smoking patterns requires strong assumptions regarding an individual's beliefs. However, simulation results that suggest small behavioral changes provide additional evidence when considered with evidence from reduced-form empirical models in which different underlying assumptions are made. Furthermore, the structural approach allows for the exploration of possible explanations for the lack of behavioral changes when confronted with personalized cardiovascular biomarker information.

The second contribution of this paper is to analyze whether current knowledge of the magnitude of health effects of smoking is free of selection bias. For example, Doll et al. (2004) use survey data of British physicians over several decades to assess the impact of cigarette smoking on mortality. Their findings suggest that lifelong smoking causes a 10-year reduction in expected longevity. However, these authors do not control for the endogeneity of smoking with respect to health outcomes, nor do they consider the possibility that individuals in different states of health may select into smoking. If, independent of smoking, smokers are of worse overall health status than nonsmokers, then standard statistical methods may overstate the effect of smoking on mortality. Indeed, Strine et al. (2005) show that smokers are more prone to alcoholism, depression, physical inactivity, pain, and sleeplessness. Yet health economists have only recently begun to model smoking and health outcomes jointly. For example, Adda and Lechene (2013) show that smoking behavior and potential longevity are correlated along observed and unobserved (to the researcher) dimensions.

The structural model of this paper, by estimating health and mortality production technologies within a model of lifetime smoking decisions, is well suited to evaluate the effect of lifelong smoking on expected longevity. I allow the unobserved errors that affect smoking and mortality to be serially correlated through a common permanent unobserved component.[9] I then simulate mortality outcomes under different patterns of

[9] The error structure is similar to recent structural models that have accounted for unobserved heterogeneity (Arcidiacono et al. 2007; Blau and Gilleskie 2008) and is based on Heckman and Singer (1984).

lifetime smoking behavior. I find that, for daily light and heavy smoking from age 18, individuals can expect roughly 3.13 and 5.41 fewer years of longevity, respectively. I contrast the estimated expected longevity loss from simulations of the structural model to the results from simulations of a reduced-form outcome (e.g., mortality) model that conditions only on observable heterogeneity. Expected longevity loss is 3.8 and 6.6 years for light and heavy smoking, respectively, in the reduced-form simulations. These results indicate a strong positive correlation between smoking tendencies and the underlying factors that influence mortality outcomes. If ignored, this leads to an overstatement of the longevity effects of cigarette smoking.[10]

This paper proceeds as follows. Section II provides evidence on smoking, information, and selection with respect to health from reduced-form analysis of the FHS data. Section III presents the structural model. Section IV describes the empirical implementation of the structural model. That section includes a discussion on the latent factor model, identification, unobserved heterogeneity, and initial conditions. Parameter estimates, model fit, and counterfactual simulations are presented in Section V. Section VI presents conclusions.

II. The Framingham Heart Study

The Framingham Heart Study is one of the longest-running panel studies in the world. The study began collecting health data in 1948 on 5,107 individuals living in Framingham, Massachusetts. They formed what became known as the Original Cohort, and in 1971, their offspring, the Offspring Cohort, began to be followed as well. Each cohort represents a unique panel study that has continued into the twenty-first century. With a stated goal of "identifying common factors that contribute to cardiovascular disease," several groundbreaking epidemiological findings have come from study data, including the relationship between smoking and cardiovascular disease (http://www.framinghamheartstudy.org/index.html).

This paper focuses on data from the Framingham Offspring Cohort because of the consistency with which health exams were administered.[11] In addition to seven health exams in which biomarker and smoking data are available, the Offspring Cohort data yield specific dates for various cardiovascular events (e.g., myocardial infarction), cancer diagnoses, and

[10] Darden, Gilleskie, and Strumpf (2017) examine this question thoroughly using data from 1948–96 from the FHS Original Cohort. Simulating a model that accounts for observed and unobserved heterogeneity, those authors find a 4.3-year reduction in expected longevity from continued smoking. Difference-in-means statistics from their data suggest that the difference is 9.3 years.

[11] Smoking and health questions varied over time in the Original Cohort, and constructing uniform measures of biomarkers proved to be difficult.

TABLE 1
SAMPLE CONSTRUCTION

Observations	Description
4,989	Framingham Heart Study—Offspring Cohort participants—limited-access sample
3,730	Sample after dropping all person/year observations of individuals who skipped one or more of the health exams
3,012	Sample after dropping all person/year observations of individuals who attrite
2,611	Final estimation sample after dropping all person/year observations of individuals with missing smoking or health variables

NOTE.—2,611 unique individuals yield 16,933 person/year observations.

mortality. In the initial exam wave, the behavioral survey also asked retrospective smoking questions that allow for the construction of lifetime smoking data through the seventh health exam.

There were 5,124 individuals in the first FHS offspring exam; of those, 4,989 individuals consented to having their records appear in a limited-access data set. The main drawbacks of both Framingham cohorts are their lack of geographic variation and the limited demographic variation: for confidentiality reasons, the sampling procedure considered only white residents of Framingham. Therefore, the FHS sample is not representative of a population beyond Framingham, Massachusetts, and limiting the sample does not come with a great loss of external validity. In this paper, I drop individuals from the analysis for missing one or more exams, attrition, or missing key smoking or health variables.[12] The resulting sample contains 2,611 individuals and 16,933 person/year observations. Table 1 provides information on sample construction.

The average length between health exams is 4 years, but across participants, exams occurred at varying time intervals. This variation is important in identification of the structural model presented below. Table 2 provides information on the mean and standard deviation number of years between exams, along with summary statistics by health exam. There is significant variation in ages across the sample: the first health exam was conducted between 1971 and 1975 for individuals ranging in age from 13 to 62. Because those who leave the study have been dropped,

[12] At some point during the seven exams, 718 individuals are lost to attrition (i.e., some reason other than death). This constitutes approximately 14.4 percent of the sample. Simple *t*-tests for difference of means suggest that those participants are slightly more likely to be women, have a 3 percentage point lower level of SBP on average, and have a statistically insignificant difference in incidence of coronary heart disease from their nonattriting counterparts. Those that drop out are, on average, slightly more likely to smoke. I restrict the sample in this way for the sake of comparing the results presented in this section to those of the structural model below. Simulation results are not sensitive to a sample with imputed missing values vs. simply dropping individuals with missing values. See the Appendix for further details.

TABLE 2
SAMPLE CHARACTERISTICS BY EXAM

Exam	Years between Exams	Individuals	Mean Age	Female	At Least Some College	Current Smoking	Cigarettes/Day Conditional on Smoking
1	. . .	2,611	36.926	.481	.579	.414	21.512
							(12.506)
2	7.888	2,544	44.554	.485	.595	.365	20.253
	(.660)						(14.304)
3	4.354	2,479	48.669	.492	.600	.266	23.423
	(.676)						(13.370)
4	3.462	2,431	51.945	.495	.602	.216	22.319
	(.583)						(12.802)
5	3.645	2,360	55.308	.502	.607	.172	21.442
	(.561)						(12.278)
6	4.074	2,276	59.053	.510	.616	.133	20.169
	(.667)						(11.998)
7	2.920	2,232	61.795	.513	.619	.115	17.816
	(.934)						(10.418)

NOTE.—Ages in the sample range from 13 in exam 1 to 89 in exam 7. The first exam was taken between 1971 and 1975. Exam time gap and smoking intensity standard deviations are in parentheses.

the number of individuals at each exam reflects only those who have survived. The baseline sample reflects a well-educated cohort for the time period that, because of selective mortality, becomes increasingly well educated and female over time. Table 2 documents a decline in smoking as sample individuals age and either quit smoking or leave the sample through death.

In the structural model below, a chronic condition is defined as an absorbing, dichotomous state.[13] The conditions that determine a chronic health state stem from data availability on the timing of events and include a variety of cardiovascular diseases and types of cancers.[14] Because of data limitations, the dichotomous variable for chronic health used throughout does not capture all diseases that are caused by smoking. For example, the FHS data do not include panel data for chronic obstructive pulmonary disease (COPD). Given that COPD is the fourth leading cause of death in the United States (see http://www.nhlbi.nih.gov/health/dci/Diseases/Copd), the omission of COPD in the chronic health indicator may understate the behavioral response to chronic health shocks. Table 3 shows that by the seventh exam, conditional on survival, roughly 20 percent of the sample have experienced a chronic health shock.

[13] Individuals with multiple chronic health shocks are observed in the data. However, while the absorbing state assumption is made for computational convenience in the structural model, Khwaja et al. (2006) make the same assumption and argue that the first health shock is the most informative given the severity of health shocks under consideration.

[14] Appendix table A1 lists all chronic conditions considered.

TABLE 3
BIOMARKERS BY EXAM

Exam	Total Cholesterol	HDL Cholesterol	SBP	Diabetes	CVD Risk Point Score	Cumulative Chronic Health
1	196.802	50.672	122.38	.018	5.758	.002
2	203.288	48.533	122.38	.028	6.322	.034
3	211.475	51.063	123.586	.038	6.703	.062
4	206.848	49.883	126.525	.052	7.156	.088
5	204.797	49.872	126.205	.072	7.229	.114
6	205.094	51.019	127.951	.100	7.605	.15
7	199.775	53.199	127.267	.121	7.435	.193

NOTE.—Entries are biomarker and chronic health means by exam. Diabetes represents the cumulative proportion of individuals with a confirmed diabetes diagnosis. Chronic health status is a binary variable for any record of cardiovascular disease or cancer.

This paper is on the smoking response to the revelation of cardiovascular biomarkers; I focus on biomarkers identified by D'Agostino et al. (2008) as significant predictors for an individual's 10-year risk for general cardiovascular disease (CVD). Indeed, those authors, using FHS data, derive a calculator that reports a person's CVD risk as a function of individual biomarker and demographic characteristics. Ten-year CVD risk is a function of systolic blood pressure (SBP), total cholesterol, HDL cholesterol, an indicator for diabetes, and an indicator for cigarette smoking. Table 3 presents summary statistics of the biomarker values by exam.

The calculator of D'Agostino et al. (2008) assigns points on the basis of the range in which each biomarker falls (e.g., a 50-year-old man with total cholesterol between 240 and 279 is assigned 3 points for the total cholesterol category). The points are summed over biomarkers and the sum maps to an overall 10-year probability of CVD. A higher point total implies a greater 10-year risk. Because the CVD calculator was developed on the basis of data from the FHS Offspring Cohort, the resulting index value nicely proxies for cardiovascular health. In this paper, I recreate the biomarker point total for 10-year CVD risk of D'Agostino et al., and I use the point total to summarize biomarkers.[15] Thus, an increase in the point total between exams implies an increase in CVD risk. I shift the distribution of risk points such that the minimum possible value is 0—consistent with a <1 percent 10-year risk of CVD. Table 3 reports the mean point level by exam. CVD risk points increase over the sample period, but because of selective mortality, the increase is nonmonotonic.

Evidence from the FHS.—To investigate the role of biomarker and other health information on cigarette smoking, I estimate reduced-form, myo-

[15] See tables 5 and 7 of D'Agostino et al. (2008) for the specific biomarker ranges and the associated points. The calculator assigns baseline points for age and smoking status. I omit points directly from these because the empirical models below explicitly control or model these factors. Appendix table A2 provides details on the construction of the point total.

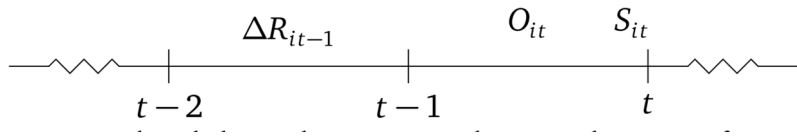

FIG. 2.—Timing of myopic smoking model. Smoking behavior between periods $t-1$ and t, S_{it}, is a function of, among other things, the most recent, observed change in biomarkers $\Delta R_{it-1} = R_{it-1} - R_{it-2}$ and the occurrence of any chronic health shocks between $t-1$ and t.

pic models of smoking behavior as a function of lagged changes in health. The modeling choices in this section are made for the sake of comparison to two sources. First, the limited economic literature on smoking responses to changes in health has estimated myopic models. For example, Clark and Etilé (2002) formulate a cigarette demand model as a function of changes in health, but they omit lead cigarette consumption—as implied by Becker et al. (1994)—due to weak instruments. Second, the reduced-form, myopic models in this section serve as a source of comparison to the dynamic structural model presented in Section III.

Because the mean time between health exams is 4 years and because the smoking questions across exams change, I model smoking behavior in exam t as a function of changes in biomarker values between exams $t-2$ and $t-1$. Thus, this section reports results from exams 3–7, in which an individual was asked if she "smoked cigarettes regularly in the last year." Figure 2 shows the time structure of the data. If individuals respond to changes in biomarkers, then the data should show changes in smoking behavior for changes in lagged and twice-lagged biomarkers. Because the data contain specific dates for chronic health shocks, smoking behavior may be responsive to shocks between periods $t-1$ and t.

Following figure 2, I estimate smoking models of the following form:

$$S_{it} = \alpha_0 + \sum_{j=1}^{5} \alpha_j \Delta R_{ijt-1} + \alpha_6 O_{it} + \alpha_7 X_{it} + \gamma_i + \delta_t + \epsilon_{it}, \qquad (1)$$

where S_{it} is an indicator of smoking behavior in the year leading to exam t; ΔR_{it-1} represents the change in biomarker j between exams $t-2$ and $t-1$; O_{it} is an indicator for a chronic health shock between $t-1$ and t; X_{it} are demographic characteristics at t; γ_i and δ_t are individual and time unobserved factors, respectively; and ϵ_{it} is an independent and identically distributed (i.i.d.) error term.[16] Table 4 presents the results from several specifications of equation (1) for the smoking participation decision. All specifications in table 4 include a lagged dependent variable

[16] Chronic health shocks are conditional on surviving at least until the next FHS exam. In some specifications, $\sum_{j=1}^{5} \alpha_j \Delta R_{ijt-1}$ is replaced with the point total for 10-year CVD risk.

TABLE 4
SMOKING REGRESSION RESULTS

	SMOKING AT EXAM t (Mean = .182)			
	(1)	(2)	(3)	(4)
Δ in total cholesterol between $t-2$ and $t-1$	.0001	.0004**	.0001	.0002
	(.0001)	(.0002)	(.0001)	(.0002)
× Female		.0000		.0001
		(.0002)		(.0001)
× Age > 36 at first exam		−.0002		.0000
		(.0002)		(.0001)
× College		−.0004**		−.0002
		(.0002)		(.0001)
Δ in HDL between $t-2$ and $t-1$	−.0004*	−.0004**	−.0003	−.0002
	(.0003)	(.0007)	(.0002)	(.0005)
× Female		−.0011**		−.0008*
		(.0005)		(.0004)
× Age > 36 at first exam		.0006		.0005
		(.0005)		(.0004)
× College		.0008		.0003
		(.0005)		(.0004)
Δ in SBP between $t-2$ and $t-1$	−.0002	.0001	−.0002*	.0000
	(.0002)	(.0004)	(.0001)	(.0003)
× Female		.0007**		.0005*
		(.0003)		(.0002
× Age > 36 at first exam		−.0007**		−.0005*
		(.0003)		(.0003)
× College		−.0004)		−.0003)
		(.0003)		(.0002)
Diabetes diagnosis between $t-2$ and $t-1$	−.0238	−.0275	−.0222	−.0173
	(.0156	(.0364)	(.0150)	(.0348)
× Female		.0320		.0009
		(.0312)		(.0302)
× Age > 36 at first exam		−.0267		−.0049
		(.0337)		(.0337)
× College		.0174		−.0014
		(.0318)		(.0300)
Chronic health stock between $t-1$ and t	−.0468**	−.0374	−.0378**	−.0326
	(.0131)	(.0312)	(.0116)	(.0317)
× Female		−.0709**		−.0778**
		(.0272)		(.0236)
× Age > 36 at first exam		.0094		.0193
		(.0282)		(.0269)
× College		.0170		.0162
		(.0265)		(.0238)
Smoking in $t-1$	.6316**	.6315**	.2078**	.2083**
	(.0114)	(.0114)	(.0133)	(.0133)
Number of years smoking as of $t-1$	.0036**	.0036**	−.0322**	−.0322**
	(.0002)	(.0002)	(.0019)	(.0019)
Age	−.0030**	−.0031**	.0048	.0050
	(.0004)	(.0004)	(.0035)	(.0035)
Age > 36 at first exam	−.0040	−.0024	. . .	. . .
	(.0085)	(.0087)		
Female	.0205**	.0216**	. . .	. . .
	(.0048)	(.0050)		
At least some college	−.0167**	−.0167**	. . .	. . .
	(.0051)	(.0053)		

TABLE 4 (*Continued*)

	Smoking at Exam t (Mean = .182)			
	(1)	(2)	(3)	(4)
Constant	.1669**	.1685**	.4095**	.3996**
	(.0253)	(.0254)	(.1738)	(.1732)
Fixed effects	No	No	Yes	Yes

NOTE.—The table reports results from linear probability models of smoking between periods $t - 1$ and t. Biomarker changes are the difference between values at exams $t - 1$ and $t - 2$. All models include exam dummies. Clustered standard errors are in parentheses. Estimates are from exams 3–7, $n = 11{,}778$.

* $p < .1$.

** $p < .05$.

and controls for exam effects. While columns 1 and 2 of table 4 report the results from ordinary least squares (OLS) regressions that ignore the time-invariant component γ_i, columns 3 and 4 report results from models that include individual-level fixed effects.

The results in table 4 suggest that the smoking response to changes in health is sensitive to the specific health change, its severity, and the subgroup in question. Focusing on the individual fixed-effects regression results in columns 3 and 4, the onset of a chronic health shock significantly lowers the probability of smoking in all specifications. At the mean of smoking (0.182), the onset of a chronic condition reduces the probability of smoking by 20.8 percent. The results of biomarker changes are mixed. Columns 3 and 4 of table 4 suggest that changes in SBP may significantly decrease the probability of smoking. A one standard deviation increase in SBP implies a 2 percent decline in the probability of smoking. For other biomarkers the results are inconclusive.

For comparison to the structural model presented below, table 5 presents results of changes in the CVD point total on smoking behavior. Individual fixed-effects results in column 3 of table 5 suggest that an increase in the CVD point index—and thus a higher risk of CVD—implies a decrease in the probability of smoking, but this effect is not statistically significant. However, that finding is specific to certain groups; those with at least some college education are statistically less likely to smoke. A 1-point increase in the CVD risk point index implies a $[(0.0008 - 0.0041)/0.182] \times 100 = 1.81$ percent decline in smoking for those with at least some college education.

It is not clear in a myopic model why an individual would quit (or continue? or increase?) smoking because of new health information. Formulating effective policy to reduce smoking requires a better understanding of the mechanisms behind smoking cessation decisions. Perhaps smokers rationally update their beliefs on the health effects of smoking but are addicted to cigarettes; thus, there is little observed change in

TABLE 5
SMOKING REGRESSION RESULTS

	SMOKING AT EXAM t (Mean = .182)			
	(1)	(2)	(3)	(4)
Δ in CVD point total between and $t-2$ and $t-1$	.0002	.0047*	−.0006	.0008
	(.0010)	(.0026)	(.0008)	(.0022)
× Female		.0019		.0025
		(.0021)		(.0017)
× Old		−.0043**		−.0014
		(.0020)		(.0017)
× College		−.0061**		−.0041**
		(.0020)		(.0017)
Chronic health stock between and $t-1$ and t	−.0479**	−.0416	−.0387**	−.0345
	(.0131)	(.0311)	(.0116)	(.0316)
× Female		−.0678**		−.0770**
		(.0272)		(.0236)
× Old		.0099		.0195
		(.0282)		(.0269)
× College		.0201		.0173
		(.0265)		(.0238)
Smoking in $t-1$	.6319**	.6318**	.2080**	.2082**
	(.0114)	(.0114)	(.0133)	(.0133)
Number of years smoking as of $t-1$	.0036**	.0036**	−.0323**	−.0322**
	(.0002)	(.0002)	(.0019)	(.0019)
Age	−.0031**	−.0031**	.0045	− .0047
	(.0004)	(.0004)	(.0035)	(.0035)
Age > 36 at first exam	−.0041	−.0027	. . .	. . .
	(.0085)	(.0087)		
Female	.0208**	.0225**	. . .	. . .
	(.0048)	(.0050)		
At least some college	−.0165**	−.0148**	. . .	. . .
	(.0051)	(.0053)		
Constant	.1677**	.1662**	.4244**	.4144**
	(.0253)	(.0254)	(.1737)	(.1731)
Fixed effects	No	No	Yes	Yes

NOTE.—The table reports results from linear probability models of smoking between periods $t-1$ and t. The biomarker point total change is the difference between the value at exams $t-1$ and $t-2$. All models include exam dummies. Clustered standard errors are in parentheses. Estimates are from exams 3–7, $n = 11{,}778$.

* $p < .1$.

** $p < .05$.

smoking behavior. Or, perhaps individuals are forward-looking with respect to biomarkers, health more generally, and mortality, but the links between smoking and biomarkers, and between biomarkers and general health and mortality, are not sufficiently strong to induce a change. Finally, regardless of whether smokers are forward-looking with respect to their health, they may simply fail to update their beliefs when presented with new information. If smokers trade off current enjoyment of smoking for increases in future health complications, then myopic models will clearly miss important behavioral channels. Furthermore, the model described above misses potentially important heterogeneity in the effect of

smoking on health, and it ignores the potential for selective mortality. In the following section, I place more structure on smoking behavior such that the different assumptions and mechanisms are made explicit.

III. Structural Model of Cigarette Consumption

In this section, I formulate a dynamic model over the life cycle in which an individual makes the smoking decision in each year (age) t and is forward-looking with respect to health.[17] The goal of the model is to capture potential mechanisms through which a smoker may (or may not) respond to personalized biomarker information. In the model, an individual undergoes repeated health exams; conditional on smoking, she learns about the idiosyncratic effect of smoking on her biomarkers. In that way, the model incorporates potentially important heterogeneity. The key trade-off is between the current (potential for) enjoyment of smoking and its uncertain, future health consequences. The biomarkers are a potentially important source of information because, along with other factors, they influence chronic health and mortality probabilities. To evaluate the potential for biomarker information to induce smoking cessation, the learning channel of the model thus works through the technology of biomarker production.

Formally, let the decision for individual i in year t be given by $d_{it} = d$, where smoking alternative d is

$$d = \begin{cases} 0 & \text{Do not smoke} \\ 1 & \text{Smoke} \leq 1 \text{ pack/day} \\ 2 & \text{Smoke} > 1 \text{ pack/day.} \end{cases}$$

The set of factors that influence individual i's smoking decision in period t are given by the state space S_{it}. Define S_{it} as follows:

$$S_{it} = \{A_{it-1}, R_{it-1}, \tau_{it-1}, \psi_{it-1}, H_{it}, X_{it}\},$$

where A_{it-1} is individual i's stock of addictive capital entering period t; R_{it-1} is an index of her biomarkers; τ_{it-1} and ψ_{it-1} are the mean and variance, respectively, of her prior belief distribution; H_{it} is her chronic health status; and X_{it} is her set of demographic characteristics.[18] Additionally influencing behavior, but not listed here, are a preference error

[17] This section outlines the basic theoretical model. Given the limitations of the data, changes to the model in the empirical implementation are discussed in Sec. IV. Furthermore, the Appendix provides derivations and details of my solution method.

[18] Rather than model the evolution of numerous biomarkers, R_{it} is the CVD risk point total. See the Appendix.

ϵ_{it} and a permanent heterogeneity term μ that are both assumed to be known to the individual but unobserved by the econometrician. Permanent unobserved heterogeneity is a random effect that enters each equation and is shifted by an equation-specific factor loading, ρ. Assumptions about these error terms that aid estimation are discussed in Section IV.

At the beginning of representative period t, an individual makes her smoking decision based on S_{it}. After making the smoking decision, an individual undergoes her period t health exam and observes her biomarker index R_t. The observed biomarker index provides a signal of information with which she updates her beliefs on her idiosyncratic effect of both past and present smoking on her biomarkers. Biomarkers, smoking, and other factors then enter the chronic health and mortality transition equations. The period ends with the realization of any chronic health and/or mortality shocks.

A. *Preferences*

Following the standard expected utility framework, the deterministic portion of per-period utility associated with chronic health state h ($h = 0, 1$) and smoking alternative d_{it} is

$$\begin{aligned}\bar{U}_{it}^{h}(A_{it-1}, d_{it}, X_{it}, \mu) &= \alpha_{0h} + (\alpha_{1h} + \alpha_{2h}A_{it-1} + \alpha_{3h}\text{Age}_{it} + \alpha_{4h}F_i + \alpha_{5h}C_i) \times \mathbf{1}[d_{it} = 1] \\ &\quad + (\alpha_{6h} + \alpha_{7h}A_{it-1} + \alpha_{8h}\text{Age}_{it} + \alpha_{9h}F_i + \alpha_{10h}C_i) \times \mathbf{1}[d_{it} = 2] \\ &\quad + \alpha_{11h} \times \mathbf{1}[d_{it-1} \neq 0] \times \mathbf{1}[d_{it} = 0] + \alpha_{12h}A_{it-1} + \rho^{Uhd}\mu.\end{aligned} \tag{2}$$

The specification accommodates nonlinearity in the effects of light and heavy smoking on utility. While $\alpha_{1\cdot}$ ($\alpha_{6\cdot}$) is the direct marginal utility of light (heavy) smoking, $\alpha_{2\cdot}$ ($\alpha_{7\cdot}$) captures the extent to which past consumption reinforces current consumption; $\alpha_{2\cdot}$ ($\alpha_{7\cdot}$) captures a part of the intertemporal trade-off in utilities. Note that higher values of A_{it} imply higher stock of addictive capital; $\alpha_{3\cdot}$ ($\alpha_{8\cdot}$) captures changes in the marginal utility of smoking across the lifespan; F is a binary variable that equals one if the person is female; and C is a binary variable that equals one if the individual has completed at least some college.[19] Specific withdrawal costs from cessation, which also capture part of the intertemporal utility trade-off, are captured by $\alpha_{11\cdot}$; $\alpha_{12\cdot}$ captures tolerance in smoking. That is, the extent to which a given level of stock affects utility is captured

[19] Because education is not being modeled, I assume that C_i reflects whether an individual's highest level of schooling was some college or greater.

here regardless of smoking behavior. Finally, permanent unobserved heterogeneity, μ, and its chronic health state and choice-specific factor loading ρ^{Uhd} shift utility.[20]

Relative preferences over smoking alternatives hinge on two main factors. First, preferences vary by the chronic health state ($H_{it} = h$). The extent to which the marginal utility of smoking varies across chronic health states remains an open question. Generally, the marginal utility of consumption of any normal good is thought to be lower in worse health states (Viscusi and Evans 1990; Gilleskie 1998; Finkelstein, Luttmer, and Notowidigdo 2013). However, if smoking provides relaxation and comfort when stricken with a chronic illness, the overall marginal utility of smoking may be larger in worse health states. Estimation of the structural parameters will therefore empirically test for the sign of the marginal utility of smoking across health states. Second, as seen below, current-period smoking affects the size of the next-period smoking stock, which in turn affects the next-period biomarker index and next-period utility. Given the dynamic nature of the model, individuals evaluate smoking alternatives while considering the future marginal utility of smoking as well as the future consequences of a higher A_{it}.[21]

B. Smoking Stock

Following the rational addiction literature, define A_{it} as the accumulated addictive stock of cigarette smoking capital. I assume that each individual has an addictive stock of zero at age 7 ($A_{i7} = 0$ for all i). Individual i's time $t - 1$ addictive stock remains zero if she has not smoked in any previous period. The technology of the stock of addictive capital is as follows:

[20] I have considered alternative specifications in which decade dummies shift the marginal utility of each smoking alternative to capture the changes in national trends and attitudes toward cigarette smoking. However, such a specification assumes that individuals have perfect foresight on future attitudes toward smoking. Also considered were nonlinear age terms to capture youth, middle-age, and elderly marginal utility shifts; however, these thresholds would be arbitrarily chosen and the computational problem with adding additional parameters is large.

[21] An important omission from the utility function is a measure of general consumption, along with the usual budget constraint. This omission is due to data limitations. Unfortunately, the FHS does not include information on income, and as all sample individuals live in Framingham (or have moved and continue to participate in the study, which is unobserved), there is no cross-sectional variation in cigarette prices. Furthermore, fig. A1 shows that, prior to 1980, there was little temporal variation in average real cigarette prices in Massachusetts. Thus, average cigarette prices would enter the structural model's utility function as cohort effects, and, given that the model already allows the value function regression coefficients to vary by year of birth, adding prices to the utility function generated unstable parameter estimates. See the Appendix for further details.

$$A_{it-1} = \begin{cases} \exp\{\delta_1 \ln(A_{it-2}) + \delta_2 \mathbf{1}[d_{it-1} = 1] \\ \quad + \delta_3 \mathbf{1}[d_{it-1} = 2]\} - 1 & \text{if } \sum_{n=0}^{t-1} d_{in} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Conditional on any past smoking, the addictive stock entering period t, A_{it-1}, is specified as a function of the previous-period stock and the previous-period decision.[22] The term δ_1 can be interpreted as one minus the depreciation rate of the stock in percentage terms. The nonlinear investments of light and heavy smoking into the smoking stock are captured by δ_2 and δ_3, respectively.

C. *Biomarkers and Learning*

At each health exam, an individual is assumed to observe a summary of her biomarker results, R_{it}. Let the evolution of this biomarker index, R_{it}, be given by the following technology:

$$R_{it} = \zeta R_{it-1} + X_{it}\phi + \theta_i A_{it} + \rho^R \mu + \nu_{it}. \tag{4}$$

The technology relates the biomarker index, R_{it}, to its lag, R_{it-1}; a vector of sociodemographic characteristics of individual i, X_{it}; the stock of addictive capital, A_{it}, after making the period t smoking decision; and time-invariant and unobserved heterogeneity, μ, and its factor loading, ρ^R. The term ν_{it} is an i.i.d. shock over individuals and time with distribution $\nu_{it} \sim N(0, \sigma_\nu^2)$. To capture heterogeneity in the effect of the smoking stock on biomarkers, each individual is endowed with a time-invariant, unknown (to both the individual and the econometrician) match value, θ_i, where θ_i is a random coefficient about which individuals learn via repeated biomarker signals, R_{it}.

Modeling learning over a technological parameter requires a number of assumptions. The assumption of rational expectations is ubiquitous in the literature on health transitions and dynamic discrete-choice models.[23] The rational expectations assumption in this context implies that an individual's beliefs about future states correspond with the actual transition probabilities; thus, an individual knows the population mean parameters that dictate technologies 3 and 4 and the distribution of any

[22] In the first period of smoking, an individual's stock evolves to either $\exp(\delta_2) - 1$ or $\exp(\delta_3) - 1$, depending on the intensity of smoking.

[23] See Aguirregabiria and Mira (2010) for a discussion of expectations in structural models, and see Gilleskie (1998), Arcidiacono et al. (2007), and Khwaja (2010) for specific examples. See Manski (2004) for a discussion of relaxing the rational expectations assumption with subjective expectations data.

shocks (i.e., ν_{it}). Rational expectations and learning imply that individual i's initial prior belief on θ_i corresponds to the population mean of the θ distribution.[24] Because the individual understands technology 4, and the distributions of θ_i and ν_{it}, and because θ_i is assumed to be time invariant,[25] R_{it} serves as a noisy signal from which an individual may update her prior on θ_i.

While the signal of information is assumed to be normally distributed, I assume that θ_i is drawn from a normal distribution that is truncated at zero: $\theta_i \sim TN(\bar{\theta}, \sigma_\theta^2, 0)$. The assumptions on the distributions of ν and θ_i imply that the Bayesian posterior belief distribution is also a normal distribution that is truncated at zero and that the belief distribution is sufficiently characterized by the truncation point and the posterior mean and variance. The assumption of a truncation point at zero imposes that an individual never believes that increasing her smoking stock, A_{it}, by smoking will improve her biomarkers. Without the truncation assumption, depending on $\bar{\theta}$ and σ_θ, some fraction of the θ distribution is guaranteed to be negative; furthermore, the truncated distribution preserves the conjugate distribution updating that yields closed-form solutions for the posterior mean and variance.[26]

Let individual i's period t prior belief on θ_i be characterized by τ_{it-1}, her prior mean, and ψ_{it-1}, her prior variance. If individual i has never smoked, then $A_{it-1} = 0$ and her prior belief is characterized by her initial priors (i.e., $\tau_{it-1} = \bar{\theta}$ and $\psi_{it-1} = \sigma_\theta^2$). Individual i uses her prior belief on θ_i to formulate her expected biomarker index, R_{it}, conditional on her smoking choice. After making the smoking choice and observing R_{it}, assuming that $A_{it} > 0$, individual i derives her posterior beliefs. This individual has two fundamental sources of information: her prior beliefs $(\tau_{it-1}, \psi_{it-1})$ and the observed results from her health exam in period t, R_{it}. Appealing to the assumption of conjugate prior and signal distributions, the period t beliefs have closed-form solutions that are given via Bayes's rule. The posterior mean and variance are[27]

$$\tau_{it} = E(\theta_i | R_{it}, A_{it}, \tau_{it-1}, \psi_{it-1}) = \frac{A_{it}^2 \psi_{it}}{\sigma_\nu^2} \hat{\theta}_{it} + \frac{\psi_{it}}{\psi_{it-1}} \tau_{it-1}, \tag{5}$$

[24] In the absence of subjective expectation data on beliefs, this assumption is standard. See Crawford and Shum (2005), Chan and Hamilton (2006), and Mira (2007) for examples.

[25] This assumption is made both for computational convenience and such that learning may take place. Without ν_{it}, an individual would perfectly learn his or her match value θ_i at the first health exam (i.e., the first realization of R_{it}). If both θ and ν were time varying, further assumptions would be needed such that an individual could disentangle a shock to θ from a shock to ν.

[26] At the estimated parameters, only 0.7 percent of the estimated distribution of θ is truncated. A previous version of this paper did not impose a truncation point at zero, and the qualitative conclusions of the paper are not sensitive to this assumption. Results are available on request.

[27] Derivations of these equations can be found in the Appendix.

$$\psi_{it} = \text{Var}(\theta_i | A_{it}, \psi_{it-1}, \sigma_\nu) = \frac{\psi_{it-1}\sigma_\nu^2}{A_{it}^2\psi_{it-1} + \sigma_\nu^2}. \quad (6)$$

Here, $\hat{\theta}_{it}$ summarizes the within-individual variation through the tth health exam. Note that these beliefs have the following appealing properties. First, the posterior mean is a weighted average of $\hat{\theta}_{it}$ and the prior mean τ_{it-1}. Second, the weight placed on the period t signal (i.e., $\hat{\theta}_{it}$) is increasing in A_{it}. That is, a higher stock of addictive capital implies a larger weight placed on the health exam results. Finally, the posterior moments of an individual for whom the stock equals zero (i.e., $A_{it} = 0$) collapse to the prior moments. When making the smoking decision, an individual uses the mean and variance in equations (5) and (6) to integrate over the positive support of the belief distribution.

D. *Chronic Health*

Let H_{it} represent an individual's overall health state defined by the presence of any chronic conditions. Let $H_{it} = h$, where outcome h is as follows:

$$h = \begin{cases} 1 & \text{if chronic condition} \\ 0 & \text{if no chronic condition.} \end{cases}$$

What differentiates H_{it} and R_{it} is "reversibility." While R_{it} changes each period, I assume that upon diagnosis of a chronic condition, an individual has the condition forever.[28] Let the probabilities of transiting to different chronic health states in period $t + 1$ be

$$\pi_{it+1}^{h=1} = \begin{cases} P(H_{it+1} = 1 | S_{it}, d_{it}, \mu) & \text{if } H_{it} = 0 \\ 1 & \text{if } H_{it} = 1. \end{cases}$$

Define the relevant probability as

$$P(H_{it+1} = 1 | S_{it}, d_{it}, \mu) = \frac{\exp(\lambda Y_{it})}{1 + \exp(\lambda Y_{it})},$$

where Y_{it} includes state variables, interactions, higher-order terms, and unobserved heterogeneity controls. Furthermore, Y_{it} includes decade dummy interactions with R_{it} to capture changes over time in how biomarkers affect the probability of chronic disease incidence (perhaps because of advances in medical technology, pharmaceuticals, etc.). In forecasting future chronic health transitions, I follow the literature referenced above

[28] This assumption captures the fact that upon having a heart attack, e.g., an individual is in a fundamentally different health state even if he does not have repeated heart attacks (Khwaja et al. 2006).

and assume that an individual understands the technology associated with the chronic health transition probability.

A natural question becomes, why do individuals in the model learn about how smoking affects biomarkers but not chronic conditions? In fact, learning is not specifically required to evaluate the effects of changes in health on smoking behavior. However, the purpose of this paper is to explore the importance of health information prior to major health shocks. Imposing that individuals understand the technology (i.e., the λ's) associated with covariates in the chronic health transition equation is the standard approach. Modeling learning enriches the standard model by incorporating heterogeneity in an important parameter. Furthermore, by modeling learning about the effect of A_{it} on R_{it}, individuals are indirectly updating their expectations about future chronic health transitions because the biomarker index enters the chronic health transition probability.

E. Mortality

For simplicity, I assume a finite horizon in the sense that, while an individual may die prior to period T, the probability of death equals one in period T. The utility of death is normalized to zero. Define an indicator for death at the end of period t, $M_{it+1} = 1$, and let its corresponding probability be given by

$$\varsigma_{it+1} = P(M_{it+1} = 1 | H_{it+1}, S_{it}, d_{it}, \mu) = \frac{\exp(\omega B_{it})}{1 + \exp(\omega B_{it})}.$$

Here, H_{it+1} is individual i's chronic health state at the end of period t.[29] Similar to Y_{it} above, B_{it} includes state interactions and higher-order terms, as well as permanent unobserved heterogeneity, μ, and decade dummy interactions with the chronic health state to capture change in medical technology.[30] Because the biomarker index enters the death transition equation directly (and indirectly through the chronic health term H_{it}), individuals are indirectly updating their expectations about death transitions conditional on their smoking choice through the learning process.

F. Summary

To summarize the structural model, consider the following: a smoker has never experienced a chronic health shock, but because of smoking, the

[29] The timing convention here is due to data aggregation. Clearly, any chronic health event occurring in period t must occur at or before the time of death, if death also occurs in t. Therefore, to accommodate the frequent observation in the data of an individual dying from a chronic health event, the appropriate chronic health data point in this equation is H_{it+1}.

[30] This assumes that individuals have perfect foresight that, given a chronic health shock, the probability of death will be different in the future.

value of her addictive stock of smoking capital is positive, $A_{it-1} > 0$. As she enters period t, her decision on whether to continue smoking (and at what intensity) depends on several factors. Given that A_{it-1} is fixed at the beginning of period t, the smoking decision affects her current utility through the direct marginal utility of smoking, α_{10} and α_{60}, and through the withdrawal from not smoking, α_{110}. Conditional on the smoking choice, she must forecast her future utility and health consequences. She understands that if she smokes, the addictive stock of smoking capital will increase ($A_{it} > A_{it-1}$) depending on the values of δ_1–δ_3, and her biomarker index R_{it} will be affected by A_{it} through her idiosyncratic effect, θ_i. She forecasts R_{it} via her belief distribution characterized by τ_{it-1} and ψ_{it-1} and her understanding of technology 4 and the distribution of ν. She also must forecast the probability of a chronic health shock. The current smoking decision influences the probability of a chronic health shock, both directly and through the updated value of R_{it}. The consequences of a chronic health shock, since it is assumed to be an absorbing state, include a potentially increased probability of death—which she must also forecast—and, conditional on survival to the next period, a different set of preference parameters, $\alpha_{\cdot 1}$.

IV. Empirical Implementation

To estimate the model described in Section III with FHS data described in Section II, there are three hurdles. First, while the data contain only seven exams over a 30-year period, the theoretical model is based on a yearly decision-making process over the life cycle. As explained below, retrospective questions in the FHS are used to construct a data set that mirrors the timing of the model. Given this expanded data set, the second hurdle is that state variables A_{it} and R_{it} must be constructed from the FHS data in such a way as to capture an individual's latent accumulated stock of smoking capital and her latent summary of biomarkers, respectively. A third hurdle is the identification of the model parameters. As described below, the structure of the model, the modeling of permanent unobserved heterogeneity, and plausibly exogenous variation in the timing of health exams across individuals together help to identify the model.

A. *Health Exam Timing*

Although individuals are observed at only seven health exams over a 30-year period, the theoretical model is based on an annual decision-making process. In the solution to the model, the final period, T, is specified to be age 100. That is, individuals are assumed to die with probability one at age 100. The yearly model is then solved recursively back to age 7, when all individuals are assumed have a smoking stock of zero (i.e., $A_{i7} = 0$

for all i). The data from Section II are expanded on the basis of answers to retrospective questions. Except for the biomarker information needed to construct R_{it}, there are data available to construct a yearly data set from age 7 until an individual either dies or completes his or her seventh exam. For any chronic health and mortality events, the data contain specific dates relative to the first exam.

Smoking behavior in the years before an individual's first exam is constructed on the basis of questions at the first and second exams. Answers to those questions reveal the first age at which one started smoking and the age at which one stopped smoking. While exams 3–7 ask whether an individual was smoking in the year prior to the exam, the first two exams ask whether an individual is "currently smoking." For the years between health exams, one can impute smoking data on the basis of individual history and adjacent health exam data. I assume that, during the interim years, agents believe with probability .25 that another FHS exam will occur. This value is chosen because the average time between exams is 4 years.[31] Further, I assume that if no exam occurs, individuals do not update their beliefs but rather use their prior beliefs to forecast biomarkers, conditional on the smoking choice. For "like smoking behavior" in adjacent exams, that behavior is assumed to continue in the interim. For "different smoking behavior" at adjacent exams, the smoking behavior is assumed to change at the midpoint between exams.[32] Measurement error from both smoking recall and imputation in interim exams, as well as the self-reported nature of the data, could potentially bias estimated smoking behavior. Indeed, transitions in smoking behavior over the life cycle are likely to be understated in the expanded FHS data. Therefore, as discussed below, the focus of the structural model is on predicting overall smoking rates over the life cycle rather than specific smoking transitions.

The expanded data set contains smoking information from age 7 until an individual's final FHS exam or death.[33] Figure 3 shows the sample probabilities for each smoking choice by age: smoking, either light or heavy, reaching its peak in the early 20s and then declining, through either cessation or death.[34]

[31] Findings are robust to different assumptions on this parameter.

[32] Health results in reduced-form simulations are not sensitive to the choice of imputation method. See the Appendix for details on the reduced-form model.

[33] There are 145,331 person/year observations in the expanded data set.

[34] This paper generally abstracts from cohort effects; however, given the variation in ages in the sample, and especially with respect to expectations over future health states, different birth cohorts may have different beliefs regarding the effects of smoking on health. The value function interpolation method used in solving the model for each sample individual does simulate the model for each age at the initial exam, so different birth cohorts are allowed to calculate future values differently. See the Appendix for further details.

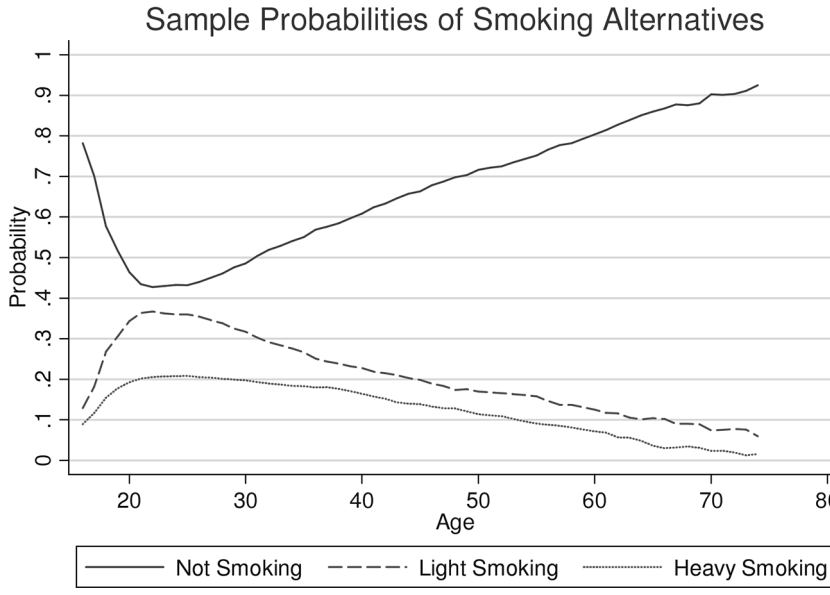

FIG. 3.—Smoking behavior by age: sample probabilities. The figure displays smoking alternative probabilities in the sample from ages 15 to 75.

B. Latent Factors

The point score derived in D'Agostino et al. (2008) is used to proxy for R_{it}, an individual's biomarker state variable in period t. The point score is an appropriate summary of biomarker results following a health exam because it (a) has a medically verified interpretation in predicting 10-year CVD risk and (b) places weights on the relative importance of different biomarkers in the production of CVD.[35]

The stock of addictive capital, A_{it}, which serves the intuitive role in Becker and Murphy (1988) of reinforcing future cigarette consumption, is a latent factor for which only proxies exist in data. Indeed, most empirical models of addiction proxy for A_{it} with the number of years of smoking, ignoring the lifetime pattern and intensity with which an individual has smoked. Furthermore, in the structural model presented in Section III, the state variable A_{it} serves two functions: reinforcing future consumption and directly influencing health.[36]

[35] A previous version of this paper used the first principal component from the biomarker correlation matrix to construct R_{it}. However, that index was difficult to interpret and the weights on each biomarker were not anchored in any outcome measure.

[36] Conceptually, we could think of two state variables: one that captures addiction through a biological dependence on nicotine and the other that captures health deterioration (e.g., an accumulation of tar in the lungs). Adda and Lechene (2001) model separate state variables for these two effects; however, those authors are unable to separately identify production technologies for the separate state variables. While restricting these

This paper employs the dynamic factor methods of Cunha and Heckman (2008) and Cunha et al. (2010) to model the latent factor, A_{it}, and to estimate technology 3. Those authors develop a dynamic factor model to estimate the technologies of cognitive and noncognitive skills in young children. Their factor model allows the data to weight the relative importance of a set of measurements, shown to be riddled with measurement error, that may proxy for cognitive and noncognitive skills. With the joint distribution of skills, the authors are able to identify parameters of an assumed skill evolution equation. In the current paper, and consistent with the timing of the structural model, four measurements of smoking history through period $t - 1$ serve as proxy variables for A_{it-1}: total number of years of smoking conditional on ever smoking, the number of years since last not smoking conditional on currently smoking (if one has quit and returned), the number of years since last smoking conditional on having quit (if one has quit and not returned), and the average intensity of individual smoking at any time conditional on ever smoking (average of light $d = 1$ and heavy $d = 2$ smoking). These endogenous measurements are available in the expanded data, and each may provide a noisy measure of both addiction and the health effects of smoking (e.g., a person with high average intensity of smoking is more likely to be addicted and to have significant health effects). Furthermore, each measurement captures a different aspect of smoking patterns through period $t - 1$.

Following the notation of Cunha et al. (2010), let measurement j for individual i in time t be given by the equation

$$Z_{ijt} = \mu_j + \alpha_{jt} A_{it} + \psi_{ijt}.$$

Up to the normalization that the loading, $\alpha_{jt} = 1$ for all t and j = total years of smoking, the loadings α_{jt} are identified from the variance-covariance matrix of the measurements.[37] This method assumes only separability of the error and that the error terms ψ_{ijt} are independent across measurements and time. Given the separability of the factor and the error, the ratio Z_{ijt}/α_{jt} is a noisy signal of the underlying factor A_{it}. Under these assumptions, the ψ's are classical measurement error terms. The likelihood function is augmented to include the density of each Z_j, and I estimate the mean μ_j and the standard deviation of ψ_j for each measurement.[38]

To estimate the parameters the stock evolution equation, the technology of the addictive stock of smoking capital is augmented with an error term, $\eta_{it} \sim N(0, \sigma_\eta)$:

---

state variables to be the same is a simplifying assumption, the correlation between nicotine dependence and health deterioration is likely high.

[37] The normalization of α_{jt} for all t and j = total years of smoking is without loss of generality. See Cunha et al. (2010, 890–94) for further details.

[38] Because each measurement is strictly positive conditional on past smoking, I assume that ψ_{ijt} is a lognormal random variable.

$$\ln(A_{it-1}) = \delta_1 \ln(A_{it-2}) + \delta_2 \mathbf{1}[d_{it-1} = 1] + \delta_3 \mathbf{1}[d_{it-1} = 2] + \rho_A \mu + \eta_{it-1}.$$

Given the interpretation of A_{it} as an accumulated addictive stock and the assumed technology, η helps to identify investment and depreciation parameters and may reflect individual-level uncertainty over the mean parameters. Furthermore, the unobserved heterogeneity term μ may reflect heterogeneity in the initial stock value after the first period of observed smoking.

C. *Identification*

Identification stems from exogenous variation in exam timing, model structure, modeling permanent unobserved heterogeneity, and the observed health and smoking transitions. The main source of exogenous variation in the data is the timing of health exams. The number of years between health exams does not directly affect health, after conditioning on observables, but observationally equivalent individuals with different time gaps between exams may select different smoking patterns. These patterns may arise because different time gaps induce variation in the belief distribution across individuals. Table 2 shows the distribution of time between exams by exam: at each exam, the standard deviation of the number of years since the previous exam is greater than 1 year.

To show the relationship between exam timing and smoking and health outcomes, table 6 presents estimates of the effect of the number of years between exams on different outcomes. In each model, controls include initial age upon entering the sample, age, education, chronic health state, and smoking history measurements.[39] Columns 1 and 2 report the results from a multinomial logit model of light and heavy smoking. Column 3 reports results from a linear probability model of a dummy variable for smoking. Column 4 reports the results from a linear regression of the biomarker index, R_{it}, on the conditioning variables listed above and the number of years between exams. The results suggest that the number of years between exams increases the probably of light smoking, and of smoking generally, but has no effect on the biomarker index.

Identification of the structural model also relies on structural assumptions; Keane (2010) notes that the advantage of the structural approach is that these assumptions are made explicit. Furthermore, Heckman and Navarro (2007), who study identification in dynamic discrete-choice models of time to treatment and the effect of treatment on outcomes, point out that curvature and cross-equation restrictions for smoking and health

[39] I focus on the number of years between exams because the number of exams completed is highly correlated with age. This reflects the classic problem of estimating age, cohort, and time effects.

TABLE 6
VARIATION IN EXAM TIMING

	Light Smoking (1)	Heavy Smoking (2)	Smoking (3)	Biomarkers (4)
Number of years since last exam	.087**	−.012	.014**	.000
	(.014)	(.018)	(.002)	(.014)
Age	−.051**	−.111**	−.010**	.077**
	(.005)	(.008)	(.001)	(.005)
Age at first exam	.025**	.075**	.006**	.068**
	(.007)	(.010)	(.001)	(.007)
Female	−.007**	−.636**	−.034**	−1.235**
	(.092)	(.126)	(.013)	(.108)
At least some college	−.467**	−.963**	−.103**	−.543**
	(.096)	(.130)	(.014)	(.113)
Chronic health state	−.011	−.085	−.002	.421**
	(.142)	(.193)	(.017)	(.162)
Constant	−.095	1.590**	.576**	1.429**
	(.241)	(.318)	(.036)	(.264)

NOTE.—Columns 1 and 2 report results from a multinomial logit model of light or heavy smoking relative to not smoking. Column 3 reports results from a linear probability model of smoking. Column 4 reports results from a regression of the biomarker index. All estimates are from exams 2–7, $n = 14{,}322$. Clustered standard errors are in parentheses.

* $p < .1$.

** $p < .05$.

together provide identification without standard exclusion restrictions. Identification of the structural model of smoking in part comes from the clearly delineated choice model with health measurements as outcomes. By structuring the model such that, within a given period, the smoking choice is made and the health consequences follow, the model restricts the pathways of causality to be consistent with economic models of choice. Counterfactual experiments presented in Section V compare the treatment effect of smoking on longevity generated by the structural model to "reduced-form" empirical models in which choice modeling is left implicit.

The specific structural parameters to be estimated include

$$\text{utility parameters:} \quad \Theta_U = \{\alpha_{0h}, \ldots, \alpha_{12h}\}_{h=0}^{1}$$

$$\text{health transition parameters:} \quad \Theta_H = \{\lambda_0, \ldots, \lambda_{11}\}$$

$$\text{death transition parameters:} \quad \Theta_M = \{\omega_0, \ldots, \omega_{14}\}$$

$$\text{smoking stock parameters:} \quad \Theta_A = \{\delta_1, \delta_2, \delta_3, \sigma_\eta\}$$

$$\text{learning and risk parameters:} \quad \Theta_R = \{\bar{\theta}, \sigma_\theta, \sigma_\nu, \phi, \zeta\}$$

$$\text{factor loadings:} \quad \Theta_\rho = \left\{ \left\{ \{\rho^{Uhd}\}_{h=0}^{1} \right\}_{d=0}^{2}, \rho^H, \rho^M, \rho^R, \rho^A \right\}$$

$$\text{measurement parameters:} \quad \Theta_\psi = \{\mu_j, \psi_j\}_{j=1}^{4}$$

$$\Theta = \{\Theta_U, \Theta_H, \Theta_M, \Theta_A, \Theta_R, \Theta_\rho, \Theta_\psi\}.$$

To identify the preference parameters, I normalize the utility of death to be zero and then consider variation in smoking behavior and health and death transitions over time. Different smoking choices across the latent smoking capital stock, age levels, gender, and education identify the preference parameters. The withdrawal parameter, $\alpha_{11\cdot}$, is identified through variation in individuals' choices after a period in which one quits. In other words, conditional on having smoked in the previous period, the reinforcement effects, $\alpha_{2\cdot}$ and $\alpha_{7\cdot}$, and the withdrawal effects, $\alpha_{11\cdot}$, encourage current-period smoking. However, withdrawal is identified separately from the reinforcement effects because even though the smoking stock variable depreciates at rate δ_1 after smoking cessation, the utility cost paid of withdrawal lasts for only one period. Finally, the direct impact of the smoking stock on utility, $\alpha_{12\cdot}$, reflects tolerance for smoking and is identified by individuals who progress from light to heavy smoking.

The main assumption that identifies the learning parameters is that θ_i follows a truncated normal distribution. While variation in the biomarker index, R_{it}, and the addictive capital stock, A_{it}, identify $\bar{\theta}$, mean of the distribution, within-individual variation in smoking helps to identify σ_θ. On identifying learning, first note that if there is no variation in θ across individuals such that $\sigma_\theta = 0$, then there cannot be learning under the assumption of rational expectations. Conditional on heterogeneity in θ across individuals, there still may not be evidence of learning if R_{it} and A_{it-1} are sufficient statistics for the current smoking choice. Evidence of learning over time could come from choices at the end of the time frame relative to the beginning, which should better reflect an individual's true match value, θ_i. Thus, A_{it-1} and R_{it} are not sufficient statistics for the current smoking choice: individuals with more reports of R_{it} have a smaller prior variance on θ_i.[40]

If agents learn from FHS health exams, then estimation of expectations and learning may be confounded by unobserved shocks to an individual's information set. For example, visiting a private physician between FHS exams would allow an individual to update her priors, and this updating would not be observed. If individuals learn about the effect of smoking on biomarkers outside of FHS exams, then the information value of FHS exams will be overstated, and the importance of biomarker information in general on smoking behavior will be understated. This possibility cannot be completely ruled out, but three factors suggest that the bias may be small. First, exogenous variation in the timing of FHS exams helps to identify the effect of changes in biomarkers as reported in

[40] Crawford and Shum (2005) make a similar argument on identifying learning within a structural model of pharmaceutical choice. Testing for learning is confounded by μ, the permanent unobserved heterogeneity term. I thank an anonymous referee for this intuition.

TABLE 7
LOGIT MODELS OF SMOKING BEHAVIOR WITH INTERIM DOCTOR VISITS

	Multinomial Logit Relative to $d_{it} = 0$				Logit Relative to $d_{it} = 0$	
Variable	Light Smoking	Heavy Smoking	Light Smoking	Heavy Smoking	Smoking	
Interim doctor checkup	−.08	−.47**	. . .	. . .	−.15	. . .
	(.16)	(.20)			(.16)	
Any interim interaction	. . .	. . .	−.24	−.25	. . .	−.23
			(.22)	(.27)		(.21)
Lagged years smoking	.06**	.01	.06**	.00	.05**	.05**
	(.01)	(.02)	(.01)	(.02)	(.01)	(.01)
Lagged years since not smoking	.22**	.26**	.22**	.26**	.23**	.23**
	(.04)	(.04)	(.04)	(.04)	(.04)	(.04)
Lagged years since last smoking	−.11**	−.41**	−.11**	−.41**	−.13**	−.13**
	(.02)	(.12)	(.02)	(.12)	(.02)	(.02)
Lagged mean intensity	.79**	3.59**	.80**	3.58**	1.12**	1.13**
	(.20)	(.27)	(.20)	(.26)	(.20)	(.20)
Health state	−.27	−.75	−.26	−.79	−.30**	−.30
	(.33)	(.41)	(.33)	(.41)	(.32)	(.32)
Age	−.09**	−.11**	−.09**	−.11**	−.09**	−.09**
	(.01)	(.02)	(.01)	(.02)	(.01)	(.01)
Female	.17	−.26**	.17	−.32	.13	.12
	(.18)	(.24)	(.18)	(.24)	(.18)	(.18)
At least some college	−.13	−.62**	−.13	−.65**	−.22	−.22
	(.18)	(.23)	(.18)	(.23)	(.17)	(.18)
Constant	.31	−2.24**	.42	−2.12**	.40	.50
	(.54)	(.76)	(.56)	(.78)	(.52)	(.54)

NOTE.—The sample includes all smoking observations from exams 4–7; $n = 9{,}299$. General practitioner checkups increased from 67.7 percent between exams 3 and 4 to 91.4 percent between exams 6 and 7. Any physician interaction increased from 89.2 percent to 94.2 percent over the same period.

* $p < .1$.

** $p < .05$.

those exams on smoking behavior. Second, data are available on whether an individual visited a physician between FHS exams. FHS consistently recorded the extent to which an individual either visited a physician in the interim for a routine checkup or had any interaction with a physician that included hospitalizations, surgeries, checkups, and so forth. These data are available at FHS exams 4–7. Comparing routine checkup rates between the FHS and the Behavioral Risk Factor Surveillance Survey during the 1980s and 1990s for similar age cohorts suggests that FHS exams may have crowded out some private, routine physicals.[41] Finally, table 7 presents the results from reduced-form multinomial logit and logit models for smoking as a function of standard controls and dummy variables for

[41] Comparison analysis results are available on request.

the two types of physician interactions. The models are run on data only at exams 4–7 (i.e., not on the expanded data). All of the estimated coefficients on the physician interaction terms are negative, but only the coefficient on a physician checkup on heavy smoking is statistically significant.

Individuals are first observed at various points in their life cycle and with a variety of health histories; failing to properly model these histories would lead to an initial conditions problem (Heckman 1981). Furthermore, the initial conditions problem might lead to dynamic selection into smoking behaviors if individuals in some permanently lower (unobserved) health state select into smoking. Solving the model generates individual probabilities of choice behavior and health/death transitions for all ages beginning at 7 (recall that data exist for all smoking, chronic health, and death events from age 7 until either death or the final health exam). I assume that each individual has a smoking stock of zero and has no chronic health problems at age 7. The only remaining initial condition is the initial biomarker index upon entering the sample. With the model, I simulate a biomarker index for each period from age 7 until the first observed health exam. Hence, I use the model to generate probabilities of individuals' health histories when they are first observed in the sample (Khwaja 2006).[42]

Finally, permanent unobserved heterogeneity enters the model in a linear fashion through the μ term and the associated factor loadings, which allow for a different effect of the unobserved μ term everywhere it enters. Rather than placing a distributional assumption on the underlying unobserved heterogeneity, I approximate its distribution with a step function and estimate the factor loadings and mass point probabilities with other parameters in the model (Heckman and Singer 1984).[43] I estimate a model with three points of support in the unobserved heterogeneity distribution. The first and third points are normalized to be 0 and 1, respectively; I estimate the second point to be between 0 and 1. Conditional on these normalizations, the factor loadings are identified by the linearity assumption and covariance restrictions across equations. Additionally, I estimate the probability weights of the mass points for the discretized distribution of the permanent unobserved heterogeneity, μ.

[42] At age 7, individuals are assumed to be in the best categories of cholesterol and blood pressure and not to have diabetes. The model then fits an evolution of biomarker values that best fit the first exam value of R. The simulated biomarker index is scaled by demographic characteristics, X_{it}, as well as the unobserved heterogeneity, μ, term and its factor loading. Furthermore, individual variation in the data at the first exam (the initial condition) helps to identify parameters of the model.

[43] A nontrivial assumption is that θ_i is assumed to be uncorrelated with the permanent heterogeneity type.

Model parameters are estimated via simulated maximum likelihood. The likelihood function includes conditional smoking choice probabilities, densities of the smoking stock and biomarker index, chronic health and mortality probabilities, and measurement error densities in the factor model. The Appendix provides details on solution of the model, the likelihood function construction, and the maximization routine.

V. Results and Simulations

Table 8 reports the main parameter estimates and their corresponding asymptotic standard errors. The estimated utility constants, α_{00} and α_{01}, for the absence of a chronic health condition and a chronic health condition, respectively, are intuitive given that the utility of death has been normalized to zero. The total marginal utility of current-period light and heavy smoking is a function of $\alpha_{1\cdot}, \ldots, \alpha_{10\cdot}$. The results are consistent with adjacent complementarity as defined in Becker and Murphy (1988). In the absence of a chronic illness, heavy smoking is found to be reinforcing (i.e., $\alpha_{70} > 0$). Indeed, I find that heavy smoking is much more "reinforcing" than light smoking. The results also suggest that the marginal utility of light smoking in the absence of a chronic condition is invariant to education but significantly larger for women. The marginal utility of heavy smoking is lower for women when free of a chronic condition but decreasing in age only for those with a chronic condition. Withdrawal from smoking, (i.e., smoking in period $t - 1$ and not smoking in period t) is negative for those without a chronic condition but, perhaps reflecting the fact that smokers quit after major health shocks, positive for those with a chronic condition. For those without a chronic condition, both the withdrawal effect and the strong reinforcement effect drive smokers to continue smoking.

Several interesting trends emerge from these results. First, note that baseline marginal utility of both light and heavy consumption is negative. As suggested by the rational addiction literature, the model cannot explain why individuals start smoking. Consider that over 90 percent of smokers in the data start smoking before age 25 and no individuals in the data under the age of 25 have a chronic condition. The estimated preference parameters in the absence of a chronic illness suggest that, for a never smoker under the age of 25, there is no incentive to begin smoking because the marginal utility of smoking is negative. Furthermore, the dynamic considerations of the model suggest that smoking will increase the probability of future chronic illness and death through the smoking stock and the biomarker index. However, upon commencing smoking, the resulting positive smoking stock, along with the withdrawal effect, drive the dynamics forward. The estimates also predict a gradual transition from light smoking to heavy smoking. Competing effects for

TABLE 8
MAIN PARAMETER ESTIMATES

Description	Chronic Condition	Parameter	Estimate	Asymptotic Standard Error
Utility parameters:				
Constants	No	α_{00}	43.811	4.074
	Yes	α_{01}	1.391	.476
Consumption—light smoking:				
Constant	No	α_{10}	−4.638	.058
Consumption × smoking stock	No	α_{20}	.100	.016
Consumption × age	No	α_{30}	.000	.000
Consumption × female	No	α_{40}	.131	.023
Consumption × college	No	α_{50}	.000	.000
Consumption	Yes	α_{11}	−5.221	.146
Consumption × smoking stock	Yes	α_{21}	.375	.032
Consumption × age	Yes	α_{31}	.000	.000
Consumption × female	Yes	α_{41}	.000	.000
Consumption × college	Yes	α_{51}	−.396	.042
Consumption—heavy smoking:				
Constant	No	α_{60}	−25.770	.369
Consumption × smoking stock	No	α_{70}	.602	.232
Consumption × age	No	α_{80}	.000	.000
Consumption × female	No	α_{90}	−.232	.030
Consumption × college	No	α_{100}	.000	.000
Consumption	Yes	α_{61}	−4.281	.733
Consumption × smoking stock	Yes	α_{71}	.420	.063
Consumption × age	Yes	α_{81}	−.053	.015
Consumption × female	Yes	α_{91}	.033	.048
Consumption × college	Yes	α_{101}	−.463	.091
Withdrawal	No	α_{110}	−7.451	.073
	Yes	α_{111}	4.868	.086
Tolerance	No	α_{120}	.000	.000
	Yes	α_{121}	.530	.084
Learning parameters:				
Mean effect		$\bar{\theta}$	.021	.004
Standard deviation of θ_i		σ_θ	.008	.003
Standard deviation of ν		σ_ν	2.266	.010
Additional biomarker index parameters:				
Lagged health marker index		ζ	.907	.002
Age in years		ϕ_1	.009	.000
Female		ϕ_2	−.185	.006
College		ϕ_3	−.063	.005
Constant		ϕ_4	.413	.011

a new smoker include the reinforcement and withdrawal effects, which both encourage more smoking, and the increased probability of chronic disease and death, which encourage cessation. Finally, the estimates suggest a small degree of individual variation in the effect of the smoking stock on the biomarker index. The estimated standard deviation of θ, σ_θ, is small relative to its mean ($\bar{\theta}$).[44]

[44] Reduced-form estimates of biomarkers on the smoking measurements suggest that an estimate of $\bar{\theta} = 0.021$ is reasonable. These results are available on request.

Table 9 provides estimates of the parameters of technology 3 and the health transition equation parameters. These estimates are not marginal effects and are difficult to interpret because each outcome (biomarker index, chronic health, and death) is a complex function of entering period states and per-period decisions. In the simulation subsection below, I describe the results of simulations that isolate the effects of each variable on the system. However, a casual interpretation of the results in table 9 does yield some interesting insights. The parameter estimates of the smoking stock evolution equation indicate that an individual's stock of smoking depreciates by a small percentage each year, and the estimated magnitude of investment return in the smoking stock is greater for heavy compared to light smoking ($\delta_2 < \delta_3$).

TABLE 9
OTHER PARAMETER ESTIMATES

Description	Parameter	Estimate	Asymptotic Standard Error
Smoking stock parameters:			
Lagged stock	δ_1	.982	.000
Investment, light smoking	δ_2	.037	.000
Investment, heavy smoking	δ_2	.049	.000
Standard deviation of η	σ_η	.051	.000
Chronic health parameters:			
Constant	λ_0	−10.363	.177
Light smoking	λ_1	.213	.036
Heavy smoking	λ_2	.329	.045
Biomarker index	λ_3	.115	.015
Light smoking × biomarker index	λ_4	.004	.002
Heavy smoking × biomarker index	λ_5	.001	.002
Biomarker index squared	λ_6	−.005	.001
1980s × biomarker index	λ_7	.000	.000
1990s × biomarker index	λ_8	.000	.000
Age	λ_9	.088	.003
College	λ_{10}	.000	.000
Female	λ_{11}	.137	.019
Mortality parameters:			
Constant	ω_0	−14.241	.244
Light smoking	ω_1	.000	.000
Heavy smoking	ω_2	.000	.001
Biomarker index	ω_3	.097	.017
Light smoking × biomarker index	ω_4	−.052	.007
Heavy smoking × biomarker index	ω_5	−.006	.003
Biomarker index squared	ω_6	−.003	.001
Chronic health state	ω_7	1.484	.098
Light smoking × chronic health state	ω_8	1.453	.127
Heavy smoking × chronic health state	ω_9	1.645	.151
1980s × chronic health state	ω_{10}	−.000	.000
1990s × chronic health state	ω_{11}	−.000	.028
Age	ω_{12}	.127	.003
Female	ω_{13}	.000	.000
College	ω_{14}	−.315	.041

As noted above, the estimated mean effect of the smoking stock on the biomarkers is positive ($\bar{\theta} = 0.021$). A greater smoking history therefore implies a higher and, thus, worse biomarker index on average. According to table 9, a higher biomarker index implies a higher probability of chronic illness (through the positive sign on λ_3), albeit at a decreasing rate ($\lambda_6 < 0$), and death (through the positive signs on ω_3), albeit at a decreasing rate ($\omega_6 < 0$). While the chronic health and mortality transition equations allow for coarse technological progress controls, estimates of interaction effects between decade binary variables and health predictors are found to be both economically and statistically insignificant.

Table 10 reports measurement error and unobserved heterogeneity parameters. The estimated standard deviations of the yearly measurements that proxy for A_{it} suggest that measurement error may be significant. For example, the standard deviation of cessation—the number of years since an individual last smoked conditional on ever smoking—is 0.967 year.

The model is estimated with three points of support for the discretized unobserved heterogeneity distribution. Estimates in table 10 suggest that individuals located to the right of the heterogeneity distribution have a greater likelihood of experiencing both chronic health and mortality shocks. Notice that the factor loading on the marginal utility of heavy smoking while in the good chronic health state ($\rho_{\mu 02}$) is large and positive, suggesting that heavy smoking is defined by larger values in the unobserved heterogeneity distribution.

In addition to positive factor loadings that dictate the effect of the permanent unobserved term on the marginal utility of smoking, larger values of μ are also associated with higher rates of mortality ($\rho_M = 1.451$). This finding suggests that individuals who are more likely to smoke are also more likely to die independently of smoking. Furthermore, larger values of the unobserved heterogeneity term are associated with an increased risk of chronic health (as defined in this paper). These correlations could arise because of inherent differences (e.g., genetics) or because of differences in the propensity to engage in correlated risks that are not modeled.

A. *Model Fit*

Panels A–C of figure 4 summarize the relationship between the model's predicted probabilities and the mean sample smoking proportions by age. For each individual, observed smoking decisions and predicted smoking probabilities for periods up to either her final exam (exam 7) or death are calculated and averaged across individuals by age.[45] The model predic-

[45] Despite the fact that the model is solved from age 7 to 100, the figure presents results only for ages 20–75. Outside of the 20–75 age range, there are insufficient data for an informative comparison.

TABLE 10
MEASUREMENT AND UNOBSERVED HETEROGENEITY PARAMETERS

Description	Parameter	Estimate	Asymptotic Standard Error
Measurement parameters:			
Measurement means:			
Total years smoking	μ_1	1.126	.001
Years since last not smoked	μ_2	1.422	.002
Years since last smoked	μ_3	2.100	.003
Individual mean intensity	μ_4	.116	.001
Measurement error standard deviations:			
Measurement means:			
Total years smoking	σ_1	.360	.001
Years since last not smoked	σ_2	.647	.002
Years since last smoked	σ_3	.967	.003
Individual mean intensity	σ_4	.295	.001
Heterogeneity parameters:			
Utility: no chronic condition:			
Not smoking	ρ_{u00}	−4.854	.206
Light smoking	ρ_{u01}	−5.987	.203
Heavy smoking	ρ_{u02}	17.821	.536
Utility: chronic condition:			
Not smoking	ρ_{u10}	−.003	.015
Light smoking	ρ_{u11}	1.188	.070
Heavy smoking	ρ_{u12}	4.696	.312
Stock	ρ_A	.000	.000
Health marker index	ρ_R	.000	.000
Chronic health	ρ_H	.926	.058
Mortality	ρ_M	1.451	.085
Mass points and probabilities:			
Mass point 1	μ_1	.000	. . .
Mass point 2	μ_2	1.640	.021
Mass point 3	μ_3	1.000	. . .
Coefficient weight on mass point 1	θ_1	.414	.170
Coefficient weight on mass point 2	θ_2	−.563	.210
Miscellaneous parameters:			
Discount factor	β	.950	. . .
Log likelihood value	$L(\Theta)$	−224,665.160	

NOTE.—Mass points 1 and 3 are fixed at 0 and 1, respectively. Mass point 2 is estimated, and its location is $\exp(1.640)/[1 + \exp(1.640)] = .838$. The corresponding probabilities of mass points 1–3 are .491, .185, and .324.

tions generated from the solution routine fit the data well even at ages for which there are not many observations. Panel D of figure 4 compares the observed sample probabilities of chronic health with the predicted health probabilities, as generated by the model at the estimated parameter values. Panel D reflects both transitions to and surviving members of the chronic health state. Solution to the model yields a predicted probability of transiting to a chronic health state of one for individuals already in that state.

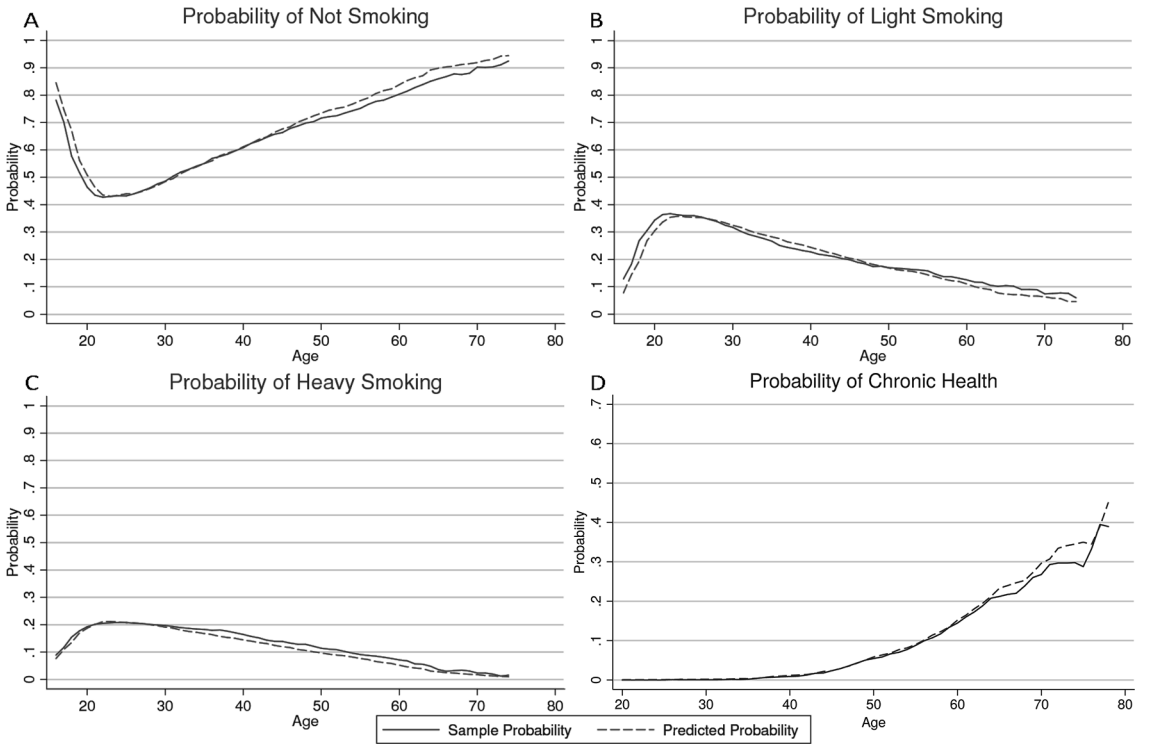

FIG. 4.—Smoking behavior by age: predicted and sample probabilities. Panels A–C display sample probabilities for each smoking alternative separately and with predicted probabilities generated from the solution routine at the parameters estimates reported in tables 8, 9, and 10. Panel D displays sample probabilities for chronic health incidence with predicted probabilities generated from the solution routine at the parameters estimates reported.

Table 11 reports sample and predicted smoking probabilities by health exam and health state. The average predicted choice probability across all health exams conditional on being in the chronic health state mirrors the observed probabilities in the data well. None of the goodness-of-fit tests are able to reject the null hypothesis that the predicted distributions are different from the observed distributions. Table 11 and figure 4 suggest that the model does a good job of predicting the overall smoking prevalence in the sample.

B. Counterfactual Simulations

By making explicit the behavioral channels through which a forward-looking individual makes the smoking decision, the structural approach allows for counterfactuals that capture how changes in information influence smoking behavior. Indeed, Heckman and Navarro (2007) note that because reduced-form analysis focuses solely on outcome equations,

TABLE 11
MODEL FIT: CHOICE PROBABILITIES

	Not Smoking		Light Smoking		Heavy Smoking	
Exam	Predicted	Observed	Predicted	Observed	Predicted	Observed
	Unconditional on Chronic Health State: 16,933 Person/Year Observations					
1	57.78	58.64	26.57	26.92	15.66	14.44
2	72.99	63.52	17.01	23.19	10.00	13.29
3	77.95	73.38	14.06	15.09	7.99	11.54
4	82.42	78.45	11.35	13.62	6.23	7.94
5	86.20	82.75	9.01	11.44	4.80	5.81
6	89.58	86.73	6.89	9.67	3.53	3.60
7	90.61	88.53	6.32	9.09	3.07	2.37
Mean	79.65	76.00	13.03	15.57	7.32	8.43
p-value	.700		.372		.598	
	Conditional on No Chronic Condition ($H_{it} = 0$): 15,432 Person/Year Observations					
1	57.70	58.63	26.62	26.94	15.69	14.43
2	72.09	63.53	17.59	23.04	10.32	13.43
3	76.58	73.33	14.96	15.18	8.47	11.48
4	80.84	78.39	12.40	13.85	6.77	7.76
5	84.55	82.58	10.10	11.63	5.35	5.79
6	87.89	86.36	8.02	9.72	4.08	3.93
7	88.56	87.85	7.72	9.60	3.73	2.55
Mean	78.31	75.81	13.91	15.71	7.77	8.48
p-value	.832		.678		.675	
	Conditional on Chronic Health ($H_{it} = 1$): 1,501 Person/Year Observations					
Mean	98.83	77.22	.50	14.11	.68	8.67

NOTE.—The table reports sample percentages in each smoking alternative and predicted smoking alternatives by health exam. *P*-values are from goodness-of-fit χ^2 tests of the null hypotheses that the predicted percentages are different from the observed percentages.

the analyst cannot evaluate the effect of changes in information on the agent's information set, expectations, decisions, or outcomes.

With the structural parameters in hand, the simulation exercise proceeds as follows. First, for each individual in the sample ($n = 2{,}611$), I replicate all observed characteristics at the initial exam 30 times (simulated $n = 2{,}611 \times 30 = 78{,}330$). To get each individual's value of the smoking stock at the initial exam, I use the observed history of smoking choices from age 7 until the initial exam, along with technology 3, and separate draws of μ_k and η_{ikt}, where k represents the kth draw out of 30.[46] For each of the 78,330 simulated individuals i, 150 sets of the unobservable variables are generated over the life cycle:[47]

$$\left\{\theta_{ik}, \left\{\nu_{ikt}, \eta_{ikt}, \left\{\epsilon_{iktd}\right\}_{d=0}^{2}\right\}_{t=7}^{100}\right\}_{k=1}^{150}.$$

Individual draws of θ are drawn from the estimated truncated normal distribution, and I integrate over the endogenous belief distribution of future biomarkers.[48] Smoking behavior and health outcomes are then simulated for each of the 78,330 observations from the initial age in which she is actually observed until death. Subsequent health exams are assumed to arrive at the mean time between exams in the data.[49]

C. *Information and Learning*

To evaluate potential mechanisms by which an individual may change her smoking behavior following the receipt of biomarker information, this section presents results from several counterfactual simulations.

To some extent, the model cannot answer the question of whether or not learning is taking place because, by placing the Bayesian structure on the process, the model imposes learning: with each signal of information (i.e., health exam), an individual's posterior variance is guaranteed to decrease conditional on a positive smoking history. More important, however, is whether the imposition of this structure changes predicted behavior. As a natural benchmark, I compare the predictions from simulation of the baseline model to results from a specification in which learning is removed. That is, individual prior beliefs are assumed to correspond to the estimated distribution of θ_i, and those beliefs remain con-

[46] Heterogeneity in the stock at the initial exam across each individual's 30 draws comes solely from μ_k and η_{ikt}.

[47] The simulated value of μ_k carries over from the initial condition.

[48] Truncated normal distribution draws are generated with Fortran code from Johnson, Kotz, and Balakrishnan (1994).

[49] Simulated individuals expect that an exam will arrive in the next period with a fixed probability of .25.

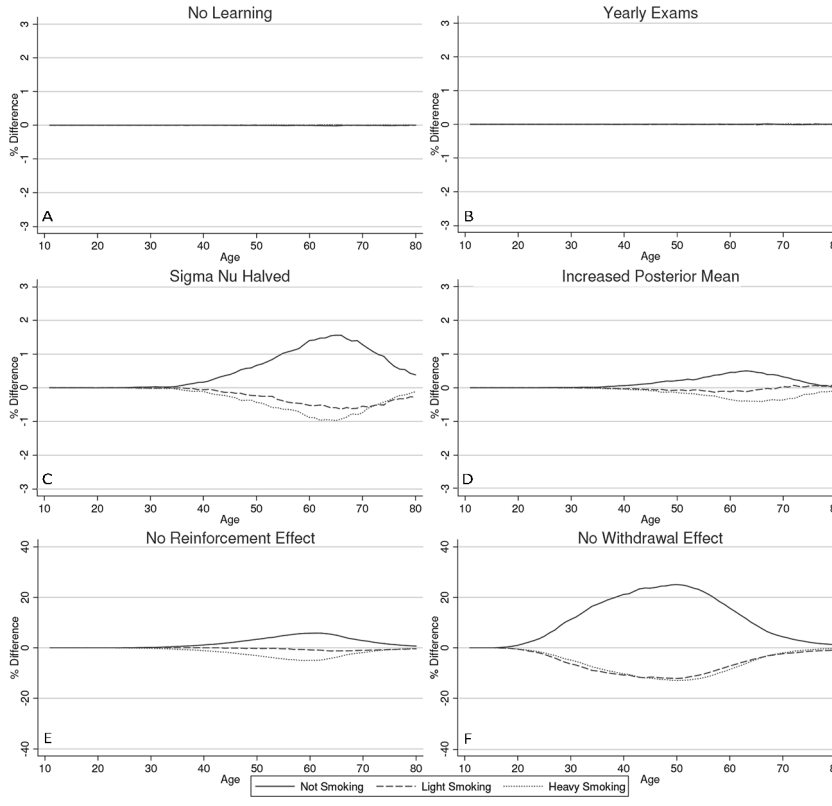

FIG. 5.—Information counterfactuals. Average difference in smoking probabilities, relative to baseline choices, by age and across different policy scenarios. Panel A assumes no learning. Panel B increases the frequency of health exams to be yearly. Panel C changes the value of $\sigma_v = 1.13$—half of its estimated value. Panel D assumes that individual posterior means are 20 times higher (worse). Panel E removes the reinforcement effect. Panel F assumes no withdrawal effect.

stant over the entire simulation.[50] Figure 5*A* presents the mean smoking percentages from the "no-learning" simulation relative to the baseline simulation. The simulation suggests that eliminating learning from the model does not change smoking behavior. Furthermore, figure 5*B* simulates the model assuming that, rather than receiving a signal of information roughly every 4 years, individuals observe their biomarkers annually. Again, the simulation finds a negligible change in smoking behavior over the entire life cycle.

[50] Removing learning effectively treats $\bar{\theta}$ like all other health transition parameters in the model; however, the simulation still allows for heterogeneity in θ_i. A similar simulation in which I remove learning and heterogeneity ($\theta_i = \bar{\theta}$ for all i) yields results that are almost identical to those that remove learning only.

One explanation for the lack of a simulated change in smoking behavior could be that biomarker information represents a very noisy signal of information. Recall that the signal of information at each health exam is muddled by a noise term that is large ($\sigma_\nu = 2.266$) relative to the distribution of θ_i. Indeed, figure 5*C* presents results from a simulation in which $\sigma_\nu = 1.133$—half of its estimated value. Interestingly, when the signal of information is less noisy, smoking declines by 2 percentage points at age 60. While still relatively small in magnitude, this result suggests that biomarker information may contain useful information regarding the health consequences of smoking, but the information may also be too inconsistent to influence smokers—even when the frequency of signals is increased as in figure 5*B*.

Another explanation could be that smokers are forward-looking and rationally update beliefs regarding θ_i, but biomarkers do not sufficiently predict poor health or mortality sufficiently to induce smoking cessation. To evaluate this possibility, I simulate the model assuming that individuals think that smoking is much worse for them than the estimates suggest. To do so, I multiply each person's simulated posterior mean by a factor of 20 in all periods. Importantly, this simulation does not change any of the technological parameters in health production. The results are presented relative to the baseline in figure 5*D*. Heavy smoking is simulated to decrease marginally in later life—up to 0.5 percentage points at age 65. While figure 5*C* suggests that biomarker information may be too noisy to induce smoking cessation, figure 5*D* suggests that individuals overestimating the effect of smoking on biomarkers may only slightly reduce smoking. More generally, figure 5*D* suggests that, relative to larger chronic health shocks, marginal biomarker changes may not be dramatic enough to change smoking behavior.[51] Two unreported simulations in which biomarkers are fixed throughout life at the 10th percentile (good health) and the 90th percentile generate a roughly 4-year difference in expected longevity, suggesting that an individual does have an incentive to manage her biomarkers; however, the small increase in biomarkers associated with an increase in the smoking stock is not found to significantly alter smoking behavior.

Finally, addiction mechanisms allowed by the model may explain the small magnitudes in figures 5*A* and 5*B*. An individual may rationally update her prior belief at every health exam, but the newly realized health benefits from smoking cessation may not outweigh the costs of quitting (or the high marginal utility of current smoking). To evaluate this possibility, I simulate the model assuming that the reinforcement effect is

[51] Figure 5*D* suggests that the implications of misspecification of beliefs, due to either functional form or unobserved information shocks, are unlikely to significantly change smoking behavior.

zero ($\alpha_{20} = \alpha_{70} = \alpha_{21} = \alpha_{71} = 0$) and, separately, the withdrawal effect is zero ($\alpha_{110} = \alpha_{111} = 0$). Results, again reported relative to the baseline model, are presented in figures 5*E* and 5*F*. In each case, smoking declines (note the larger vertical scale). While removing the reinforcement effect reduces heavy smoking by 6 percentage points at age 63—mainly through a reduction in heavy smoking—removing the withdrawal effect decreases overall smoking by over 25 percentage points at age 50. Given the magnitude of the reinforcement and withdrawal effects, it is not surprising that simulations that alter biomarker information fail to encourage smoking cessation.

D. Smoking and Selection

I use the estimated model to address how smoking affects the age of chronic health onset and death. Figure 6*A* displays, by age, the percentage of the simulated sample with a chronic condition while forcing individuals to (1) never smoke, (2) smoke lightly from age 18, and (3) smoke heavily from age 18.[52] Under these same forced behaviors, figure 6*B* shows, by age, the percentage of the simulated sample that remain alive.

The results in figure 6 confirm the findings in Sloan et al. (2003) that the detrimental effects of smoking occur largely after the age of 50. Indeed, the gap in the percentage of the sample in the chronic health state between never smokers and heavy smokers widens from less than 2 percentage points at age 50 to more than 10 percentage points at age 70. Similarly, while the difference in those surviving to age 50 between heavy and never smokers is negligible, that gap widens to 20 percentage points at age 70. These results are smaller than those in Doll et al. (2004). Those authors find a difference of approximately 28 percentage points at age 70 when considering never smokers and smokers. Importantly, while Doll et al. condition their results only on decade of birth and gender, I report results that are conditional on both observed and unobserved heterogeneity.

Given the relatively young age of the FHS Offspring Cohort, I observe death only in 394 individuals out of an estimation sample of 2,611, so I must rely on the model to generate chronic health and mortality results. To highlight the importance of incorporating unobserved heterogeneity and modeling smoking as a choice, table 12 reports the mean age of onset chronic health and mortality outcomes from simulation of the structural model and from simulation of a reduced-form outcome model.[53] Results from the reduced-form model in column 3 suggest differences

[52] Recall from the structural model that, upon transiting to a chronic health state, an individual remains in that state for life.

[53] See the Appendix for details on the reduced-form analogue to the structural model.

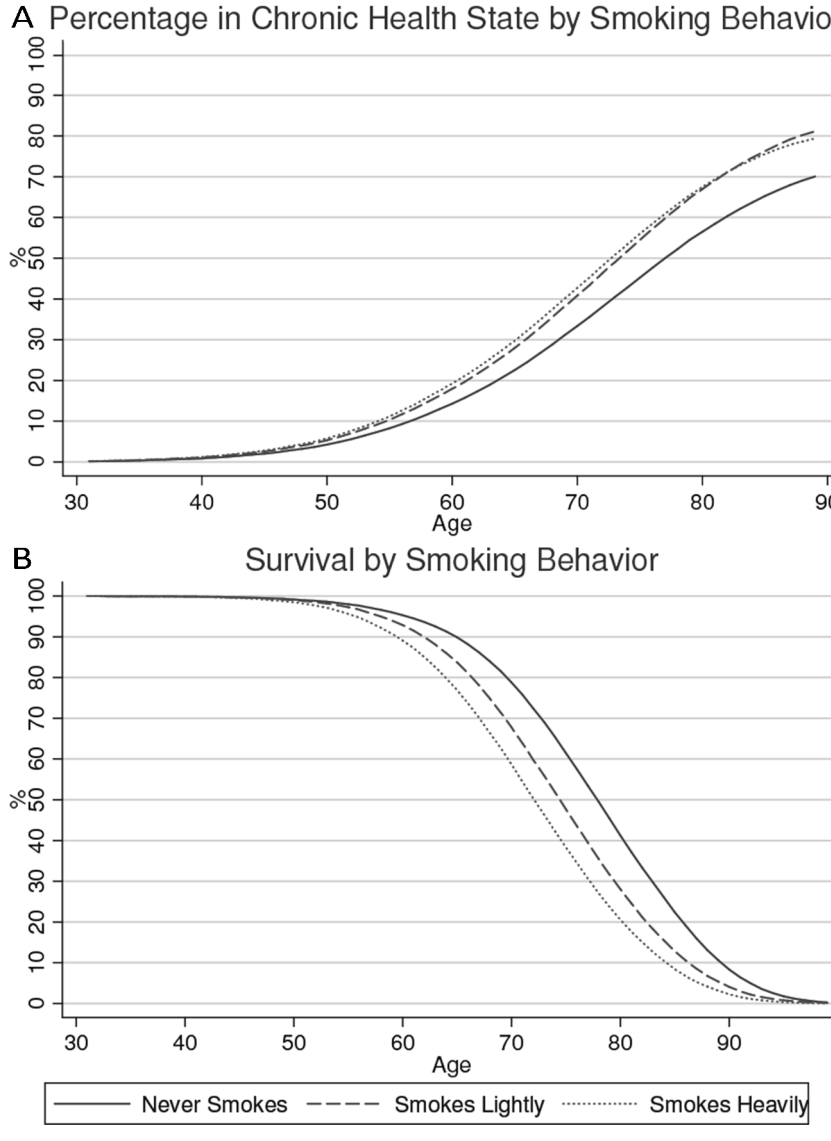

FIG. 6.—Simulated health outcomes. Percentage of simulated sample A in the chronic health state and B surviving, by age and smoking behavior. Individuals are simulated to never smoke, smoke lightly from age 18, and smoke heavily from age 18.

in mean age of chronic health onset of 1.55 and 3.73 years for continued light and heavy smoking relative to never smoking, respectively. The reduced-form simulation results in column 4 suggest differences in mean age of death of 3.77 and 6.56 years relative to never smoking, respectively.

TABLE 12
AGE OF CHRONIC HEALTH ONSET AND DEATH

	Structural		Reduced-Form	
Variable	Mean Age of H_{it} Shock (1)	Mean Age of Death (2)	Mean Age of H_{it} Shock (3)	Mean Age of Death (4)
Never smokes	70.13	77.80	71.69	83.24
	(12.27)	(9.68)	(18.37)	(12.30)
Smokes ≤ 1 pack/day from age 18	69.30	74.67	70.14	79.47
	(12.24)	(9.54)	(18.54)	(12.92)
Smokes > 1 pack/day from age 18	68.05	72.39	67.96	76.68
	(12.03)	(9.62)	(17.94)	(13.16)

Columns 1 and 2 of table 12 report the mean age of onset of a chronic health shock and mortality from simulation of the structural model that takes into account unobserved heterogeneity (μ_i). Here, light and heavy smoking reduce the mean age of chronic health onset by 0.83 and 2.05 years, respectively. Those numbers for mean age of death are 3.13 and 5.41. The smaller differences in means in the structural model relative to both the data and the reduced-form models reflect the positive correlation in the unobserved heterogeneity in the utility function and the chronic health and mortality equations.[54] Doll et al. (2004) report that smoking shortens the life span by 10 years, but in addition to ignoring both observed and unobserved heterogeneity with respect to smoking and health, those authors do not take into account intensity of smoking in their calculations. Figure 6 demonstrates that, conditional on smoking, the intensity with which one smokes is an important factor in explaining longevity. My results suggest that ignoring this heterogeneity may lead to an overstatement of the health effects of continued smoking.

VI. Discussion

This paper studies the role of personalized health information in the decision to smoke cigarettes for participants of the Framingham Heart Study—Offspring Cohort. Reduced-form, myopic models of smoking behavior suggest small, and generally statistically insignificant, effects of recent changes in biomarkers on smoking behavior. To investigate further, this study develops and estimates a dynamic stochastic model of smoking behavior that extends the classic rational addiction model with health and learning about health. By estimating the structural parameters of the model, I examine preferences and expectations in the trade-off be-

[54] The finding of positive selection is consistent with the comprehensive study of smoking and mortality in the FHS Original Cohort of Darden et al. (2017). Those authors find an average reduction in expected longevity of 4.3 years from continued smoking.

tween smoking and the potential for future health shocks. The structural approach also allows for counterfactual simulations that (*a*) assess the importance of biomarker information in the decision to smoke cigarettes and (*b*) capture the direct effect of smoking, and smoking cessation, on different health outcomes while controlling for unobserved heterogeneity.

Estimates of the structural parameters suggest that there exists heterogeneity in the effect of the accumulated smoking stock on an index of biomarkers. Despite this heterogeneity, learning about an individual's own place in the distribution of this effect, at least under the assumptions of the model and with respect to the FHS data, does not appear to significantly change smoking behavior over time. Both reduced-form and simulation evidence suggest that biomarkers are likely ill-suited to encourage smoking cessation because they are too noisy to provide reliable signals as to the health effects of smoking. Even in the face of a relatively clear and homogeneous signal of the health consequences of smoking (figs. 5*C* and 5*D*), smoking is simulated to decrease by a maximum of 2.5 percent. This finding is very small compared to simulations in which the addiction mechanisms of reinforcement and withdrawal are removed.

This paper also presents evidence of positive selection with respect to smoking and mortality by estimating the correlation in permanent unobserved heterogeneity between these outcomes. As an example, if heavy drinking is correlated with smoking and mortality, then failing to allow for this correlation may overstate the direct effect of smoking. Indeed, with respect to confounding factors, Doll et al. (2004, 4) state, they "are unlikely to have influenced greatly the absolute difference between the overall mortality rates of cigarette smokers and lifelong non-smokers." No evidence is given for this claim. Simulations of the structural model confirm the positive selection and suggest that, when controlling for unobserved heterogeneity, the effects of smoking on mortality outcomes may be less extreme than previously estimated. I find that smoking heavily from age 18 can reduce life expectancy by 5.41 years relative to lifelong nonsmokers and by 3.13 years for light smokers (less than or equal to one pack per day) from age 18. I compare my results to those of Taylor et al. (2002), Doll et al. (2004), and Brønnum-Hansen et al. (2007), which find overall longevity loss from daily smoking to be 7.4–10.5, 10, and 8.7–10.4 years, respectively. The absolute difference between the estimates of those authors and mine may be due to the unrepresentative nature of the Framingham sample, but simulations of a reduced-form model estimated with Framingham data suggest that light and heavy smoking are associated with a 3.77- and 6.56-year loss in longevity, respectively—less than the other studies, but still greater than my preferred results.

Admittedly, estimating smoking responses to health changes over the life cycle is difficult. Indeed, the structural model makes a number of unrealistic assumptions: the distribution of the effect of the smoking stock on the biomarker index is normally distributed (truncated at zero); individuals update their (rational) expectations only at FHS health exams, and not from other physician visits; individuals have time-consistent preferences; price effects are negligible; and the expanded data capture smoking transitions. However, the simple reduced-form model results in tables 4 and 5 also contain strong assumptions (e.g., myopic time preference). Rather than assuming away the role of expectations, this paper makes explicit the assumptions regarding expectations and learning such that I can conduct counterfactual simulations to investigate the mechanisms that drive smoking behavior.

The two main questions of this paper should guide future work. First, do sources of information, personalized or otherwise, exist that effectively convince individuals to stop smoking before the realization of major, chronic health shocks? Second, what are the sources of unobserved heterogeneity that are shown to be correlated across preferences for smoking and health.

Appendix

Solution and Estimation

Following Rust (1987), the current-period utility is augmented by an i.i.d. preference shock that is alternative and health state specific, ϵ_{it}^{dh}. In the empirical implementation below, ϵ_{it}^{dh} is simply an additive econometric error; however, in the theoretical model, ϵ_{it}^{dh} is given a structural interpretation as an unobserved state variable (Aguirregabiria and Mira 2010). Subtracting the additive error term, the alternative specific lifetime value function in health state h, conditional on unobserved heterogeneity μ, is

$$\bar{V}_d^h(S_{it}|\mu) = \bar{U}_{it}^h(A_{it}, d_{it} = d, R_{it}, X_{it}, \mu) + \beta\left\{(1 - \varsigma_{it+1})\sum_{a=0}^{1}\pi_{it+1}^{a}E_{it}[V^a(S_{it+1}|\mu)|d_{it} = d]\right\}.$$

Here, $V^a(S_{it+1}|\mu)$ is the maximal expected lifetime utility of being in health state a in period $t + 1$; β is the annual discount rate, and it is assumed to be fixed at $\beta = 0.95$.[55] The value function is conditional on the unobserved heterogeneity component μ. Assuming that ϵ_{it}^{dh} has an extreme value type I distribution, the

[55] Preferences are therefore assumed to be time consistent. Fixing the discount rate is standard in dynamic structural models because it is often poorly identified (Aguirregabiria and Mira 2010). Recent experimental evidence has shown that smokers are no less time consistent over monetary incentives. Male smokers have higher discount rates than male nonsmokers, but there is no similar correlation in females (Harrison, Lau, and Rutstrom 2010).

maximal (EMAX function) expected lifetime utility has the following closed-form solution:

$$V^h(S_{it+1}|\mu) = EC + \ln\left\{\sum_{d=1}^{D} \exp[\bar{V}_d^h(S_{it+1}|\mu)]\right\} \quad \forall t, \forall h, \tag{A1}$$

where EC = Euler's constant. Individuals are able to calculate the optimal smoking choice and the corresponding value in equation (A1) given the state vector S_{it} and expectations. The choice probabilities are dynamic logit probabilities generated by the choice-specific value functions:

$$P(d_{it} = d|S_{it}, \mu) = \frac{\bar{V}_d^h(S_{it}|\mu)}{\sum_{b=0}^{2} \bar{V}_b^h(S_{it}|\mu)}. \tag{A2}$$

The computational hurdle in calculating the conditional choice probabilities is to solve for the integrated Bellman (EMAX) equation. Technically, the EMAX equation must be calculated for all possible points s in the state space S_{it}. However, given the long time frame of the model and the mixed discrete/continuous nature of the state space, I employ a variant of the Keane and Wolpin (1994) value function interpolation method for approximating the value function. This method amounts to drawing from the state space, calculating the resulting EMAX function for each draw, and interpolating the EMAX function for all other points. The end goal of this procedure is to generate choice probabilities for each individual i, in each time period t, conditional on the unobserved heterogeneity μ and a trial set of the parameters, to enter the likelihood function. The iterative solution method proceeds in two main steps: model simulation and individual-specific solution. While the first solution step yields value function regression coefficients from model simulation, the second step uses these coefficients to calculate the individual-specific, integrated Bellman equation that enters the choice probabilities.

The first step of the solution method is to solve the model for a group of simulated individuals. The goal of this step is to generate a set of regression coefficients that map from the state space to the value function. Because the time horizon is finite ($T = 100$), the model is solved using backward induction, which avoids iterating on the value function itself. Starting in the final period T, I draw n state vectors and sequences of past smoking behavior $\mathbf{D}_{iT-1} = \{d_{i1}, \ldots, d_{iT-1}\}$.[56] Each of these n draws represents one simulated individual. For each of the n draws, I construct the main equations of the model for period T. Note that because the probability of death at the end of period T equals one, each of the choice-specific value functions in period T simply equals the current-period utility from the smoking alternative. Next, I posit a relationship between the n calculated value functions and a set of regressors. The regressors include the drawn state variables in addition to interaction and higher-order terms. The regression of the value function values on the state variables generates coefficients that are specific to time period T. When calculating the expected value function in period $T - 1$ (i.e., age 99), I use the regression coefficients from period T to approx-

[56] In practice, I set $n = 150$.

imate the expected future value function. This process is repeated for all periods back to age 7, $t = 7$.

The first-stage process is conditional on three factors. First, the simulation is conducted for each possible age at which an individual may have taken her first health exam.[57] Second, I discretize the support of the unobserved heterogeneity distribution into K points.[58] Thus, I have a full set of value function regression coefficients (from age 7 to 100) for each age at the initial health exam and for each unobserved type, μ. Finally, the value function regression coefficients are also conditional on the trial set of parameters used to solve the model.[59]

The second main step, conditional on the same trial set of parameters and using the above regression coefficients, involves solving the model for each individual. In each period in which individual i undergoes a health exam, he or she must forecast the future evolution of the state variables and the resulting values associated with all current and future smoking decisions. Luckily, however, because the value function regression coefficients approximate the next-period value function, I must construct the expected value of the next-period state variables conditional on only the current-period smoking decision. Because the chronic health and mortality logit probabilities have closed-form expressions, assuming rational expectations makes the next-period chronic health and mortality transition expectations straightforward. I numerically integrate over the belief distribution and the latent stock of addictive capital with 150 sets of simulation draws.

This process is complicated by the fact that data for the biomarker index, R_{it}, exist only in years in which an FHS health exam was taken. For time periods in which no health exam was taken, in addition to integrating over the future values of the smoking stock and biomarker index, I must also integrate over the current-period value of the biomarker index. In this case, I use the same method as integrating over the future biomarker index, using only different draws (from those used to integrate over the future term) of the i.i.d. error term ν. All other probabilities in the model are constructed conditional on the drawn value of the current-period biomarker index and averaged. For each individual, the model is solved backward from age 100 to generate conditional choice probabilities, biomarker and smoking stock densities, and chronic health and mortality transition probabilities conditional on each draw. This yields 150 sets of smoking and health transition probabilities and 150 individual contributions to the likelihood function.

The parameters of the model are estimated via simulated maximum likelihood. The expanded FHS data contain years for all chronic health and mortality events, as well as noisy measures of smoking behavior in every year until either death or the seventh exam. In constructing the likelihood function, consider first the contribution of individual i. For ease of exposition, define the binary variable y as follows:

[57] In the data, the ages range from 13 to 62. There exists great variation in the data in the timing of the health exams. However, in the simulation, regardless of age at the initial health exam, I use the average number of years between exams to avoid having to simulate the model for all possible combinations of exam sequences.

[58] In practice, I set $K = 3$.

[59] This method generates bias because the error from the value function regression is not incorporated into the standard errors of the parameter estimates. However, the interpolation greatly reduces the computational burden.

$$y_{it} = \begin{cases} 1 & \text{if an exam was taken in year } t \\ 0 & \text{if no exam was taken in } t. \end{cases}$$

Furthermore, let w index the draw from the Monte Carlo simulator used to integrate over agent expectations and unobserved state variables. Define $Q_{wit}^{y=1}|\mu$ as individual i's likelihood contribution in period t for draw w when $y_{it} = 1$ and conditional on unobserved heterogeneity term μ. Let

$$Q_{wit}^{y=1}|\mu = \prod_{d=0}^{2}\left\{P(d_{wit} = d|s_{it}, \mu) \times \left(\prod_{j=1}^{4}\Psi_{jt}\right) \times \Omega_t \times \left[\prod_{h=0}^{1}(\pi_{t+1}^{dh}|\mu)^{\mathbf{1}[H_{it+1}=h]}\right]\left[\prod_{m=0}^{1}(\varsigma_{t+1}^{dm}|\mu)^{\mathbf{1}[M_{it+1}=m]}\right]\right\}^{\mathbf{1}[d_{wit}=d]}. \tag{A3}$$

Here, $P(d_{wit} = d|s_{it}, \mu)$ are the conditional smoking choice probabilities generated from solution of the model from draw w in equation (A2). For individual i, four smoking history measurement equations from the dynamic factor method enter the likelihood function. The probability density function of measurement j (e.g., total years of smoking) is given as

$$\Psi_{jt} = f(\psi_{ijt}|A_{it}) = \frac{1}{\sigma_\psi}\phi((\log Z_{ijt} - \mu_j - \alpha_{jt}A_{it})/\sigma_{\psi_j}).$$

The measurement equations are not indexed by w, the draw of the Monte Carlo simulator, because each measure j comes directly from the data. The normality assumption for η_{it} implies that the density of A_{it} can be written as

$$\begin{aligned}\Lambda_t &= f(\eta_{it}|A_{it-1}, d_{it-1}, \mu) \\ &= \frac{1}{\sigma_\eta}\phi((\log A_{it} - \delta_1 \log A_{it-1} - \delta_2 d_{t-1} - \rho_A\mu)/\sigma_\eta),\end{aligned}$$

where $\phi(\cdot)$ is the density of the standard normal distribution. For years in which an individual took a health exam, the biomarker index R_{it} is observed; however, unless a health exam was taken in the year directly before t, the lagged value of the biomarker index is unobserved. In practice, I use the biomarker index evolution equation to insert the conditional average of the lagged value of R_{it}, R_{it-1}, when it is unobserved. The biomarker contribution to the likelihood function is therefore

$$\begin{aligned}\Omega_t &= g(\nu_{it}|R_{it-1}, X_{it}, A_{it}, \mu) \\ &= \frac{1}{\sigma_\nu}\phi((R_{it} - \zeta R_{it-1} - X_{it}\phi - \theta_i A_{it} - \rho^R\mu)/\sigma_\nu).\end{aligned}$$

Finally, π_{t+1}^h represents the probability of transiting to health state h in period $t + 1$ and ς_{t+1}^m is the probability of transiting to death state m in period $t + 1$.

In periods in which $y_{it} = 0$, the biomarker index, R_{it}, is unobserved. However, all right-hand-side terms in the biomarker technology are observed in these "off"

years because of retrospective questions with the exception of the lagged value of the biomarker index when the previous period did not contain a health exam. In periods with no health exam, I integrate over R_{it} using draws from the Monte Carlo simulator, w. For each draw of the simulator, R_{wit} is constructed, and all other probabilities in the model are calculated conditional on the simulated value of R_{it}. Other probabilities in the model are conditional on the simulated value of R_{wit} for years in which $y_{it} = 0$. In the period directly after a health exam, the lagged value of the biomarker index (i.e., from the exam and not the simulated term) is used in the construction of R_{wit}. Define $Q_{wit}^{y=0}|\mu$ as individual i's likelihood contribution in period t when $y_{it} = 0$:

$$\prod_{d=0}^{2}\left\{P(d_{wit} = d|s_{it}, \mu) \times \left(\prod_{j=1}^{4}\Psi_{jt}\right) \times \left[\prod_{h=0}^{1}(\pi_{wt+1}^{dh}|\mu)^{\mathbf{1}[H_{it+1}=h]}\right]\left[\prod_{m=0}^{1}(\varsigma_{wt+1}^{dm}|\mu)^{\mathbf{1}[M_{it+1}=m]}\right]\right\}^{\mathbf{1}[d_{wit}=d]}. \tag{A4}$$

Here, chronic health and mortality probabilities are conditional on the simulated value of R_{wit}, and the density of the biomarker index has been dropped because it is unobserved.

The total conditional (on μ) likelihood contribution from individual i for all time periods $7, \ldots, T_i$, where T_i is either the period of an individual's death or her final exam, is

$$L_{wi}(\Theta|\mu) = \prod_{t=7}^{T_i}\left[\prod_{y=0}^{1}(Q_{wit}^{y}|\mu)^{\mathbf{1}[Y_{it}=y]}\right]. \tag{A5}$$

Because of the discretized distribution of the unobserved heterogeneity, each individual's unconditional contribution will be a finite mixture of likelihoods. Given K points of support in the estimated distribution of μ, the unconditional likelihood function contribution for individual i is

$$L_{wi}(\Theta) = \sum_{k=1}^{K}\xi_k L_{wi}(\Theta|\mu_k), \tag{A6}$$

where ξ_k is the estimated probability weight placed on mass point k. Individual i's contribution to the likelihood is then the average contribution over the w draws

$$L_i(\Theta) = \frac{1}{150}\sum_{i=1}^{150}L_{wi}(\Theta).$$

The full sample log likelihood function is

$$L(\Theta) = \left[\sum_{i=1}^{N}\log L_i(\Theta)\right]. \tag{A7}$$

The parameters in Θ are estimated via a nested solution method (Rust 1987). The inner algorithm solves the dynamic model for each individual conditional on a given set of parameters and for all mass points of the unobserved heterogeneity distribution. Using the resulting probabilities, the outer algorithm calculates the

unconditional likelihood function, $L(\Theta)$, and attempts to improve the likelihood value via the gradient method of Berndt et al. (1974).

Bayesian Updating

To derive the posterior mean and variance formulas in equations (5) and (6), I assume rational expectations such that an individual's initial belief upon entering the sample regarding her true θ_i is the population distribution

$$E_0(\theta_i) = \tau_{i0} = \bar{\theta},$$
$$V_0(\theta_i) = \psi_{i0} = \sigma_\theta^2.$$

Consider an individual in period t with smoking stock A_{it}. For ease of exposition, assume that an individual takes a health exam each period. Under the assumptions in Section III, define the signal of information as

$$\kappa_{it} = R_{it} - \zeta R_{it-1} - X_{it}\phi - \rho^R\mu = \theta_i A_{it} + \nu_{it}.$$

When deriving the posterior beliefs in period t, an individual considers only her prior beliefs $(\tau_{it-1}, \psi_{it-1})$ and her signal of information k_{it}. According to Bayes's rule, the posterior distribution, f_t, of θ_i is given as

$$f_t(\theta_i|\kappa_n, \tau_{it-1}, \psi_{it-1}) \propto \{f_{t-1}(\theta_i)g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu)\} \times \mathbf{1}\{\theta_i > 0\}. \tag{A8}$$

Note that while $g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu)$ conveys information about κ^{it}, an individual knows A_{it} and, because θ_i is time invariant, can therefore infer information about θ_i over time. First consider $g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu)$:

$$g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu) = \frac{1}{(2\pi\sigma_\nu^2)^{1/2}} \exp\left[\frac{-1}{2\sigma_\nu^2}(\kappa_{it} - \theta_i A_{it})^2\right].$$

Note that because we are concerned with the distribution of θ_i, any term that does not include θ_i can be treated as part of the normalizing constant. Thus, $g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu)$ is proportional to

$$g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu) \propto \exp\left[\frac{-1}{2\sigma_\nu^2}(-2\theta_i\kappa_{it}A_{it} + \theta_i^2 A_{it}^2)\right].$$

Simplifying and completing the square yields

$$\propto \exp\left[-\frac{A_{it}^2}{2\sigma_\nu^2}\left(\theta_i - \frac{\kappa_{it}A_{it}}{A_{it}^2}\right)^2\right].$$

Notice that the term subtracted from θ_i is the within (individual i) variation OLS estimate of θ_i from the nth signal of information. Define $\hat{\theta}_{it} = \kappa_{it}A_{it}/A_{it}^2$. Substituting for $\hat{\theta}_{it}$, we have that

$$g(\kappa_{it}|A_{it}, \theta_i, \sigma_\nu) \propto \exp\left[-\frac{A_{it}^2}{2\sigma_\nu^2}(\theta_i - \hat{\theta}_{it})^2\right]. \tag{A9}$$

Now consider the prior probability distribution of θ_i:

$$f_{t-1}(\theta_i) = \begin{cases} \dfrac{1}{(2\pi\psi_{it-1})^{1/2}} \exp\left[\dfrac{1}{2\psi_{it-1}}(\theta_i - \tau_{it-1})^2\right] & \text{if } \theta_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The nice aspect of the conjugate distribution assumption is that we can characterize the posterior distribution sufficiently with closed-form expressions for the posterior mean and variance. Therefore, we have to characterize only that part of the posterior density that captures the mean and variance. In that light, consider the product of the exponential portions of equation (A9) and the positive support of the prior distribution after rearranging terms and absorbing those without θ_i into the normalizing constant:

$$f_t(\theta_i) \propto \left\{ -\frac{1}{2\psi_{it-1}\sigma_\nu^2}\left[\theta_i^2(A_{it}^2\psi_{it-1} + \sigma_\nu^2) - 2\theta_i\left(A_{it}^2\psi_{it-1}\hat{\theta}_{it} + \sigma_\nu^2\tau_{it-1}\right)\right]\right\}.$$

After rearranging and completing the square, we have the kernel of a normal distribution representing the posterior distribution:

$$f_t(\theta_i) \propto \left\{ -\frac{A_{it}^2\psi_{it-1} + \sigma_\nu^2}{2\psi_{it-1}\sigma_\nu^2}\left[\theta_i - \left(\frac{A_{it}^2\psi_{it-1}\hat{\theta}_{it} + \sigma_\nu^2\tau_{it-1}}{A_{it}^2\psi_{it-1} + \sigma_\nu^2}\right)^2\right]\right\}. \tag{A10}$$

The posterior mean and variance are

$$\tau_{it} = E(\theta_i|\kappa_t, \tau_{it-1}, \psi_{it-1}) = \left(\frac{A_{it}^2\psi_{it-1}}{A_{it}^2\psi_{it-1} + \sigma_\nu^2}\right)\hat{\theta}_{it} + \left(\frac{\sigma_\nu^2}{A_{it}^2\psi_{it-1} + \sigma_\nu^2}\right)\tau_{it-1},$$

$$\psi_t = \mathrm{Var}(\theta_i|\psi_{it-1}, \sigma_\nu) = \frac{\psi_{it-1}\sigma_\nu^2}{A_{it}^2\psi_{it-1} + \sigma_\nu^2}.$$

Rearranging these equations yields the posterior mean and variance equations above.

Prices

Cigarette taxes have played a large role in the observed decline in smoking in the United States, especially in young and light smokers. Chaloupka and Warner (2000) and Chaloupka, Yurekli, and Fong (2012) review the literature on cigarette tax and price elasticities and conclude that excise taxes are effective in controlling tobacco use. However, the same paper by Chaloupka and Warner and more recent research by Callison and Kaestner (2014) find a small estimate of the price elasticity of demand for adult smokers of roughly 0.035. In fact, Callison and Kaestner find elasticities that are not much larger (0.3–0.7) for those 18–34. While the national average tax per pack of cigarettes has increased dramatically since 1990, 48 million Americans smoked cigarettes in 2009, and figure 1 shows that the observed decline in smoking has leveled off.[60]

[60] The average state-plus-federal tax rate on a pack of cigarettes has increased from $0.30 in the early 1990s to over $2.00 in 2009. See http://www.cdc.gov/tobacco/.

Figure A1 presents a time series of the real (1982–84 cents) average retail price in cents per 20-cigarette pack for Massachusetts.[61] As shown in the figure, the period 1955–80 shows little increase in average real prices. Figure A1 is important because it shows that there is little variation in real cigarette prices when individuals in my sample were relatively young—exactly when they would be most sensitive to prices. By the 1990s, when cigarette taxes began to rise exponentially, most of my sample had either quit smoking or died.

Reduced-Form Model

This section presents a simple "reduced-form" model to which I compare the structural model. The four estimated equations are a multinomial logit model for light and heavy smoking relative to not smoking, an OLS regression of the biomarker index, and logit models for chronic health onset and mortality. Each equation is estimated separately, and because of data restrictions, the sample sizes for each equation differ.

Because the biomarker index is constructed outside of the structural model, data exist for R_{it} in each of the seven exams conditional on an individual surviving. The biomarker index is estimated on data from FHS exams 2–7 ($n = 14{,}322$ person/exams) as a function of its lag, demographic characteristics, and the four smoking history measurements: total number of years smoking entering the period (total), total number of years smoking entering the period since last not smoking (tenure), total number of years since last smoked (cessation), and the average intensity smoking when smoking (average).

Smoking is modeled with multinomial logit specification for light and heavy smoking relative to not smoking. Smoking is a function of demographics, smoking history measurements entering the period, and the chronic heath state. The model is estimated on the expanded data ($n = 145{,}331$ person/years) from age 7 until the last FHS exam or death.

The logit equation for the onset of chronic health is estimated for all person/year observations after the first exam (at which no individuals were in the chronic health state) and until the period of onset. The equation is a function of demographics, smoking history measures, and the biomarker index. In periods in which there was no health exam, I impute the biomarker index with the last observed value at a health exam. The mortality logit equation specification is similar but also includes the measure of whether an individual has transited to the chronic health state.[62]

Similar to the structural model simulations, the estimates of the reduced-form model are used to simulate health outcomes from age 7 until age 100. One thousand smoking and health shocks are drawn, and the results are averaged by age. The chronic health and mortality equations are simulated under forced smoking patterns, and the biomarker index is allowed to evolve according to its estimated equation. These simulations generate the mortality and chronic health results in table 12.

[61] I thank Koleman Strumpf for constructing the price data set. See Darden et al. (2017) for more information.

[62] Estimates of these equations are available on request.

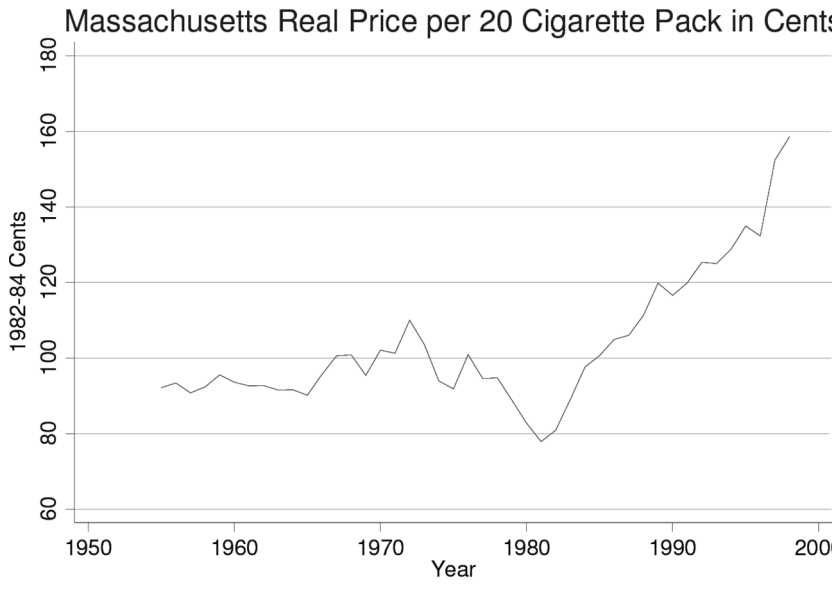

FIG. A1.—Average real cigarette prices (cents per pack) in Massachusetts over time. Average real prices are in 1982–84 cents per 20-cigarette pack in Massachusetts. Price data are from the *Tax Burden on Tobacco: Historical Compilation*. Consumer Price Index data are from the "All Urban Consumers" series with base period 1982–84. See Darden et al. (2017) for more information.

The reduced-form model is also used to conduct sensitivity analysis to (1) the choice of sample and (2) the method of imputation. Figure A2 reports smoking probabilities from reduced-form simulations under four different samples. Sample A is my preferred sample in which individuals with missing values of smoking and health variables are dropped and smoking is imputed in interim years assuming that behavior changes at the midpoint between exams. Sample B keeps individuals with missing values and imputes these values with a flexible regression imputation technique. The smoking imputation is the same in sample A and sample B. Sample C keeps those with missing values but imputes smoking in interim years with last observed smoking behavior. Sample D drops individuals with missing values and imputed interim smoking with the last observed smoking observation.

Each figure represents a baseline simulation in which smoking, biomarkers, and chronic health and mortality evolve according to their estimated reduced-form models. Goodness-of-fit tests fail to reject the null hypotheses that the simulated results are different. Reduced-form estimates and test results are available by request.

Chronic Conditions

Table A1 lists the chronic conditions are available in the FHS offspring data.

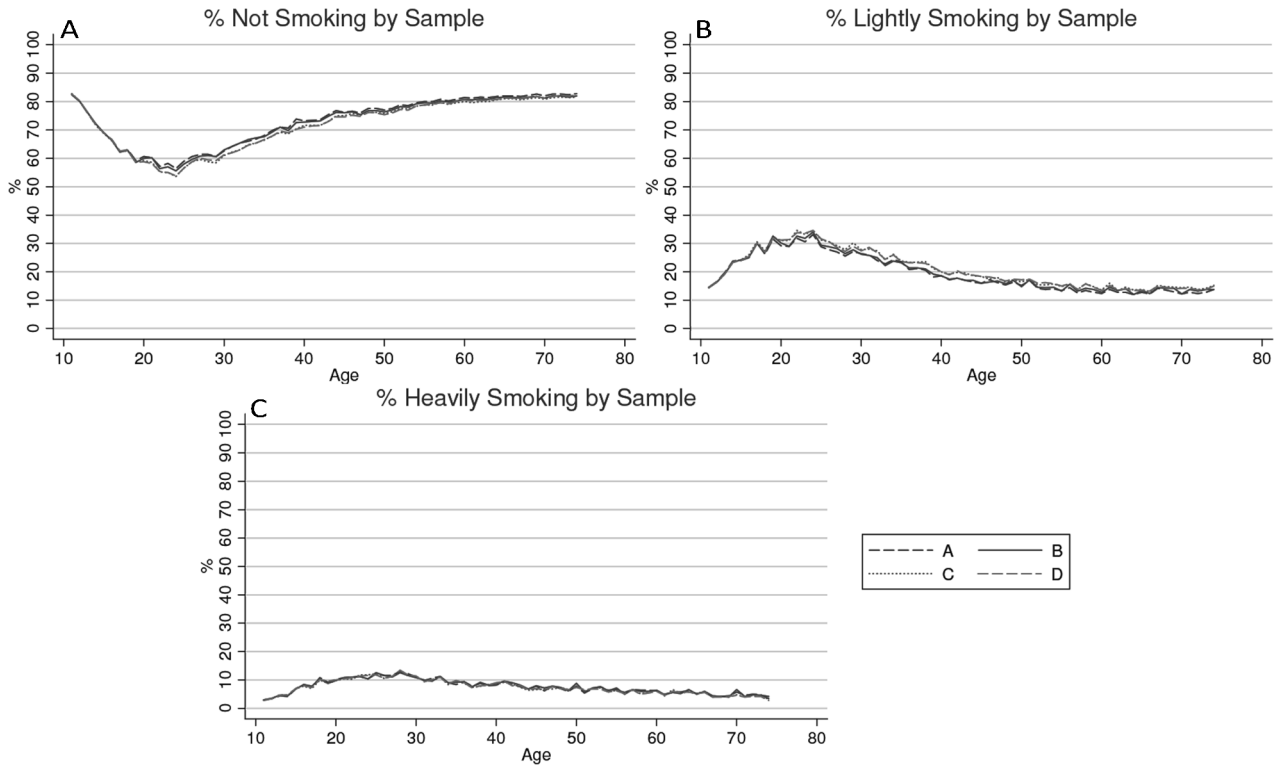

FIG. A2.—Simulated smoking behavior by sample and smoking imputation method. Percentage of simulated sample in each smoking alternative by age, sample construction, and smoking imputation method.

TABLE A1
CHRONIC CONDITIONS THAT INDUCE THE CHRONIC HEALTH STATE

Coronary heart disease:	Cancer (*Cont.*):
Myocardial infarction	Esophagus
Angina pectoris	Stomach
Cerebrovascular accident (CVA):	Small intestine
Atherothrombotic infarction	Colon
Transient ischemic attack	Rectum
Cerebral emolism	Liver
Subarachnoid hemorrhage	Gallbladder
Other CVA	Pancreas
Intermittant claudication	Retroperitoneum
Congestive heart failure	Peritoneum
Cancer:	Nasal cavities
Lip	Ear
Tongue	Larynx
Major salivary gland	Trachea bronchus
Gum	Lung
Floor of mouth	Pleura
Mouth other	Hematopoietic
Oropharynx	Bones and joints
Hypopharynx	Female and male breast
Prostate	Uterus
Testes	Cervix
Penis	Placenta
Other male genital	Corpus uteri
Urinary bladder	Ovary
Kidney	Other female genital
Eye	Thyroid
Brain	Other endocrine
Lymph nodes	Melanotic skin

Biomarker Index Construction

Table A2 shows the point values assigned for different biomarkers for women. These values follow directly from table 5 of D'Agostino et al. (2008). For example, a diabetic woman who does not take blood pressure medication with HDL of 36, total cholesterol of 180, and SBP of 135 is observed with 6 points for R_{it}, the biomarker index.

TABLE A2
BIOMARKER INDEX POINT TOTALS FOR WOMEN

Points	HDL	Total Cholesterol	SBP Not Treated	SBP Treated	Diabetic
−3			<120		
−2	60<				
−1	50–59			<120	
0	45–49	<160	120–129		No
1	35–44	160–199	130–139		
2	<35		140–149	120–129	
3		200–239		130–139	Yes
4		240–279	150–159		
5		280<	160<	140–149	
6				150–159	
7				160<	

NOTE.—Points totals for women come directly from table 5 of D'Agostino et al. (2008).

References

Adda, J., and V. Lechene. 2001. "Smoking and Endogenous Mortality: Does Heterogeneity in Life Expectancy Explain Differences in Smoking Behavior?" Discussion paper, Dept. Econ., Univ. Oxford.

———. 2013. "Health Selection and the Effect of Smoking on Mortality." *Scandinavian J. Econ.* 115 (3): 902–31.

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

Arcidiacono, P., H. Sieg, and F. A. Sloan. 2007. "Living Rationally under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking." *Internat. Econ. Rev.* 48 (1): 37–65.

Becker, G. S., M. Grossman, and K. Murphy. 1994. "An Empirical Analysis of Cigarette Addiction." *A.E.R.* 84 (3): 396–418.

Becker, G. S., and K. Murphy. 1988. "A Theory of Rational Addiction." *J.P.E.* 96 (4): 675–700.

Berndt, E., B. Hall, R. Hall, and J. Hausman. 1974. "Estimation and Inference in Nonlinear Structural Models." *Ann. Econ. and Soc. Measurement* 3:653–65.

Bernheim, B. D., and A. Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *A.E.R.* 94 (5): 1558–90.

Bize, R., B. Burnand, Y. Mueller, M. R. Walther, and J. Cornuz. 2009. "Biomedical Risk Assessment as an Aid for Smoking Cessation." *Cochrane Database of Systematic Reviews* (2): CD004705. DOI:10.1002/14651858.CD004705.pub3.

Blau, D. M., and D. B. Gilleskie. 2008. "The Role of Retiree Health Insurance in the Employment Behavior of Older Men." *Internat. Econ. Rev.* 49 (2): 475–514.

Brønnum-Hansen, H., K. Juel, M. Davidsen, and J. S. Rensen. 2007. "Impact of Selected Risk Factors on Expected Lifetime without Long-Standing Limiting Illness in Denmark." *Preventive Medicine* 45:49–53.

Callison, K., and R. Kaestner. 2014. "Do Higher Tobacco Taxes Reduce Adult Smoking? New Evidence of the Effect of Recent Cigarette Tax Increases on Adult Smoking." *Econ. Inquiry* 52 (1): 155–72.

Chaloupka, F. J., and K. E. Warner. 2000. "The Economics of Smoking." In *Handbook of Health Economics*, vol. 1B, edited by Anthony J. Culyear and Joseph P. Newhouse, 1539–1627. Amsterdam: Elsevier.

Chaloupka, F. J., A. Yurekli, and G. T. Fong. 2012. "Tobacco Taxes as a Tobacco Control Strategy." *Tobacco Control* 21:172–80.

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

Clark, A., and F. Etilé. 2002. "Do Health Changes Affect Smoking? Evidence from British Panel Data." *J. Health Econ.* 21 (4): 533–62.

Crawford, G. S., and M. Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* 73 (4): 1137–73.

Cunha, F., and J. J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *J. Human Resources* 43:738–82.

Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

D'Agostino, R., et al. 2008. "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study." *Circulation* 117 (6): 743–53.

Darden, M., D. Gilleskie, and K. Strumpf. 2017. "Smoking and Mortality: New Evidence from a Long Panel." *Internat. Econ. Rev.*, forthcoming.

Doll, R., R. Peto, J. Boreham, R. Gray, and I. Sutherland. 2004. "Mortality in Relation to Smoking: 50 Years' Observations on Male British Doctors." *British Medical J.* 328:1519–28.

Doonan, R. J., A. Hausvater, C. Scallan, D. P. Mikhailidis, L. Pilote, and S. S. Daskalopoulou. 2010. "The Effect of Smoking on Arterial Stiffness." *Hypertension Res.* 33:398–410.

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." *J. European Econ. Assoc.* 11 (S1): 221–58.

Gepner, A. D., M. E. Piper, H. M. Johnson, et al. 2011. "Effects of Smoking and Smoking Cessation on Lipids and Lipoproteins: Outcomes from a Randomized Clinical Trial." *American Heart J.* 161:145–51.

Gilleskie, D. B. 1998. "A Dynamic Stochastic Model of Medical Care Use and Work Absence." *Econometrica* 66 (1): 1–45.

Gilleskie, D. B., and K. S. Strumpf. 2005. "The Behavioral Dynamics of Youth Smoking." *J. Human Resources* 4:822–66.

Gruber, J., and B. Köszegi. 2001. "Is Addiction 'Rational'? Theory and Evidence." *Q.J.E.* 116 (4): 1261–1303.

Gul, F., and W. Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica* 69:1403–35.

———. 2007. "Harmful Addiction." *Rev. Econ. Studies* 74:147–72.

Hai, R., and J. J. Heckman. 2015. "A Dynamic Model of Health, Education, and Wealth with Credit Constraints and Rational Addiction." Manuscript, Soc. Sci. Res. Network.

Harrison, G. W., M. I. Lau, and E. E. Rutstrom. 2010. "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark." *J. Health Econ.* 29:708–17.

Heckman, J. J. 1981. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time Discrete Data Stochastic Process." In *Structural Analysis of Discrete Data and Econometric Applications*, vol. 4, edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press.

Heckman, J. J., and S. Navarro. 2007. "Dynamic Discrete Choice and Dynamic Treatment Effects." *J. Econometrics* 136:341–96.

Heckman, J. J., and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2): 271–320.

Hollands, G. J., S. C. L. Whitwell, R. A. Parker, et al. 2012. "Effect of Communicating DNA Based Risk Assessments for Crohn's Disease on Smoking Cessation: Randomised Controlled Trial." *British Medical J.* 345.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions.* 2nd ed. New York: Wiley.

Kaestner, R., M. Darden, and D. Lakdawalla. 2014. "Are Investments in Disease Prevention Complements? The Case of Statins and Health Behaviors." *J. Health Econ.* 36:151–63.

Keane, M. P. 2010. "Structural vs. Atheoretic Approaches to Econometrics." *J. Econometrics* 156:3–20.

Keane, M. P., and K. I. Wolpin. 1994. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *Rev. Econ. and Statis.* 76 (4): 648–72.

Khwaja, A. 2006. "A Life Cycle Analysis of the Effects of Medicare on Individual Health Incentives and Health Outcomes." Manuscript. https://www.researchgate.net/publication/228776492_A_Life_Cycle_Analysis_of_the_Effects_of_Medicare_on_Individual_Health_Incentives_and_Health_Outcomes.

———. 2010. "Estimating Willingness to Pay for Medicare Using a Dynamic Life-Cycle Model of Demand for Health Insurance." *J. Econometrics* 156 (1): 130–47.
Khwaja, A., F. A. Sloan, and S. Chung. 2006. "Learning about Individual Risk and the Decision to Smoke." *Internat. J. Indus. Org.* 24:683–99.
Manski, C. 2004. "Measuring Expectations." *Econometrica* 72:1329–76.
McClure, J. B. 2001. "Are Biomarkers a Useful Aid in Smoking Cessation? A Review and Analysis of the Literature." *Behavioural Medicine* 27:37–47.
Mira, P. 2007. "Uncertain Infant Mortality, Learning, and Life-Cycle Fertility." *Internat. Econ. Rev.* 48 (3): 809–46.
Orphanides, A., and D. Zervos. 1995. "Rational Addiction with Learning and Regret." *J.P.E.* 103 (4): 739–58.
Parks, G., T. Greenhalgh, M. Griffin, and R. Dent. 2008. "Effect on Smoking Quit Rate of Telling Patients Their Lung Age: The Step2Quit Randomised Control Trial." *British Medical J.* 336.
Rust, J. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033.
Sloan, F. A., V. K. Smith, and D. H. Taylor. 2003. *The Smoking Puzzle: Information, Risk Perception, and Choice.* Cambridge, MA: Harvard Univ. Press.
Smith, V. K., D. H. Taylor, F. A. Sloan, F. R. Johnson, and W. H. Desvousges. 2001. "Do Smokers Respond to Health Shocks?" *Rev. Econ. and Statis.* 83 (4): 675–87.
Stead, L. F., G. Bergson, and T. Lancaster. 2004. "Physician Advice for Smoking Cessation." *Cochrance Database of Systematic Reviews* (2): CD000165. DOI:10.1002/14651858.CD000165.pub3.
Strine, T. W., C. A. Okoro, D. P. Chapman, et al. 2005. "Health-Related Quality of Life and Health Risk Behaviors among Smokers." *American J. Preventive Medicine* 28:182–87.
Taylor, D. H., V. Hasselblad, S. J. Henley, M. J. Thun, and F. A. Sloan. 2002. "Benefits of Smoking Cessation for Longevity." *American J. Public Health* 92 (6): 990–96.
Viscusi, W. K. 1990. "Do Smokers Underestimate Risks?" *J.P.E.* 98 (6): 1253–69.
Viscusi, W. K., and W. N. Evans. 1990. "Utility Functions That Depend on Health Status: Estimates and Economics Implications." *A.E.R.* 80 (3): 353–74.
Viscusi, W. K., and J. K. Hakes. 2008. "Risk Beliefs and Smoking Behavior." *Econ. Inquiry* 46 (1): 45–59.