Reference Health and the Demand for Medical Care

Matthew Harris[*] Jennifer Kohn[‡]

January 6, 2016

Abstract

This paper demonstrates the utility individuals receive from health is affected by health in previous periods, which we term reference health. We demonstrate the relevance of reference health to economic modeling theoretically, empirically and with a policy relevant simulation. We estimate a joint conditional distribution of medical care and consumption nested in a finite mixture framework. Including reference health improves our ability to match individuals in the top 5 percent of medical care spenders by 65 percent. Using our estimated parameters, we show that models that omit reference health will understate by 50 percent the cost-savings of healthy aging initiatives.

Keywords: Reference Dependence, Human Capital, Demand for Medical Care, Health Dynamics, Semi-Parametric, Conditional Density Estimation

[*]University of Tennessee, Haslam College of Business

[†]Drew University, Department of Economics and Business

[‡]We are grateful to Robert Patrick for his collaboration on the theory. We are also grateful to David Bradford, Jim Burgess, Andrew Ching, Chris Cronin, Michael Darden, Donna Gilleskie, William Neilson, Nicholas Papageorge, Meghan Skira, Robert Town, Josh Kinsler, Forrest Spence, and the participants at the 2014 Annual Health Econometrics Workshop and 2015 Annual Health Economics Conference. We also gratefully acknowledge participants at numerous conferences and seminars who commented on both theoretical and empirical iterations of this work. We remain responsible for all errors.

I Introduction

Economists have shown that an individual's health is a factor in many decisions of interest including labor supply, investment in financial assets, investment in education and other human capital, consumption of goods and services, and the demand for medical care. However, health is fundamentally different from other economic inputs. Clearly, health is different from exogenous, static explanatory variables (e.g., gender and race). Health also differs from many time-varying individual characteristics (e.g., education, work experience, marital status or number of children) in that health evolves in a less linear and predictable manner. Finally, unlike other forms of human capital, economists have treated health as both a consumption and an investment good.

Empirical models that include health in the utility function typically include only contemporaneous health, implicitly assuming contemporaneous health captures all information relevant to economic decisions (Grossman, 1972; Gilleskie, 1998; Viscusi and Evans, 1990; Hall and Jones, 2007; Edwards, 2008; Bound, Schoenbaum, Stinebrickner, and Waidmann, 1999; Yang, Gilleskie, and Norton, 2009; Yogo, 2009; Khwaja, 2010; DiNardi, French, and Jones, 2010; Hugonnier, Pelgrin, and StAmour, 2013).[1] However, diverse literatures have successfully incorporated past realizations of key variables to explain economic puzzles ranging from the equity premium puzzle (Constantinides, 1990; Dai and Grischenko, 2014), labor supply (Koszegi and Rabin, 2006, 2009; Crawford and Meng, 2011) and addiction (Becker and Murphy, 1988; Darden, 2015). The reference dependence literature has shown that individuals value contemporaneous wealth differently if they were more (less) wealthy in the past. Similarly, individuals may value contemporaneous health differently if they were more (less) healthy in the past.

We combine elements from the literature on reference dependent utility with the literature on dynamic models of investment in health to address the economic puzzle of why some individuals demand high amounts of medical care. We incorporate past realizations of health (which we term reference health) into the utility function in a dynamic model of health, consumption, and the demand for medical care. Consider two individuals with the same contemporaneous level of health. Consistent with the literature on reference dependence, we hypothesize that an individual who was previously in better health would have greater marginal utility for health improvement than an individual who was previously at the same level of health. Greater marginal utility from health justifies higher medical care spending. While our present application is the demand for medical care, evidence that reference health affects the marginal utility from contemporaneous health has implications for any economic model where health is

[1] This list is clearly a subset of the vast relevant literature.

a factor (e.g., investment decisions, labor force participation, smoking, exercise, etc).

Understanding the demand for medical care, particularly among heavy users, is highly policy relevant as medical care spending accounts for 18 percent of U.S. GDP. Half of this figure (nine percent of GDP) is consumed by five percent of the population. Contemporaneous health alone has surprisingly little explanatory power in predicting who will be in the top five percent of medical care consumers. Fewer than 20 percent of these high spenders report poor health, while 7.5 percent report excellent health (Schoenman, 2012). Conversely, not all those in poor health consume large amounts of medical care. Nearly 40 percent of medical care spending by individuals in poor health is incurred by the top five percent of this group (Claxton, Kamal, and Cox, 2014). Models that include only contemporaneous health are therefore unlikely to explain why some individuals demand large amounts of medical care. By incorporating reference health, our model can better explain the variation in medical care spending among individuals in identical contemporaneous health.

To motivate our empirical analysis we first present a theoretical model in the Grossman (1972) tradition that illustrates how incorporating reference health in the utility function justifies high medical care spending. Using data from the US Health and Retirement Survey (HRS), we then demonstrate that conditional on contemporaneous health, medical care spending is indeed correlated with reference health. We posit that reference health affects the demand for medical care through the utility function. Prior to presenting the results of our full model, we consider an extensive array of alternative explanations. Alternative mechanisms by which higher reference health can increase medical care spending include: reference health merely reflecting health shocks, costs of learning to access the health system, start up costs associated with new health conditions, measurement error of health, or reference health affecting the marginal productivity of medical care rather than the marginal utility of health. Finding no empirical evidence for these alternative hypotheses, we turn to our full dynamic specification.

Our econometric methods address the particular challenges of estimating a joint dynamic model with correlated unobservable heterogeneity and highly skewed health and expenditure distributions. We estimate the joint demands for medical care and consumption in a finite mixture framework that allows for permanent and time varying unobservable heterogeneity (Cameron and Trivedi, 1986; Pohlmeier and Uhlrich, 1995; Cameron and Johansson, 1997; Deb and Trivedi, 1997; Gurmu, 1997; Shen, 2013). We use a conditional density estimator (Gilleskie and Mroz, 2004) that models the density of health and allows explanatory variables to have different impacts over the entire support of the distribution of health and medical care spending. We operationalize reference health as the average of the individual's past two realized health states. We offer sensitivity analysis and a discussion of this construction of reference health with respect to both the theory and data limitations. We use the initial conditions and estimated parameters of the model to forward simulate an individual's demand for medical care, consumption, and health transitions over the sample period. We compare the predictions of our model that includes both contemporaneous and reference health with predictions from a model that includes only contemporaneous health, medical care and consumption.

We find strong empirical evidence that reference health directly affects the demand for medical care at all levels of spending. Conditional on contemporaneous health, an individual whose reference health is 10 percent greater (or equivalently, who experiences a 10 percent decline from reference health) will increase her medical care spending by 14 percent on average. As we hypothesize, the effect of reference health is strongest at the high end of the medical care spending distribution. Conditional on contemporaneous health, an individual whose reference health is 10 percent greater is 22 percent more likely to be among the top 5 percent of medical care consumers. For validation, we use estimated parameters of the model to demonstrate that models with reference health outperform models without reference health in matching the demographics of individuals in the top of the medical care spending distribution.

Finally, we use our estimates to simulate different paths of medical care spending associated with different trajectories of health over a twelve year span. These simulations clearly illustrate that while contemporaneous health matters, the path of health matters too. Comparing simulations with and without reference health demonstrates that models that omit reference health underestimate by half the potential cost savings from policies such as health aging initiatives that would prevent more volatile trajectories of health.

The rest of the paper proceeds as follows: Section II motivates our empirical specification with a theoretical model. Section III describes the data used in estimation, provides descriptive evidence in support of our main hypothesis, and evidence regarding alternative hypotheses. Section IV details our estimation strategy and econometric identification. Section V contains the results of our model and simulation, and Section VI concludes.

II Theoretical Motivation

To motivate our empirical specification we extend the Grossman (1972) model by including reference health as an element of an individual's utility to illustrate how individuals may make decisions based on reference health. The model developed here is purely for expository purposes and is not intended to stand on its own.[2] Rather, the purpose here is to illustrate how incorporating reference health can provide additional intuition and inform empirical models with greater explanatory power. We start with a dynamic life cycle specification and then simplify to a two-period model to derive a comparative static that will most clearly illustrate the effect of reference health on the demand for medical care.[3]

The objective of the individual is to maximize her expected present discounted utility from health, H_t, and non-medical consumption, z_t, given her reference health, R_t. Specifically, in our empirical specification we operationalize reference health as the average of the past two lags of health, $R_t = \frac{H_{t-1}+H_{t-2}}{2}$, and we test the sensitivity of our results to this functional form. However, to ease our theoretical exposition we maintain a general recursive functional form for reference health as denoted in II.1[4]

$$\begin{aligned} &U(z_t, H_t; R_t) \\ &R_{t+1} = f(R_t, H_t) \end{aligned} \tag{II.1}$$

[2]Please see Kohn and Patrick (2012) for a complete exposition of the model in a continuous time optimal control framework.

[3]We make several modeling trade-offs consistent with our focus and data. See Hugonnier, Pelgrin, and StAmour (2013) Table 3 for an excellent summary of the major modeling choices in much of the literature. First, we do not model the choice of insurance. Insurance choice is clearly endogenous to health status and the demand for medical care, but also highly constrained by employment, income (Medicaid) and in the U.S., age (Medicare over age 65). In our data nearly all individuals are insured for the full sample period and we have checked for, but do not observe discontinuities in medical care usage at age 65. Second, we do not model labor decisions, particularly retirement. While we do not add separate equations for these decisions, we do include control variables on employment, insurance status, age, and income in the model. Third, we do not model non-medical care health investments (e.g. exercise, health-related consumption such as smoking or diet) for which we do not have adequate data. Rather, we employ permanent and time-varying discrete factor random effects to mitigate concerns about these omitted health inputs. Section IV details our econometric strategy.

[4]There is an extensive literature exploring functional forms for "reference" values. Prior work on habit persistence has used exponentially weighted sums of past values, one period lags, two period lags, etc. (Constantinides, 1990; Ryder and Heal, 1973; Campbell and Cochrane, 1999). The literature on reference dependent utility has also offered competing hypotheses on how the reference point is formed. In the subset of the reference dependence literature that focuses on internal reference points, Koszegi and Rabin (2006) use rational expectations to form the reference point, Baucells, Weber, and Welfens (2011) tests several different combinations of past values. We have evaluated different formulations of reference health under adaptive expectations and offer a discussion in Section B and Table V.

Higher values of H_t and R_t reflect higher values of contemporaneous and reference health respectively. We assume that utility is concave in health and consumption. The marginal utility of reference health is negative, $U^R < 0$.[5] Individuals with higher levels of reference health receive less utility from any health and consumption bundle.

The critical relationship is that of health and reference health, specifically the sign of the cross partial, $U^{H,R}$. We hypothesize that higher levels of reference health increase the marginal utility of health, $U^{H,R} > 0$. Two individuals with identical contemporaneous health will value improvements to their health differently depending on their health in previous periods. Intuitively, individuals who have been in good health will want more good health in the future, whereas individuals in poor health become accustomed to poor health. Thus, individuals in poor health experience disutility from two sources, first, their poor health, and second the loss of health from their reference levels. The literature on reference dependence refers to this second source as gain/loss utility. This intuition is consistent with a rational expectations story as used by Becker and Murphy (1988) in the rational addiction literature, a habit formation story as used by Constantinides (1990) in the life cycle consumption literature, and the reference dependence literature as used by Koszegi and Rabin (2006). Indeed, when an individual's reference point is defined by her own past experience and relative utility is defined as the difference between contemporaneous and reference levels, then the implications of reference dependence and habit persistence are very similar. After specifying the rest of the mode,l we demonstrate that the relationship between medical care and reference health is determined by the sign of $U^{H,R}$.

Individuals maximize utility subject to constraints on the evolution of health and wealth. Individuals can improve their health by investing in medical care, m_t, with diminishing marginal returns. Contemporaneous health also affects the production of health, yielding an investment function $I(m_t, H_t)$.[6] An individual's wealth evolves, consistent with the extant literature, as the difference between interest on current wealth, $(1 + r)W_t$, plus income, Y_t, minus expenditures on consumption and medical care, $-z_t - p_m m_t$, with the price of non-medical consumption normalized to 1.[7] Wealth, W_t, must remain above some minimum level that does not preclude debt. Health is not bounded from below. If H_t falls below some minimum value of health, however,

[5] Partial derivatives are denoted by superscripts.

[6] The assumption of a health production function with diminishing returns to medical care is consistent with Ehrlich and Chuma (1990) and Galama (2011).

[7] Making income a function of health does not change the implications of the model with respect to reference health. For a complete exposition of the effect of health-dependent income see Kohn and Patrick (2012)

the individual dies. The constraints of the model are:

$$
\begin{aligned}
H_{t+1} &= H_t + I(m_t, H_t) + \delta_t \\
W_{t+1} &= (1+r)W_t + Y_t - z_t - p_m m_t \\
W_t &\geq W_{min}, \qquad \forall t \\
H_t &\leq H_{max}, \qquad \forall t
\end{aligned}
\tag{II.2}
$$

Next-period health is also subject to a stochastic shock, δ_t, about which we make no distributional assumptions. In the context of our data, which is comprised of older individuals, we consider the shock to be negative in most of our discussion.[8] The health shock likely has two properties that motivate our empirical specification. First, it is likely to be persistent. Individuals who experience declines in health are likely to experience similar events in subsequent periods. Second, the health shock is unlikely to be independent with respect to the demand for medical care and consumption. Our empirical strategy addresses both concerns.[9]

We express the constrained optimization problem recursively:

(II.3)

$$
\begin{aligned}
V_t(W_t, H_t; R_t) &= \max_{z_t, m_t} U(z_t, H_t; R_t) + \beta E V_{t+1}(W_{t+1}, H_{t+1}, R_{t+1}) \\
&= \max_{z_t, m_t} U(z_t, H_t; R_t) + \beta E V_{t+1}\Big((1+r)W_t + Y_t - z_t - p_m m_t, H_t + I(m_t, H_t) + \delta_t, f(R_t, H_t)\Big)
\end{aligned}
$$

where the dynamic constraints are substituted for the state variables in the second line of II.3. The first order conditions for the choice variables, z_t, consumption, and m_t, medical care, are:

$$
\begin{aligned}
z_t: \quad & U_t^z = r\beta E_t V_{t+1}^W \\
m_t: \quad & E_t V_{t+1}^H I_t^m = r\beta E_t V_{t+1}^W p_m
\end{aligned}
\tag{II.4}
$$

The first order conditions from II.4 show that individuals choose an optimal level of consumption when the marginal utility of consumption equals the discounted marginal value of wealth in the next period. Critically, these equations in conjunction with the envelope conditions below show that the marginal value of wealth in the next period is itself a function of both health and reference health in the next period. Since medical care is a derived demand from the demand for health, individuals choose an optimal

[8]Consistent with Kahneman and Tversky (1979) and Koszegi and Rabin (2006), the effects of a positive health shock will have an effect of the opposite sign but smaller magnitude. See Kohn and Patrick (2012) for a more detailed derivation.

[9]There is a strand of literature that examines uncertainty in the productivity of medical care (Dardanoni and Wagstaff, 1990). Empirically, we cannot separate ineffective medical care from a separate health shock shock. For the purpose of theoretical exposition, we simplify the stochastic element in health production to one additive shock. See Section IV.D.

amount of medical care to equate the expected marginal benefits of health to the expected marginal costs in the form of wealth. These first order conditions therefore yield two insights that motivate our empirical approach. First, any argument that affects the demand for consumption must be included in the demand for medical care. Second, since reference health affects the marginal value of health, it also affects the marginal value of wealth. Therefore, we can express optimal z_t and m_t:

$$\begin{aligned} z_t^* &= z(H_t, R_t, W_t, Y_t; \mathbf{X_t}) \\ m_t^* &= m(H_t, R_t, W_t, Y_t; \mathbf{X_t}) \end{aligned} \tag{II.5}$$

where $\mathbf{X_t}$ is a vector of exogenous variables such as age, race, etc. The envelope conditions on the health, reference health, and wealth states are:

$$\begin{aligned} H_t: \quad V_t^H &= U_t^H + \beta E_t \Big(V_{t+1}^H (1 + I_t^H) + V_{t+1}^R f_t^{H_t} \Big) \\ R_t: \quad V_t^R &= U_t^R + \beta E_t V_{t+1}^R f_t^R \\ W_t: \quad V_t^W &= (1+r)\beta E_t V_{t+1}^W \end{aligned} \tag{II.6}$$

The envelope conditions for health and wealth can then be substituted into the first order condition for medical care to yield an equilibrium expression for medical care demand, with the marginal benefits on the left-hand side equaling the marginal cost on the right:

$$E_t \Big[U_{t+1}^H + \beta E_t \big(V_{t+2}^H (1 + I_{t+1}^H) + V_{t+2}^R f_{t+1}^{H_{t+1}} \big) \Big] I_t^m = (1+r)\beta E_t V_{t+1}^W p_m \tag{II.7}$$

While full comparative dynamics in a three-state optimal control model are not easily computed, we can obtain accessible insights that inform our empirical exercise from a comparative static at the cost of some non-trivial simplifying assumptions. Since our purpose here is purely expository, we simplify the dynamic model to only two periods. In doing so, we appeal to the Principle of Optimality (Caputo, 2005, p.81) which suggests that any decision along an optimal path must also be optimal with regard to the state variables that resulted from prior optimal decisions. In other words, we can look at just a one period decision along an optimal path and consider the initial reference health state that was optimally determined in the prior period to be exogenous for the one-period decision. Thus, for the purpose of the comparative static, the individual enters time t knowing her health, H_t, and reference health, R_t, which was optimally determined in the prior period. She chooses the combination of consumption and medical care $[z_t, m_t]$ and subsequently receives a health shock prior to the start of the next period. Simplifying II.7 to only two periods, t and t+1, leaves a modified equilibrium condition for medical care demand:

$$m_t^* : E_t\Big[U_{t+1}^H\Big] I_t^m - U_t^z p_m = 0 \tag{II.8}$$

The total derivative of equation II.8 with respect to m_t and R_t can be arranged to form the comparative static change in medical care demand resulting from a change in reference health:

$$\frac{\partial m^*}{\partial R_t} = -\frac{E_t\left(U_{t+1}^{HR}\frac{\partial R_{t+1}}{\partial R_t}\right) I_t^m - U_t^{zR} - U_t^{zz} p_m \frac{\partial z_t}{\partial R_t}}{E_t U_{t+1}^H I_t^{mm} - U_t^{zz}\left(\frac{\partial z^*}{\partial m_t}\right) p_m} \tag{II.9}$$

This comparative static illustrates how reference health affects the demand for medical care and motivates our empirical analysis. First, note that the sign of the comparative static takes the sign of the numerator because of the preceding negative sign and the negative denominator.[10] Our focus is on the first term in the numerator with the cross-partial of health and reference health. If individuals consider reference health in their demand for medical care, they do so with respect to the marginal utility of health. The presence of the possible relationship between reference health and consumption in the comparative static further motivates our joint estimation of the demands for consumption and medical care with all elements of the model (health, reference health and consumption) in each demand equation. The sign of the cross partial of utility between consumption and reference health is not known by assumption and is therefore an empirical question. Regardless of the sign of U^{zR}, this effect is unlikely to dominate as the second and third terms in the numerator will partially offset each other.[11]

To summarize the empirical predictions of the model: if $U^{H,R} > 0$ as we have hypothesized, then the effect of reference health on medical care demand will be positive as well. Empirically we evaluate $\frac{\partial m^*}{\partial R_t}$, which we treat as a referendum on the sign, magnitude, and significance of $U^{H,R}$. Furthermore, the comparative static suggests that the effect of reference health will be greater at higher levels of medical care spending. At high levels of medical care, I^{mm} in the denominator will be a smaller magnitude, thereby magnifying the effect of reference health on the demand for medical care. Our

[10]The negativity of the denominator follows from the assumptions of concavity of the investment and utility functions and the binding budget constraint.

[11]U^{zR} and $\frac{\partial z_t}{\partial R_t}$ will have the same sign. Given that U^{zz} is negative by concavity, the second and third terms will be of opposite signs.

full distribution estimator enables us to test this hypothesis by flexibly permitting the marginal effect of reference health to vary over the distribution of medical care spending. Thus, including reference health in the model can help explain why individuals with the same level of contemporaneous health nonetheless demand different amounts of medical care.

There are other plausible explanations for a positive association between reference health and the demand for medical care, including that reference health may affect the demand for medical care through the health production rather than the utility function. Our theoretical model assumes that contemporaneous health encompasses all necessary information for the purpose of health production. We relax this assumption by including reference health in our empirical specification of the evolution of health and testing for its significance. We also conduct several reduced form tests for alternative explanations for why reference health may be positively correlated with medical care spending in the next section. We find no empirical evidence that any of these alternative explanations drive our primary results.

III Data and Descriptive Evidence

The data for our empirical analysis are taken from the RAND files of the Health and Retirement Study (HRS), a longitudinal biennial panel survey of individuals 50 years old and over, from 1992-2010. Several features of the RAND HRS data set make it ideal for investigating the effect of reference health on the demand for medical care. First, the HRS data contain all relevant variables (detailed health data, out of pocket medical care spending, income, wealth, etc.) over a 20 year sample period of sufficient length to capture the dynamic evolution of health, demand for medical care, and consumption. Second, using HRS data avoids the insurance-induced complications of within-year spending variation (e.g. individuals crossing deductibles) that must be addressed in more high-frequency data, such as Medical Expenditure Panel Survey or Survey of Income and Program Participation.

Finally, the fact that the HRS is comprised of older individuals yields four additional advantages for our empirical work. First, older individuals are more likely to exhibit differences in their overall health trajectory, whereas 40 year olds are likely to exhibit highly stable health trajectories for extended periods. Second, the fact that individuals in our sample are near or past retirement age reduces concerns about the investment benefits of health or the endogeneity of income with respect to health. Third, concerns about insurance choices are at least partially mitigated by Medicare. Finally, an alternative explanation for our primary finding is that individuals with higher reference health may be less experienced with the health care system, and therefore less educated and efficient consumers of medical care. While we provide empirical evidence against this hypothesis, the age distribution of our sample also makes this alternative explanation less likely.

Our empirical model, detailed in Section IV, requires two periods of initial conditions. We therefore restrict the sample to those individuals in the primary cohort who are observed for at least three periods, yielding a sample of 173,312 observations of 25,872 individuals. We observe individuals for an average of 7 waves (14 years). Tables I and II display summary statistics.

The RAND HRS includes two measures of medical care expenditures: out of pocket medical care expenditures and total medical care expenditures. We use out of pocket medical care expenditures for three reasons. First, total medical expenditures were only reported for the first six waves, cutting our effective sample by more than half. Second, the 'total medical expenditure' question asked respondents to recall the total amount incurred, with no external validation. As insured individuals are notoriously insulated from their true total costs, these responses were simply unreliable.[12] Third, while total expenditures are certainly policy relevant, the out of pocket expense is the more relevant cost to the individual. Even among insured individuals, those over (under) 65 still spend approximately 16 percent (10 percent) of disposable income on medical care (Desmond, Rice, Cubanski, and Neuman, 2007; Banthin and Bernard, 2010).

The RAND HRS files include many discrete categorical variables for level of difficulties with activities of daily living (ADL's), instrumental activities of daily living (IADL's), chronic conditions, self-reported health, and other data on the respondent's health. While each of these measures is important, no single measure provides a complete picture of the individual's health state. Since health is both a dependent variable and an explanatory variable, we need a measure of health that is as comprehensive as possible to avoid omitted variable bias. Preferably, our measure will be continuous, as discrete responses may introduce significant error into the dynamics of our model.[13]

[12] Reported total medical expenditures has 16 times the variance of out of pocket medical expenditures. The two variables share a correlation coefficient of 0.14

[13] Consider an individual who self reports health on a five point Likert scale with latent values of 4.3, 4.0, 3.7, 3.4 in periods 1-4. The discrete measure of self-reported health would likely be reported as 4, 4, 4, 3. Relying just on discrete values of self-reported health rather than our continuous index would result in erroneous calculation of health, reference health, and the dynamics of health. We would find that health is unchanging through the first three periods, then falling during the fourth. All the effects of mismeasuring our dependent variable would be imparted on the covariates used to calculate transition probabilities, leading to biased estimates. See Contoyannis, Jones, and Rice (2004) for a full discussion of the problems associated

We therefore convert these discrete categorical variables into a single continuous measure of health using Multiple Correspondence Analysis (MCA).[14] Table III contains the variables and weights used to construct the health variable used in estimation.

Figure 1 shows the distribution of the difference between contemporaneous and reference health values in our sample.[15] As our sample is comprised of older individuals, it is unsurprising that the median value of the change in health from reference health is negative. At the median, individuals' health declines by 1.6 percentage points per year from reference levels. As the figure shows, large negative changes are more common than large positive changes yielding a mean change of -3.2 percentage points. While individuals' health tends to decline overall, there are also plenty of instances where individuals in the sample experience improvements in their health relative to recent health history.

Table IV provides descriptive evidence that medical care spending does indeed vary with the change in health from reference health. We divide health and reference health into quintiles and calculate the mean change in log medical care spending for each quintile combination. For each quintile of health, the mean change in log out of pocket medical care expenditures is monotonically increasing in reference health. In every row of Table IV, the fifth column (containing individuals in the top quintile of reference health) has the greatest change in out of pocket medical expenditures, followed by the fourth column, etc. Note that for each quintile of reference health, the mean change in log out of pocket expenditures is monotonically decreasing in health. The single largest value in the table is in the cell where individuals were in the top quintile of reference health, but the lowest quintile of contemporaneous health. This descriptive evidence is consistent with our theory that reference health provides relevant information about the demand for medical care.

We calculate consumption of the non-medical good by subtracting the change in non-housing financial wealth and out of pocket medical expenses from income. This calculated variable represents consumption-net-of-savings. However as most individuals in the data set are approaching or past retirement age, dissaving is more common than saving. Additionally, we cannot capture the effects of capital gains or losses.[16]

with self-assessed health.

[14]MCA is used to transform discrete variables into a single continuous variable, whereas Principal Components Analysis is used to transform continuous variables into a single continuous variable. See Kohn (2012) for a full description of the MCA health index methodology.

[15]We operationalize reference health as the average of the first two lags of health. Section B Table V contains sensitivity analysis regarding this specification.

[16]The median person in our sample has non-housing financial wealth of \$10,500 and a person in the 90^{th} percentile of wealth has \$250,000 in non-housing financial assets. Except for the upper tail of the wealth distribution, unobserved capital gains are a minimal concern in calculating consumption.

We estimate the joint conditional distribution of three variables: health, out of pocket medical care expenditures, and consumption of an aggregate non-medical care good. From the descriptive statistics and the graphs of the kernel densities of these variables (see Figure 2), the distributions of medical care expenditures and non-medical consumption are skewed right, but the distribution of the health index is skewed left.[17] Most people are fairly healthy, and medical care expenditures and consumption are driven by right-tailed income and wealth distributions. In our sample, the top 5 percent of medical care consumers account for 46 percent of all medical care spending, consistent with the population observation that the top 5 percent account for nearly half of medical care spending. These skewed distributions underscore the importance of modeling the full distributions, rather that just estimating the conditional mean.

A Conditional Mean Results and Evidence on Alternative Explanations

Prior to detailing our full empirical specification (see section IV and equation IV.1), we provide evidence that higher levels of reference health are associated with greater medical care spending at the conditional mean. These results demonstrate that our full model results are not artifacts of our econometric methods. Next, since there can be no natural experiment or suitable instrument for reference health, we consider an extensive set of alternative explanations to augment our econometric identification. While our motivation focuses on the effect of reference health through the marginal utility of health, there are clearly several alternative mechanisms by which health in previous periods could affect the demand for medical care. These alternatives include: the effect of reference health is actually the effect of a health shock or otherwise attributable to initial diagnosis and treatment costs of new conditions; the effect of relative inexperience with the medical care system; reference health is simply correcting for measurement error in contemporaneous health; or reference health actually affects medical care expenditures through the production function. We find no evidence to support any of these alternative explanations.

Table V contains results from a regression of log out of pocket expenditures on various specifications for reference health. While we only report the coefficients for contemporaneous health and our specifications of reference health, these regressions also control for lagged out of pocket medical expenditures, income, wealth, demographic

[17] There is considerable multi-modality at the top of the health index distribution as many individuals report having just one ADL, or one chronic health condition, or “very good” health.

variables, insurance status, and random effects as in our full model (equation IV.1).[18]

Regardless of the specification, the effect of reference health on current medical care spending is always positive and significant. The result from the specification of reference health employed in the full dynamic model (average of past two periods) is shown in column 6. The estimated coefficient on this specification of reference health is smaller than the coefficient estimated using a three period average (column 7). Because of concerns about initial conditions, sample size and a desire not to overstate our results, we have chosen the specification for reference health that yields the smallest effect. Thus, while the true construction of reference health is unknown, these results provide evidence that the variation of interest in our full model is present in a conditional mean estimation, and that our results are not an artifact of our empirical specification of reference health.

Is the 'reference health' effect due to individuals experiencing sudden health shocks or initial costs of treating new health conditions?

The results in Table V are not consistent with the health shock or initial diagnosis/treatment cost story, but they are consistent with the reference health hypothesis. Consider the results in columns (1)-(4) with lags up to $t-4$. Recalling that our data are biennial, these lagged values of health are therefore health two, four, six, and eight years prior to the period in which the individual decides how much medical care to consume. Suppose that an individual is diagnosed with a new condition that requires large initial expenditures for diagnosis and treatment. Suppose further that it is these initial costs that create the empirical relationship between 'reference health' and contemporaneous expenditures on medical care. We might expect that among two individuals in the same contemporaneous health, the one with the higher levels of *recent* prior health might have larger expenditures if she is in the initial phase of diagnosing or treating this new condition.

If we examine the results in column (1), that story seems plausible because the coefficient on the first lag of health is positive and significant. However, if initial treatment costs rather than reference health are driving our results, higher health in the most recently observed period should always positively affect medical care expendi-

[18]Including fixed effects in regressions with lagged dependent variables requires a dynamic panel estimator (e.g., Arellano-Bover-Blundell-Bond). We have checked that our results still hold under dynamic panel specifications. We are aware of the complications of LDV regressions and emphasize that these results are intended to be descriptive in nature and motivate our full empirical model rather than stand alone as causal estimates.

tures even when we add higher order lags of health. Moreover, the higher order lags should be insignificant if the effect of reference health is limited to the onset of *new* conditions. Examining the results in columns (2), (3), and (4), we observe the opposite effect. When we add higher order lags, the first lag of health becomes insignificant while higher order lags are highly significant. As further evidence, we hypothesize that the individual's reference health may be determined by her highest value of health as in column (5).[19] When the reference point is defined by the maximum value of lagged health, the coefficient on H_{t-1} is insignificant. The first lag of health, which should capture the onset of new conditions, is not statistically significant in any of these specifications. Thus, this table provides evidence that the effect of interest can be attributable to reference health, but not initial costs of treatment or effects of sudden shocks.

Is the effect of reference health operating through the production function?

Another alternative hypothesis is that reference health enables consumers to realize greater gains from medical care. In other words, contemporaneous health (which we do include in our theoretical health production function) does not tell us everything we need to know about the effect of health on health production. For example, perhaps those who have been in good health consistently for a long period of time are healthier with respect to co-morbidities or other complicating health factors that are not captured in contemporaneous health. In this case, both contemporaneous health and reference health should significantly increase the marginal productivity of medical care in the health production function. Consider the linear analog to our expression for the evolution of health in the full model, IV.2:

$$H_{t+1} = \beta_0 + \beta_1 H_t + \beta_2 R_t + \beta_3 m_t + \beta_4 H_t m_t + \beta_5 R_t m_t + \beta_6 \mathbf{X}_t + \epsilon_t \tag{III.1}$$

If reference health increases the marginal productivity of medical care (leading to increased medical care spending), the coefficient on the interaction term, β_5, should be positive and statistically significant. Additionally, reference health may influence medical care spending by affecting expectations about health in future periods. For example, an individual with health and reference health values of (A, B) respectively,

[19]These values typically occur in the first or second period we observe an individual due to the age composition of the sample. This is consistent with Baucells, Weber, and Welfens (2011). The use of the highest observed value as a reference point is also consistent with the literature on salience. (Bordalo, Gennaioli, and Schleifer, 2013)

where $B > A$, may believe her health will continue to worsen and therefore consume more medical care. If higher levels of reference health are associated with lower expected future health, β_2 should be negative and significant.

Neither of these alternative explanations to the utility mechanism are supported by our results. Table VII contains the relevant results of a dynamic system estimator (Arellano-Bover-Blundell-Bond) of equation III.1, the linear analogue to our expression for the evolution of health in the full model.[20] The coefficient on $R_t m_t$ is negative and insignificant for all specifications of reference health, implying reference health does not affect the marginal productivity of medical care. Additionally, the coefficient on R_t is positive and significant, implying that higher levels of reference health, even controlling for medical care expenditures, are associated with better expected health in the next period. This finding directly contradicts the notion that reference health increases medical care spending through the expectations channel.

These results imply that if operating solely through the production function, higher levels of reference health would lead individuals to demand *less* medical care. In our theoretical model, we assume that utility is concave with respect to health, meaning healthier individuals receive less marginal utility from improving their health. As β_2 is positive, higher levels of reference health predict better health in the next period, meaning that individuals with higher reference health expect less marginal utility from improving their health. As β_5 is negative and insignificant, individuals with higher levels of reference health do not experience greater marginal productivity from medical care. Because higher reference health does not increase the health returns to medical care spending, but does lead to decreased marginal utility from health gains, the total effect of the production mechanism of reference health on medical care demand is negative. As our estimated effect of reference health on the demand for medical care is positive, these results lend additional support to the hypothesis that higher reference health increases the demand for medical care through the utility function.

While these single equation models do not address the endogeneity of medical care expenditure to the extent of our full model, they do support our inclusion of reference health in the production function in our full model.[21] We discuss the effect of reference health through the production function when discussing the results of our full model.

Is reference health merely correcting measurement error in contemporaneous health?

[20] We must use a dynamic system estimator in the health production specification because reference health is a function of lagged dependent variables. Additionally, m_t is itself a function of higher order lags of the dependent variable.

[21] R_t is included as a stand alone argument, but is not interacted with m_t in our full model.

If reference health merely adds missing information to contemporaneous health, then we would expect reference health and contemporaneous health to have the same negative effect on medical care spending. Suppose that true health, H_t^*, is unobserved. Instead, we observe H_t, which is measured with error. If R_t provides additional information about true health, H_t^*, then we would expect that individuals with higher values of reference health ($H_t < R_t$) are in greater true health than individuals for whom $H_t = R_t$. As such, we would expect that individuals with greater levels of reference health would be healthier and therefore demand less medical care.

We find the opposite. Referring again to Table V, reference health and contemporaneous health have different signs. Similarly, if contemporaneous health is measured with error, values of health recorded closer to the current period should provide more missing information about contemporaneous health than older measures. Again Table V shows the opposite. Specifications of reference health using observations more distant in time retain comparable levels of information as those closer in time as indicated by the size and significance of the coefficients. In fact, the specification of reference health that uses the largest observed value of health (typically a more distant value) has the largest estimated magnitude, while the first lag is small and insignificant.

Is the effect of reference health attributable to individuals with higher reference health being less efficient consumers of medical care?

Consider two individuals who are both in good health. Until recently, one of them was in great health, and we observe that person spending more on medical care than his counterpart. Is this a 'reference health' effect, or is it simply that the person who was until recently in great health is a less savvy consumer of medical care? First, note that our results in Table V do control for lagged medical care spending. We can further investigate this "efficiency" or "learning" alternative hypothesis by splitting our sample into groups on the basis of medical care expenditures last period and comparing the estimated effects of reference health on medical care spending this period. If inexperience with the system explains our reference health effect, we should see the reference health effect concentrated among individuals who were small or non-consumers of medical care spending last period. Similarly, if experience is driving this result, we should see a non-result (insignificant effect) of reference health on medical care spending among individuals who were already spending some money on medical care and would presumably understand something about utilization practices.

Our results do not support the "learning-to-use-the-health-system" explanation.

Table VI reports the coefficients and standard errors of the estimated effects of reference health as specified in columns (5)-(7) of Table V when we split the estimation sample by amount spent on medical care last period. Out of pocket medical care spending here is our proxy for experience with the medical system. The lowest tercile spent less than $400 out of pocket in the prior year. The upper tercile spent over $1,800 out of pocket in the same time frame. The estimated effect of reference health, regardless of specification, does not statistically change in magnitude or significance when considering individuals with little or substantial experience with the medical care system. In fact, when comparing coefficients in the first and third rows, the effect of reference health among 'less experienced' individuals is not statistically significantly different from the more experienced individuals in the upper tercile. Finally, recall that our sample is comprised of individuals near or past retirement age, who are likely familiar with the health care system.

In summary, we have shown that there is a relationship in the conditional mean of a simple lagged dependent variable regression between reference health and contemporaneous out of pocket medical care spending, conditioning on the set of covariates used in our full model. We have shown that this relationship is robust to specification changes for reference health. As the results vary with specification changes, they do so in ways that are consistent with the reference health mechanism. We have shown that several reasonable alternative explanations are not supported by the data. Furthermore, we have chosen a specification of reference health that yields the smallest effect at the conditional mean. We are therefore confident that the results from our full model represent a lower bound on the true effect of reference health on the demand for medical care.

IV Econometric Methodology

Our discussion proceeds in four stages. First, we present our estimating equations and show that they follow from our theoretical motivation. Second, we discuss econometric identification. Third, we provide a brief explaination of conditional density estimation. Finally, we incorporate discrete factor random effects, our mechanism for addressing the non-independence in the econometric errors.

A Estimating Equations

In keeping with the theory, the individual makes a joint choice of medical care and consumption to maximize expected discounted lifetime utility. She does so aware of her demographics, contemporaneous and reference health, and the functional forms of utility and the health production function. Because reference health is an argument in the utility function, reference health is included in the expressions for both the demand for consumption and the demand for medical care.

The empirical joint demand for consumption and medical care is drawn from equation II.5 with a few necessary innovations. First, we add lagged medical care and consumption. Adding lagged medical care (and consumption) to the empirical model is necessary to capture the dynamic nature of the model since the individual's choice of consumption at any time, t, is a function of past decisions. Second, we add econometric errors into the expressions for z_t^* and m_t^*. Thus, the individual's demand for consumption and medical care can be expressed as:

$$\begin{aligned} z_t^* &= z(H_t, R_t, W_t, Y_t, m_{t-1}, z_{t-1}, \mathbf{X_t}, \epsilon_t^z) \\ m_t^* &= m(H_t, R_t, W_t, Y_t, m_{t-1}, z_{t-1}, \mathbf{X_t}, \epsilon_t^m), \text{ where} \end{aligned} \tag{IV.1}$$

where $R_t = \frac{H_{t-1}+H_{t-2}}{2}$. The individual's choices for z_t^* and m_t^* affect health at the start of the next period:

$$H_{t+1} = \alpha(H_t, R_t, m_t, z_t, \mathbf{X_t}, \epsilon_t^H) \tag{IV.2}$$

This empirical health transition equation follows the theory with the addition of the econometric error and reference health. Recall that the theory assumed that contemporaneous health contained all necessary information for health production, and our single equation tests confirmed that reference health appears to increase the marginal productivity of medical care. Our single equation results did, however, imply that higher levels of reference health predicted better health in the next period. We therefore include reference health in our full econometric model to evaluate the importance of this mechanism.

B Econometric Identification

The econometric identification of this model comes through three sources: exclusion restrictions, timing, and non-linearities in the functions. Table VIII contains the full list of variables included in each expression. For the health transition expression, the information from the individual's lagged consumption of medical and non-medical goods is captured by the individual's contemporaneous health, consumption and medical care spending, which are themselves functions of past consumption decisions. Conditional on the individual's realized contemporaneous health state, only contemporaneous consumption of medical and non-medical goods and a stochastic shock determine the individual's health in the next period. Similarly, the individual's income and wealth should affect her demand for medical and non-medical goods, but not otherwise affect her health in the next period.

To appeal to the timing assumption, we must specify endogenous initial conditions for the health states that form the reference health and initial demand for medical and non-medical goods:

$$\begin{aligned} H_1 &= H_1^i(\mathbf{X_1}, \mathbf{X^i}) \\ H_2 &= H_2^i(\mathbf{X_2}, \mathbf{X^i}, H_1) \\ m_2 &= m_2^i(\mathbf{X_2}, \mathbf{X^i}, H_1) \\ z_2 &= z_2^i(\mathbf{X_2}, \mathbf{X^i}, H_1) \end{aligned} \tag{IV.3}$$

$\mathbf{X^i}$ is a set of variables only included in the initial conditions: whether the respondent was a veteran, the respondent's number of living parents, the current/final age of those parents, and a vector of occupational stress measures.[22] Ceteris paribus, individuals who worked in occupations which were more physically demanding, required heavy lifting, or exposed them to more environmental risk should have worse health and consume more medical care at the time we first observe them. Details on the construction of these occupational stress measures are discussed further in Appendix A.

C Conditional Density Estimation

We employ Conditional Density Estimation (CDE) to estimate the joint distribution of medical care expenses, consumption, the health transition process, and initial conditions (Gilleskie and Mroz, 2004). CDE provides three advantages over conditional mean estimation. First, CDE enables us to match any moment of the distribution of each variable. Second, CDE does not require parametric assumptions on the distribution of the error terms, enabling us to flexibly model left or right skewed distributions.[23] Third, CDE permits the marginal effect of explanatory variables (including reference

[22] All variables included in the expressions for m_t, z_t, H_t when $t > 2$ are also included in the initial conditions.

[23] Section III shows that both medical care spending and consumption are left skewed while health is right skewed.

health) to vary over the support of the dependent variable for each equation in the model. Thus, we can capture whether the effects of reference health are stronger in the top tail of the medical care spending distribution or at the extensive margin. Similarly, we can compare the marginal effects of contemporaneous and reference health in different regions of the distribution of non-medical consumption.

CDE utilizes a sequence of conditional logit probability functions to approximate the density of the outcome of interest. First, we divide each dependent variable, y, into K quantiles containing equal numbers of observations in each cell.[24] For each interval, the k^{th} interval is defined by $[y_{k-1}, y_k)$. We define y_0 as the smallest observation and $y_K = \infty$.

The conditional probability that the dependent variable is observed in the k^{th} interval, given that it is not observed in intervals 1 through $k-1$ can be expressed as:

$$\lambda(k,x) = p[y_{k-1} \leq Y < y_k | x, Y \geq y_{k-1}] = \frac{\int_{y_{k-1}}^{y_k} f(y|x)dy}{1 - \int_{y_0}^{y_{k-1}} f(y|x)dy} \tag{IV.4}$$

Thus, the $\lambda(k,x)$ serves as a discrete hazard function, given the cut off points k, the upper and lower bounds on Y and covariates x. As a hazard function, the probability that Y falls into the k^{th} interval is given by:

$$p[y_{k-1} \leq Y < y_k | x] = \lambda(k,x) \prod_{j=1}^{k-1} [1 - \lambda(j,x)] \tag{IV.5}$$

As suggested by Gilleskie and Mroz (2004), we use a sequence of logit probabilities to form the hazard function, and thus the probability that our random variables of interest fall into a given cell. Additionally, we interact each covariate x with a function of the interval number, $\gamma_k = -ln(K-k)$ and γ_k^2. These interactions between the γ_k terms and the covariates are what permit the marginal effect of the variable of interest to vary over the support of the dependent variable. For each expression in equations IV.1-IV.2 and for each cell $k \in \{1, \ldots, K\}$, we can form a linear function of our covariates and γ terms:

$$g^j(k,x) = \mathbf{X^j}\beta_1^j + \mathbf{X^j}\gamma_k\beta_2^j + \mathbf{X^j}\gamma_k^2\beta_3^j + \epsilon_t^j \; \forall j \in \{z, m, H, H_1, H_2, z_2, m_2\} \tag{IV.6}$$

[24] The optimal number of quantiles, K^*, can be determined empirically as discussed in Gilleskie and Mroz (2004). We have verified that our results are not sensitive to the number of quantiles used. Additionally, while the distributions of non-medical consumption, medical care, and the health index are multi modal, they are sufficiently continuous to where single values do not 'straddle' quantile boundaries. See Figure 2 for a visual representation

With some abuse of notation, $\mathbf{X^j}$ includes of all variables in expression j. This function is quadratic in the function of the cell indicators, γ. The order of the polynomial determines how many times the sign of the marginal effect of x_k on the hazard probability can change over the support of the dependent variable.[25] From equation IV.6, we can therefore form the logit probabilities used to form the hazard function:

$$\lambda^j(k,x) = \frac{e^{g^j(k,x)}}{1+e^{g^j(k,x)}} \tag{IV.7}$$

and these terms are subsequently combined to form the estimable probabilities in equation IV.5.

D Discrete Factor Random Effects

We must address two likely issues with the econometric errors in our joint dynamic model. First, as discussed in the theoretical model, there is likely to be persistence in the outcomes and the unobservable factors that influence medical care consumption, non-medical consumption, and health dynamics (French and Jones, 2004; Cohn and Yu, 2012; Kohn and Liu, 2013). Second, the errors of these expressions are almost certainly dependent for two reasons. There is likely to be correlation between the persistent aspects of the error terms, and there is likely correlation in the variation of observed medical care, non-medical consumption and health around the overall persistent trends.

Therefore, for each expression, we utilize a flexible random effects estimation technique that permits time-invariant and time-varying unobserved heterogeneity without imposing distributional assumptions on the error term. We approximate the joint distribution of both permanent and time-varying unobservables with a step function (Heckman and Singer, 1984). In Monte Carlo simulations, the discrete factor random effects estimator has been shown to reduce bias relative to the assumption of joint normality in the distribution of unobserved heterogeneity (Mroz, 1999).

We include the time-invariant, permanent unobserved heterogeneity component for each expression and allow these time-invariant components to be correlated with one another. As our expressions include lagged dependent variables, these permanent components capture persistence in the error terms rather than heterogeneity in levels. For example, individuals who heavily value the future may be likely to consistently invest in more medical care, engage in lower consumption, and enjoy persistently smaller

[25]Including higher orders yields greater flexibility in estimating the marginal effects over the distribution at a cost of increasing the number of parameters to be estimated. We have verified that our results hold in linear and cubic polynomial functions of the cell indicator. As expanding to a cubic polynomial did not improve performance, we have reported the results from the quadratic specification.

declines in health. Alternatively, individuals who are genetically predisposed to poor health may consume ever increasing amounts of medical care and yet experience more rapidly deteriorating health. The time varying component of heterogeneity is meant to capture changes that affect unobservable factors on a per-period basis. These time varying components are assumed independent of the permanent components and are designed to capture correlation in the idiosyncratic variation around the overall trend. We therefore decompose the errors in each expression into three components:

$$\epsilon_t^j = \mu^j + \nu_t^j + e_t^j \qquad \forall j \in z, m, H \tag{IV.8}$$

where μ^j captures the permanent heterogeneity for each expression, ν_t^j captures the time-varying component, and e_t^j represents the remaining i.i.d. Type-1 Extreme Value error necessary to formulate the logit hazard probabilities. Errors for the initial conditions expressions do not include time-varying heterogeneity.

The likelihood function includes eight expressions: the per-period demand for medical care, non-medical consumption, the health transition equation, a per-period probability of death, two initial conditions equations for health (initial health and second period health, in order to formulate reference health) and initial conditions for the demand for medical care and consumption. The full estimation procedure consists of a joint CDE estimation nested in a finite mixture framework. The individual's contribution to the likelihood function is:

(IV.9)

$$\begin{aligned}
L_i(\Theta, \Psi, \Pi) = \sum_{k=1}^{K} \pi_k \Bigg[& \prod_{j_{h1}=1}^{J_{h1}} P(H_1 = j_{h1} | \mu_k^{H_1})^{1(H_1 = j_{h1})} \prod_{j_{h2}=1}^{J_{h2}} P(H_2 = j_{h2} | \mu_k^{H_2})^{1(H_2 = j_{h2})} \\
& \times \prod_{j_{m2}=1}^{J_{m2}} P(m_2 = j_{m2} | \mu_k^{m_2})^{1(m_2 = j_{m2})} \prod_{j_{z2}=1}^{J_{z2}} P(z_2 = j_{z2} | \mu_k^{z_2})^{1(z_2 = j_{z2})} \\
& \times \prod_{t=3}^{T_i} \sum_{l=1}^{L} \psi_l \Big[\prod_{j_m=1}^{J_m} P(m_t = j_m | \mu_k^m, \nu_{lt}^m)^{1(m_t = j_m)} \prod_{j_z=1}^{J_z} P(z_t = j_z | \mu_k^z, \nu_{lt}^z)^{1(z_t = j_z)} \\
& \times \prod_{j_H=1}^{J_H} P(H_t = j_m | \mu_k^H, \nu_{lt}^H)^{1(H_t = j_H)} \prod_{D=0}^{1} p(death = D | \mu_k^D, \nu_{lt}^D)^{1(death = D)} \Big] \Bigg]
\end{aligned}$$

where Θ is the vector of parameters to be estimated from equations IV.1, IV.2, and IV.3; ψ_l are the mixing parameters for the time-varying heterogeneity, and π_k are the mixing parameters for the permanent heterogeneity. K and L represent the number of mass points for the distribution of permanent and time-varying heterogeneity, respectively, and t indexes the waves in the data for each individual. The terminal time, specified T_i, reflects that not all individuals are observed in the sample for the same number of periods. $J_{h1}, J_{h2}, J_{m2}, J_{z_2}, J_z, J_m$, and J_H are the number of cells for each conditional density estimation. The model is estimated in parallel using MPI, employing full information maximum likelihood methods with a BHHH algorithm. We have estimated the model with three mass points in the support of permanent heterogeneity and two points of time-varying heterogeneity.[26]

V Results

A Full Model Results

Tables IX and X report marginal effects.[27] Because the estimates are for parameters in non-linear hazard probabilities, they are not directly interpretable without numerical simulation.[28] Recall that one benefit of using CDE is that we can estimate marginal effects at different points of the distribution of the dependent variable. We report three marginal effects: one for the bottom quartile of the dependent variable, one for the top quartile, and a third for the interquartile range. Each marginal effect is the mean percentage change in the value of the dependent variable, conditional on the dependent variable being observed in that quartile. Some of the variation in these marginal effects stems from explanatory variables having different local effects on the probability of an outcome being observed in a given part of the distribution. For some variables, the marginal effect is positive in the low end of the distribution of the dependent variable, and negative in the top portion of that same distribution, or vice versa.

For interpretation, consider the marginal effect of having insurance on medical care consumption (left column of Table IX). The marginal effect of having insurance on out of pocket medical care expenditures is negative (-3.1 percent) among those observed in the lowest quartile of medical care consumers. The marginal effect of having insurance on out of pocket medical expenditures is still negative in the interquartile range (-2.6 percent), but the marginal effect of having insurance is positive in the top quartile

[26]Adding additional mass points does not significantly increase the likelihood function, as vetted with an LR test (p-value = 0.21).

[27]Marginal effects are calculated by replicating each observation in the data 80 times. We forward simulate the data using the observed values of the exogenous variables and estimated parameters of the model. We then change a variable of interest, and re-simulate the data still using the estimated parameters of the model and holding all other exogenous variables fixed. Continuous variables are increased by adding 10 percent to their previous value. Marginal effects of binary variables are calculated by simulating twice - once with all observations set to zero, once with all observations set with a value of one.

[28]Parameter estimates are therefore included in a separate appendix available from the authors.

of medical care consumers. We interpret this as evidence that the price elasticity of medical care demand may be increasing in individuals' medical care expenditures. Alternatively, those who spend the most on care may also face the most binding budget constraint. For another example, the effect of education on medical care expenditures is negative across the distribution, consistent with education increasing health productivity. However the magnitude of the marginal effect of education increases from -1.0 percent at the bottom quartile to -11.5 percent at the top. This monotonic but non-linear effect over the distribution is intuitive as those who spend more have more to gain (or in this case save) by being more savvy and productive consumers of medical care.

B Empirical Evidence on the Effects of Reference Health

B.1 Results for Medical Care Expenditures

Controlling for contemporaneous health, reference health significantly affects medical care spending. By simulating the model, we find that conditional on contemporaneous health, an individual with a 10 percent higher value of reference health is 22 percent more likely to be in the top 5 percent of medical care spending. Over the entire distribution of medical care spending, a 10 percent higher value of reference health relative to contemporaneous health, $(R_t - H_t)$, increases expected medical care consumption by 14 percent. Recall from Section II that $U^{H,R} > 0$, suggesting that an increase in reference health increases the marginal utility from health, thereby increasing medical care spending. Our findings support our theoretical hypothesis. Relative to contemporaneous health, a one percentage point increase in reference health increases the probability of being in the top 5 percent by 2 percent, and increases expected medical expenditures by 1.3 percent. However, a 20 percentage point increase in reference health leads to a substantially larger 48 percent probability of being observed in the top 5 percent, and a 31.2 percent increase in medical care expenditures. The marginal effects of lower reference health are smaller in magnitude than those of higher reference health. A 10 percentage point reduction in reference health decreases the probability that an individual is observed in the top 5 percent by 19 percent, and overall medical expenditures decrease by 13 percent.

For perspective, we compare the effects of reference health to the pure effect of contemporaneous health alone. If we decrease contemporaneous health and reference health by 10 percent each (thereby holding $R_t - H_t$ constant) the probability that an individual is observed in the top 5 percent of the medical care spending distribution increases by 44 percent. Therefore, while contemporaneous health is clearly important, reference health compounds the effect of poor health on medical care spending.

B.2 Reference Health and the Production of Health

Using single equation methods, we showed in Section A that there was no evidence to suggest that reference health affected the marginal productivity of medical care. While there was a positive effect of reference health, independent of medical care, on health in future periods, this result supports our hypothesis that reference health influences the demand for medical care through the utility function. In our full empirical model, the estimated effect of reference health on health in future periods is still positive, but the estimated marginal effects are much smaller. In the interquartile range, a 10 percentage point increase in reference health increases expected H_{t+1} by 1.4 percent. This effect is less than one-fifth of the magnitude of the effect of contemporaneous health on H_{t+1}. This result is intuitive: health is highly persistent, and individuals who have been in historically better health are expected to have better health in the future. Among individuals who are expected to be in good health (top quartile), contemporaneous health has far more explanatory power than reference health. Among individuals expected to be in poor health, the marginal effects of reference health and contemporaneous health are of the same approximate magnitude. For individuals expected to be in poor health, a recent history of *good* health predicts better prospects for recovery.[29]

These results provide additional evidence that the positive effect of reference health on the demand for medical care operates through the utility function. While reference health likely does affect expectations about future health, remediate measurement error, and predict health in the future, all evidence suggests these mechanisms of reference health should create a *negative* effect on the demand for medical care. These results rely on the assumption that utility is concave in health and the earlier finding that reference health does not affect the productivity of medical care. Consider the hypothesis that reference health corrects for measurement error in contemporaneous health in the health transition equation. For example, take two individuals with contemporaneous health values of 0.7. One individual has a reference health value of 0.7, the other has a reference health of 0.9. Our results indicate that the person with the higher reference health will be healthier next period, independent of their use of medical care. The person with the higher reference health, by virtue of being in better health, will face

[29] As our full model reduces the magnitude of the effects of reference health on future health, the effect of $m_t R_t$ would, in all likelihood, remain insignificant.

lower expected marginal utility from consuming medical care. Because reference heath does not affect the marginal productivity of medical care, the person with greater reference health should therefore consume less medical care. However, our results indicate that $\frac{\partial m^*}{\partial R_t}$ is positive. Our results in the previous section therefore likely understate the utility-specific effect of reference health.

Consider also the hypothesis that our effect of reference health operates through the shock, or change, rather than reference dependent utility. An individual experiencing a decline in health from reference levels may feel compelled to consume more medical care lest their downward trajectory continue. Our results directly contradict this hypothesis. Consider again individuals with (H_t, R_t) pairs of $(0.7, 0.7)$ and $(0.7, 0.9)$. Our results indicate that the second individual, the one who experienced the decline, is expected to be healthier next period. Perhaps this individual is consuming more medical care, but the endogeneity of this choice is addressed by our econometric methods.

C Matching the Individuals and Demographics of the Top 5 Percent

To evaluate the impact of including reference health in our model, we again simulate the model, replicating each observation in the data 80 times. In simulation, we use the model to generate predictions and then compare the predictions of the model to the observed data.[30] For comparison, the lagged dependent variable regressions in Table V explain approximately 6 percent of the variation in medical care consumption. Using 6 percent as a bench mark, our model represents improvement if it generates a match rate that is 6 percent greater than random.

Simulating the model as described above, our model generates a 14.5 percent match rate between predicted and observed individuals in the top 5 percent of medical care consumers. We also have estimated this model under a specification which does not include reference health. Removing reference health also requires removing lagged values of medical care and non-medical consumption to avoid biased estimates if these coefficients absorb the omitted effects of reference health. This specification with only contemporaneous variables generates an 8.5 percent match rate of predicted and observed individuals in the top 5 percent of medical care spending. Therefore, our preferred specification outperforms the alternative specification by 65 percent in matching

[30]We randomly draw each replication's permanent type for all periods, a time-varying joint shock for each period, and an idiosyncratic draw from the uniform distribution. We then use the individual's exogenous variables and the estimated parameters of the model to forward simulate the individual's health state transitions and decisions to consume medical care and non-medical goods. We compare the averaged outcomes of these simulated individuals to the observed decisions and outcomes in the data.

individuals in the upper tail of the medical care spending distribution.

To gauge our model's effectiveness in fitting the full distribution of medical care consumption, we create an indicator variable for whether an individual's predicted medical care consumption is within a 10 percentile range of the individual's observed medical care consumption. For our model, this +/- 10 percentile match rate is 36 over the top quartile and 29 percent over the full distribution. When an individual's predicted and observed health indices are within 10 percentiles of one another, the match rate for the top quartile improves to 41 percent and total match rate improves to 32 percent.

Finally, we compare the summary statistics for individuals predicted to be in the top 5 percent under our specification and the above alternative specification with the observed data. These summary statistics are in Table XI. We match the observed data more closely than the specification without reference health or lagged dependent variables on most dimensions. We better match the top 5 percent on contemporaneous health. In the observed data, mean contemporaneous health among the top 5 percent of medical care consumers is 0.636, compared to our predicted mean health in the top 5 percent of 0.456 and the alternative specification prediction of only 0.26. Our predicted mean age in the top 5 percent is three years younger than the alternative model, thus, we better match young people in the top 5 percent. We also better match years of schooling, gender, and lagged medical care.

D Simulation of Medical Care Spending for Different Health Trajectories

We use the estimated parameters of two specifications of our model, one with reference health and one without, to compare medical care spending estimates. These simulations illustrate key features of the reference health model and provide evidence of the policy relevance associated with the improved estimates. Specifically, one implication of the reference health model is that individuals who experience sharp declines in health are more likely to consume an amount of medical care in the upper tail of the spending distribution. Healthy aging initiatives are therefore likely to create two types of cost savings. First, individuals will be healthier for more of their lives. Second, individuals are less likely to experience large decreases in health from their reference health levels prompting outlier medical care spending. We therefore suggest that omitting reference health from models of medical care spending will underestimate the cost savings of healthy aging, leading to Type II errors of failing to invest in healthy aging initiatives when they would result in meaningfully different spending trajectories and aggregate lifecycle spending.

We simulate medical care expenditures for three different health trajectories: a sharp, moderate and gradual decline in health. Each path starts with a health of 0.95 (corresponding to our health index values ranging from 0 - 1) and ends five periods later at a value of 0.45 while preserving a mean value of health of 0.7 over the simulation period. We simulate the model 400 times for each individual and report the average out of pocket medical care expenses (in units of $100,000) for each period at each level of health and then sum the expenses over the simulation periods. The spending estimates for each period and trajectory are reported in Table XII.

The simulation results illustrate several key features of the reference health model. First, in period one when there is no effect of reference health, there is no difference in the estimated expenses either between the models or between the trajectories. Second, looking at all of the estimates at the lowest level of health, 0.45, demonstrates the ability of the model with reference health to provide more accurate estimates. The model without reference health provides practically the same estimate for all periods and all trajectories when health is 0.045 (spending between 0.154 - 0.156). By contrast, spending estimates from the model with reference health range from 0.119 to 0.17 when individuals are in the lowest level of health. This result reflects the empirical reality that not all those in poor health are equally high spenders on medical care. Third, the highest spending comes in the period of the sharpest health decline from a health of 0.95 to 0.45 (period 4 for sharp decline) when we include reference health in the model. This result is consistent with our theory that higher reference health causes higher spending conditional on the level of contemporaneous health.

The results of our simulation illustrate the policy relevance of including reference health in models of individual decision making. First, accurate estimates of the demand for medical care are critical for budgeting, financing, and capacity planning on all levels: state, national, insurer, provider and even individual. Second, calling the gradual decline the 'healthy aging' trajectory, the simulation without reference health suggests that healthy aging reduces mean medical care expenditures by 5.8% over the simulation period. By contrast, the model with reference health predicts cost savings of 13.2% (0.466 compared to 0.537). Thus, our results indicate that modeling the savings from healthy aging initiatives without reference health will understate the impact by more than half. Population aging trends make this result highly policy relevant. Failure to understand the effect of reference health and account for it's impact on aging individuals' medical care spending will lead to substantial under investment in healthy aging initiatives.

VI Conclusion

Health is a key variable in many econometric models that estimate outcomes from retirement decisions to portfolio choice to the demand for medical care. To our knowledge, extant dynamic models that include health include contemporaneous health only. Using contemporaneous health makes an implicit assumption that the current value of observed health encompasses all relevant information for the choice being modeled. Intuitively it is reasonable to expect that the value of health, like the value of wealth modeled in other literatures, is relative to past realizations of health rather than absolute. If the value individuals attribute to contemporaneous health depends on individuals' reference health, then reference health would be a significant input to any decision that also includes health as a factor.

We demonstrate that reference health substantially affects individual's demand for medical care. Theoretically, we demonstrate that incorporating reference health in the individuals' optimization problem can explain why not all individuals in poor health demand large amounts of medical care while some in good health do. Our comparative static subsequently shows that the effect of reference health would be greater at higher levels of medical care demand. We also show that our dynamic theoretical model informs our empirical specification of the joint demands for medical care and consumption that include all elements of the model.

Empirically, we demonstrate that reference health is significant in a joint model of the demands for medical care and consumption. Our econometric methods address the challenges of estimation of a dynamic model with skewed variables of interest, persistence, and multiple sources of unobservable heterogeneity. We nest the estimation of joint demands in a finite mixture freamework that allows for both permanent and time varying unobservable heterogeneity and use a conditional density estimator which allows us to estimate marginal effects of the covariates, including reference health, at different points of the distribution of our dependent variables. Our conditional density estimation allows us to demonstrate that the effect of reference health is strongest in the top tail of the medical care spending distribution, consistent with the implication of the comparative static in our theoretical model. The marginal effect of a 10 percent decline in health from reference levels is a 22 percent increase in the probability of being in the top 5 percent. Our model including reference health posts a 65 percent improvement in matching individuals in the top 5 percent over an alternative specification with only contemporaneous variables. We offer evidence from a basic lagged dependent variable regressions that our findings are robust to different specifications of reference health. We also offer reduced form evidence that alternative explanations of health shocks, initial treatment costs, consumer inexperience, and production function concerns are not supported by the data.

From a policy perspective, our most important finding is that including reference health in dynamic models will reduce type II errors when evaluating healthy aging policies. Models that include only contemporaneous health overstate the importance of poor health in predicting high medical care spending, but completely ignore the primary benefits of healthy aging, namely making declines in health occur more gradually. Incorporating reference health in the model of medical care expenditures enables us to see that the time path of health matters in determining average costs among older individuals.

There are several interesting avenues for future work. One is to model reference health with respect to a social rather than individual measure of health. Modeling a social reference point is in line with a large clinical and economic literature on social network effects on obesity and risky behaviors such as smoking. A social reference point for health suggests potentially using a game theoretic model where an individual's choice is influenced by and in turn influences the choices of others as described by Schelling (2006). A social reference point for health may also help explain differences in medical care demand across regions and countries.

References

Banthin, J., and D. M. Bernard (2010): "Changes in financial burdens for health care: National estimates for the population younger than 65 years, 1996-2003," *Journal of the American Medical Association*, 296(22), 2712–2719.

Baucells, M., M. Weber, and F. Welfens (2011): "Reference-point formation and updating," *Management Science*, 57(3), 506–519.

Becker, G. S., and K. M. Murphy (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, 96(4), 675–700.

Bordalo, P., N. Gennaioli, and A. Schleifer (2013): "Salience and Consumer Choice," *Journal of Political Economy*, 121(5), 803–843.

Bound, J., M. Schoenbaum, T. Stinebrickner, and T. Waidmann (1999): "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, 6(2), 179–202.

Cameron, A., and P. Johansson (1997): "Count Data Regression Using Series Expansions: With Applications," *Journal of Applied Econometrics*, 12, 203–223.

Cameron, A., and P. Trivedi (1986): "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, 1(1), 29–53.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

Caputo, M. R. (2005): *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press, Cambridge, UK.

Claxton, G., R. Kamal, and C. Cox (2014): "How health spending patterns vary by demographics in the U.S.," Discussion paper, Kaiser Family Foundation, http://www.healthsystemtracker.org/2014/12/how-health-spending-patterns-vary-by-demographics-in-the-u-s/.

Cohn, S., and W. Yu (2012): "The concentration and persistence in the level of health expenditures over time: Estimates for the U.S. Population 2008-2009," *AHRQ Statistical Brief no. 354*.

Constantinides, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium," *Journal of Political Economy*, 98(3), 519–543.

Contoyannis, P., A. Jones, and N. Rice (2004): "The Dynamics of Health in the British Household Panel Survey," *Journal of Applied Econometrics*, 19, 473–503.

Crawford, V. P., and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101(5), 1912–1932.

Dai, Q., and O. V. Grischenko (2014): "An Empirical Investigation of Consumption-Based Asset Pricing," *Quarterly Journal of Finance*, 04(1).

Dardanoni, V., and A. Wagstaff (1990): "Uncertainty and the demand for medical care," *Journal of Health Economics*, 9(1), 23–38.

Darden, M. E. (2015): "Smoking, Expectations, and Health: A Dynamic Stochastic Model of Lifetime Smoking Behavior," *Working Paper*.

Deb, P., and P. Trivedi (1997): "Demand for medical care by the elderly: a finite mixture approach," *Journal of Applied Econometrics*, 12, 313–356.

Desmond, K., T. Rice, J. Cubanski, and P. Neuman (2007): "The Burden of Out-of-pocket Health Spending Among Older Versus Younger Adults: Analysis from the Consumer Expenditure Survey 1998-2003," *Medicare Issue Brief, The Henry J. Kaiser Family Foundation.*

DiNardi, M., E. French, and J. Jones (2010): "Why do the elderly save? The role of medical expenses," *Journal of Political Economy*, 112(2), 379–444.

Edwards, R. (2008): "Health Risk and Portfolio Choice," *Journal of Business and Economic Statistics*, 45, 472–485.

Ehrlich, I., and H. Chuma (1990): "A model of the demand for longevity and the value of life extensions," *Journal of Political Economy*, 98, 761–782.

French, E., and J. B. Jones (2004): "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Econometrics*, 19(6), 705–721.

Galama, T. (2011): "A Contribution to Health Capital Theory," *RAND Working Paper.*

Gilleskie, D. B. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66(1), 1–46.

Gilleskie, D. B., and T. A. Mroz (2004): "A flexible approach for estimating the effects of covariates on health expenditures," *Journal of Health Economics*, 23(2), 319–418.

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

Gurmu, S. (1997): "Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization," *Journal of Applied Econometrics*, 12(1), 225–242.

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

Heckman, 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.

Hugonnier, J., F. Pelgrin, and P. StAmour (2013): "Health and (other) asset holdings," *Review of Economic Studies*, 80(2), 663–710.

Jones, A. M., A. Laporte, N. Rice, and E. Zucchelli (2014): "A synthesis of the Grossman and Becker-Murphy models of health and addiction; theoretical and empirical applications," Health Economics and Data Group, University of York.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292.

Khwaja, A. (2010): "Estimating willingness to pay for Medicare using a dynamic life-cycle model of demand for health insurance," *Journal of Econometrics*, 156(1), 130–147.

Kohn, J. L. (2012): "What is health: a multiple correspondence health index.," *Eastern Economic Journal*, 38(2), 223–255.

Kohn, J. L., and J. Liu (2013): "The Dynamics of Medical Care Use in the British Household Panel Survey," *Health Economics*, 22(6), 687–710.

Kohn, J. L., and R. H. Patrick (2012): "Health and wealth: a dynamic demand for medical care.," Working paper, National Institute for Health Care Management, http://papers.ssrn.com/sol3/papers.cfm?abstract-id=994723.

Koszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121(4), 1133–1165.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99(3), 909–936.

Mroz, T. A. (1999): "Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome," *Journal of Econometrics*, 92(2), 233–274.

Pohlmeier, W., and V. Uhlrich (1995): "An econometric model of the two-part decision making process in the demand for health care," *Journal of Human Resources*, 30(2), 339–361.

Ryder, H. E., and G. M. Heal (1973): "Optimal Growth with Intertemporally Dependent Preferences," *The Review of Economic Studies*, 40(1), 1–31.

Schoenman, J. (2012): "The concentration of health care spending," Brief, National Institute for Health Care Management, http://www.nihcm.org/pdf/DataBrief3%20Final.pdf.

Shen, C. (2013): "Determinants of Health Care Decisions: Insurance, Utilization, and Expenditures," *Review of Economics and Statistics*, 95(1), 142–153.

Viscusi, W. K., and W. N. Evans (1990): "Utility Functions That Depend on Health Status: Estimates and Economic Implications," *American Economic Review*, 80 (3), 353–374.

Yang, Z., D. B. Gilleskie, and E. C. Norton (2009): "Health Insurance, Medical Care, and Health Outcomes," *Journal of Human Resources*, 44(1), 47–114.

Yogo, M. (2009): "Portfolio Choice in Retirement: Health Risk and the demand for annuities, housing, and risky assets," *NBER Working Paper*, 15307.

Table I: Number of Observations Per Individual

Number of Individuals	**Number of Waves Observed**
2,389	3
4,568	4
1,856	5
1,778	6
4,380	7
1,381	8
2,355	9
7,120	10
Total Individuals	Average Observations Per Individual
25,827	6.949

The relatively high numbers of individuals observed for 4 & 7
are due to HRS adding respondents at waves 7 and 4 respectively

Table II: Summary Statistics

Variable	**Mean**	**S.D.**	**Min**	**Max**
Demographic Variables				
Female	0.575	0.494	0	1
Black	0.145	0.352	0	1
Hispanic	0.087	0.282	0	1
Other Non-White	0.021	0.142	0	1
Health Index	0.767	0.172	0	1
Age	67.190	10.526	50	109
Married	0.635	0.481	0	1
Widowed	0.192	0.394	0	1
Number of Children	3.166	1.977	0	8
Western Region	0.168	0.374	0	1
Midwest Region	0.240	0.427	0	1
Northeast Region	0.159	0.365	0	1
Number of Living Parents	0.272	0.530	0	2
Mothers Age (or Age at Death)	75.263	14.944	16	113
Fathers Age (or Age at Death)	71.390	14.397	12	113
Death	0.061	0.239	0	1
Education/Human Capital				
Highest Grade Completed	12.086	3.349	0	17
High School Graduate	0.693	0.461	0	1
Attended College (1+ years)	0.382	0.486	0	1
College Graduate	0.180	0.385	0	1
Tenure at longest job (years)	20.728	11.842	0	1
Veteran	0.237	0.423	0	1
Strength Required (primary occupation)	0.654	0.925	0	7
Physical Demand (primary occupation)	1.370	8.492	0	6
Exposure Factors (primary occupation)	0.291	0.284	0	3
Financial Information				
Insured	0.872	0.344	0	1
Non-Housing Wealth (100K units)	0.942	2.223	0	15.02
Annual Income (100K units, top coded)	0.491	0.538	0	5
Annual Out of pocket med. exp. (100K units)	0.029	0.102	0	12.06
Calculated Annual Consumption (100K units)	0.661	0.464	0	12.04
Individuals in Data Set	25,872			
T Number of Observations	173,312			

Table III: Health Index Weights

Variable	Weight
Self-Assessed Health	
Excellent	1.241
Very Good	0.802
Good	0.145
Fair	-1.056
Poor	-2.810
Index of Activities of Daily Living	
0	0.392
1	-1.677
2	-2.497
3	-3.00
4	-3.401
5	-3.489
Number of Chronic Health Conditions	
0	1.079
1	0.568
2	-0.047
3	-0.729
4	-1.484
5	-2.418
6	-3.277
7	-4.306
8	-4.317
CESD Mental & Emotional Index	
0	0.807
1	0.180
2	-0.467
3	-0.947
4	-1.227
5	-1.539
6	-1.987
7	-2.501
8	-2.854

Table IV: Cross-Tabs: Mean Change in Log OOP Medical Expenditures by Quintiles of Health and Reference Health

Quintile of Health	Quintile of Ref. Health 1	2	3	4	5
1	0.007	0.015	0.039	0.056	0.081
2	0.002	0.004	0.004	0.013	0.019
3	-0.001	0.000	0.003	0.004	0.008
4	-0.002	0.000	0.002	0.002	0.006
5	–	-0.001	0.001	0.002	0.002

There were only seven observations for which the individual had a reference health in the first quintile and health in the top quintile

Table V: Simple Lagged Dependent Variable Regressions, Log OOP Medical Care Expenditures

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
H_t	-0.073	-0.082	-0.089	-0.095	-0.086	-0.083	-0.091
	(0.003)***	(0.003)***	(0.004)***	(0.004)***	(0.004)***	(0.003)***	(0.004)***
H_{t-1}	0.008	-0.003	-0.003	-0.008	-0.007		
	(0.003)***	(0.003)	(0.004)	(0.005)	(0.004)		
H_{t-2}		0.014	0.014	0.009			
		(0.003)***	(0.004)***	(0.004)**			
H_{t-3}			0.014	0.009			
			(0.004)***	(0.004)*			
H_{t-4}				0.021			
				(0.004)***			
$H_{t'<t}^{max}$					0.037		
					(0.006)***		
$\overline{(H_{t-1}, H_{t-2})}$						0.012	
						(0.004)***	
$\overline{(H_{t-1}, H_{t-2}, H_{t-3})}$							0.025
							(0.005)***
Constant	0.022	0.023	0.015	0.012	0.006	0.023	0.015
	(0.002)***	(0.003)***	(0.003)***	(0.002)***	(0.004)	(0.003)***	(0.003)***
N	139,233	112,610	87,780	66,080	88,730	112,610	87,780

All regressions include controls for age, sex, race, number of children, marital status, income, wealth, lagged medical care, lagged non-medical consumption, and insurance status.
Estimates and Standard Errors for control variables are available upon request.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table VI: Marginal Effects of Reference Health on Logged OOP Medical Expenditures. Lagged Dependent Variable Specification.
Sample split by tercile of lagged medical care expenditure

	Reference Specification		
Variable	$\frac{1}{2}\sum_{j=1}^{2} H_{t-j}$	$\frac{1}{3}\sum_{j=1}^{3} H_{t-j}$	$H_{t'<t}^{max}$
Lower Tercile	0.011	0.025	0.036
	(0.005)**	(0.006)***	(0.006)***
Middle Tercile	0.019	0.032	0.038
	(0.005)***	(0.007)***	(0.007)***
Upper Tercile	0.015	0.024	0.030
	(0.007)**	(0.009)***	(0.010)***

All regressions include same controls as Table V including
lagged medical care, lagged non-medical consumption, and insurance status.
Estimates and Standard Errors for control variables are available upon request.

Table VII: Dynamic Panel Estimation of Health - Evidence on the Effect of Reference Health on the Productivity of Medical Care: Dependent Variable H_{t+1}

	Reference Specification		
Variable	$\frac{1}{2}\sum_{j=1}^{2} H_{t-j}$	$\frac{1}{3}\sum_{j=1}^{3} H_{t-j}$	$H_{t'<t}^{max}$
H_t	0.595	0.585	0.532
	(0.012)***	(0.011)***	(0.011)
R_t	0.193	0.131	0.374
	(0.010)***	(0.011)***	(0.086)***
m_t	-0.061	-0.051	-0.037
	(0.055)	(0.067)	(0.072)
$H_t * m_t$	0.172	0.219	0.172
	(0.086)**	(0.086)**	(0.068)**
$R_t * m_t$	-0.136	-0.129	-0.126
	(0.108)	(0.119)	(0.095)

All regressions include controls for age, sex, race, number of children,
marital status, income, wealth - same arguments as in our full model.
Models are estimated using Arellano-Bover-Blundell-Bond regressions (xtdpdsys)
and robust standard errors
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table VIII: Table of Variables Included in each CDE expression

Variable	Per-period Med. Care	Per-period Cons.	Per-period Health	Initial Med. Care	Initial Cons.	Initial Health	Prob. of Death
Age	X	X	X	X	X	X	X
Age^2	X	X	X	X	X	X	
Black	X	X	X	X	X	X	X
Female	X	X	X	X	X	X	X
Education	X	X	X	X	X	X	
Region Indicators	X	X	X	X	X	X	X
# of Kids	X	X		X	X	X	X
Insured	X	X		X	X	X	
Married	X	X		X	X	X	X
Income	X	X		X	X		
Wealth	X	X		X	X		
H_t	X	X	X				X
H_t^R	X	X	X				X
m_{t-1}	X	X					
z_{t-1}	X	X					
m_t			X				X
z_t			X				X
$H_t * z_t$			X				
$H_t * m_t$			X				
Education*m_t			X				
Veteran Status				X	X	X	X
Living Parents				X	X	X	
Mother's Age				X	X	X	X
Father's Age				X	X	X	X
Physical Work				X	X	X	
Hazard Exposure				X	X	X	
Strength Required				X	X	X	

Figure 1: Distribution of the difference in contemporaneous health and reference health

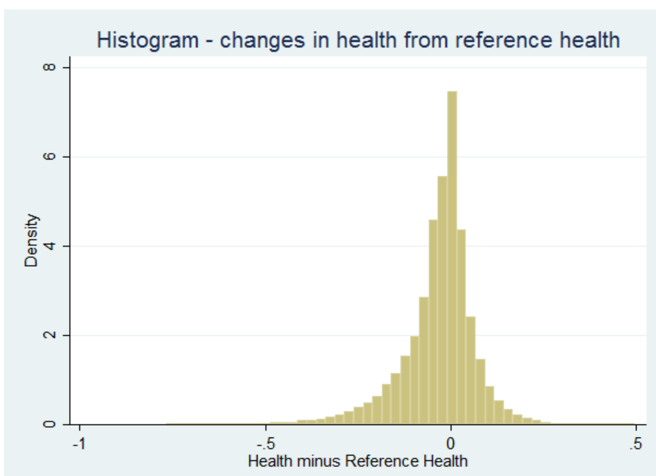

Table IX: Marginal Effects of Key Variables on Medical Care and Consumption

Variables	Medical Care			Consumption		
	Bottom Quartile	Inter Quartile	Top Quartile	Bottom Quartile	Inter Quartile	Top Quartile
Health	-8.8%	-22.3%	-30.3%	5.5%	3.3%	0.3%
Reference Health	10.5%	14.1%	18.5%	3.8%	4.0%	1.4%
Non-Housing Wealth	0.1%	0.2%	0.1%	1.5%	1.1%	0.7%
Income	-0.1%	-0.2%	-0.4%	29.4%	6.5%	2.8%
Insurance	-3.1%	-2.6%	3.5%	18.2%	7.4%	1.4%
Number of Children	-1.4%	-1.8%	-2.4%	10.4%	0.7%	0.05%
Age	4.1%	2.0%	-5.8%	-10.3%	-14.6%	-10.3%
Marital Status	20.0%	4.6%	-0.7%	18.3%	17.1%	3.7%
Widowed	12.7%	15.5%	12.6%	3.0%	1.8%	0.9%
Black	-31.8%	-23.8%	-9.8%	-2.5%	-21.6%	-16.5%
Female	14.2%	-10.1%	-18.0%	1.7%	2.2%	2.7%
Years of Schooling	-1.0%	-6.8%	-11.5%	4.8%	4.2%	1.5%
Lagged Medical Care	1.1%	1.3%	0.5%	0.1%	0.1%	0.2%
Lagged Consumption	-2.5%	-0.8%	1.1%	1.0%	-1.0%	-1.0%

Table X: Marginal Effects of Key Variables on H_{t+1}

Variables	Health		
	Bottom Quartile	Inter Quartile	Top Quartile
Health	4.4%	7.7%	9.1%
Reference Health	4.2%	1.4%	0.8%
Age	-4.1%	-2.3%	-1.4%
Black	-2.4%	-1.8%	-0.7%
Female	12.2%	6.5%	2.2%
m_t	-0.4%	-1.6%	-1.4%
z_t	-0.3%	-0.5%	-0.7%
$H_t * m_t$	1.2%	2.4%	2.5%
Years of Schooling * m_t	0.1%	0.0%	-0.1%
$H_t * z_t$	0.0%	0.0%	0.0%

Figure 2: Kernel Density of Predicted and Observed Health Index, Medical Care, and Consumption Expenditures Distribution

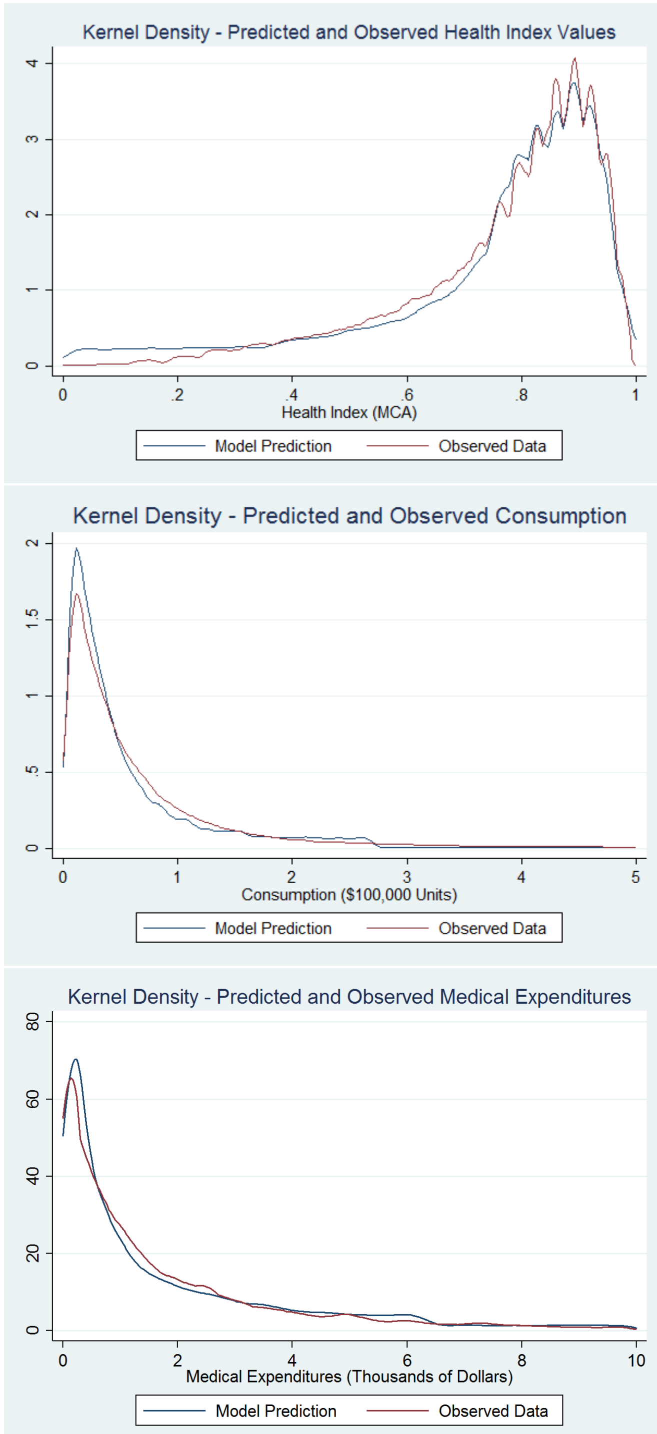

Table XI: Matching Observed Means of the Top 5%

Variable	Observed Data	Preferred Model	Contemp. Variables Only
Age	70.70	**70.77**	73.39
Health	0.636	**0.456**	0.26
Reference Health	0.078	**0.073**	-0.004
Years of Schooling	12.42	**12.13**	11.67
Female	0.635	**0.646**	0.611
Married	0.575	0.487	**0.546**
Income	0.497	0.460	0.432
Lagged Medical Care	0.037	**0.033**	0.055

Numbers in bold indicate they are statistically significantly closer to the values observed in the data by 5%, as determined by two-tailed test for difference in mean.

Table XII: Simulation Results - Average Medical Care Expenditures under Gradual, Moderate, and Sharp Declines in Health in Models with and without reference health

	Period						
	1	2	3	4	5	6	**Total**
			Sharp Decline				
H_t	0.95	0.95	0.95	0.45	0.45	0.45	
m_t with ref. health	0.035	0.036	0.038	0.170	0.141	0.119	0.537
m_t without ref. health	0.035	0.035	0.035	0.154	0.156	0.155	0.570
			Moderate Decline				
H_t	0.95	0.95	0.70	0.70	0.45	0.45	
m_t with ref. health	0.035	0.036	0.077	0.071	0.143	0.131	0.494
m_t without ref. health	0.035	0.035	0.084	0.084	0.156	0.156	0.545
			Gradual Decline				
H_t	0.95	0.85	0.75	0.65	0.55	0.45	
m_t with ref. health	0.035	0.047	0.063	0.081	0.105	0.134	0.466
m_t without ref. health	0.035	0.052	0.073	0.097	0.126	0.156	0.537

A Formation of Occupational Demands

The RAND HRS files contains limited information on respondents' employment history: a categorical response where individuals' self-identify as having primarily been engaged in one of 17 occupations and a variable for how long the individual held that occupation. The coarse occupation categories listed in the HRS correspond to subgroups of occupations from the 1980 Census Occupation Codes.

Data on the indices of occupational demands come from the 1991 Dictionary of Occupational Titles.[31] The DOT is produced by the US Department of Labor and was designed to provide information on the skills/abilities required to perform an occupation. The DOT contains information on 12,686 "occupations" that are better characterized as 'tasks'. A supplement to the 1991 DOT contains information on the demands occupations place on individuals. There is a 5-category rating for required strength: Sedentary, Light, Medium, Heavy, and Very Heavy, corresponding to lifting/manipulating varying weights on the job with varying frequency. We numerically code Sedentary occupations with a strength requirement of 1 and a Very Heavy occupation with a strength requirement of 5.

In addition to strength, the DOT also contains information on the frequency with which occupations require climbing, balancing, stooping, kneeling, crouching, crawling, reaching and handling. For each occupation, each requirement was coded with one of the following values:

$$\begin{aligned}
&0 = \text{Not required}\\
&1 = \text{Required up to } 1/3 \text{ of the time}\\
&2 = \text{Required between } 1/3 \text{ - } 2/3 \text{ of the time}\\
&3 = \text{Required more than } 2/3 \text{ of the time}
\end{aligned} \tag{A.1}$$

To form the numerical measure of physical demands of a given DOT occupation, we take the average over the numerical values assigned to each of the strength values in that occupation.

The DOT also contains information on the frequency with which occupations expose individuals to various adverse environmental conditions or other occupational hazards, including: weather, extreme cold, extreme heat, humidity, noise levels, vibration, poor breathing conditions, proximity to moving mechanical parts, electrical

[31] The 1991 DOT had changed little from the 1977 edition of the DOT. While there are newer, better data sets on occupational requirements (e.g., the O*NET) the 1977-1991 DOT is more relevant to the time where individuals in the HRS were working.

shock, unprotected heights, radiation, explosives, and caustic chemicals. These occupational hazards are coded with the same frequency values as the physical demands. The hazard exposure index for a DOT occupation is calculated by averaging over the reported frequencies for each occupation.

The job requirements supplement to the 1991 DOT also details which 1980 Census Occupation Code encompasses each DOT occupation. Calculating an HRS occupation index value is then a matter of averaging the strength, non-strength physical, and hazard exposure numeric values over all DOT occupations within each COC/HRS occupation.

The HRS also reports the amount of time a respondent worked in their selected occupation. If individuals who worked in physically intensive or hazardous occupations have more worn bodies by the time we observe them in the sample, individuals who worked more years in these arduous occupations should have more proverbial miles on them. We therefore multiply the HRS occupation values for strength required, physical demands, and hazard exposure by the respondent's reported number of years worked in that occupation. It should be noted that these variables, used for exclusion restrictions in initial conditions are intended as ordinal, rather than cardinal variables. The interpretation of having a hazard exposure of "5" is irrelevant. For our purposes, it is sufficient that individuals who have worked more years in more hazardous/arduous occupations experience better/worse health and higher/lower medical expenditures at the time of first observation.

B Parameter Estimates from Conditional Density Estimation

Table XIII: CDE Parameter Estimates for Per-period Medical Care Expenditure

Variables	**X**		**Xγ**		**Xγ^2**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Constant	-4.808	***0.200	1.098	***0.194	0.035	0.058
Health	1.944	***0.134	-0.825	***0.134	-0.543	***0.035
Reference Health	-1.038	***0.146	-1.012	***0.164	-0.330	***0.046
Non-Housing Wealth	0.006	0.006	-0.000	0.006	-0.006	***0.002
Income	0.201	***0.04	0.079	*0.041	-0.026	**0.011
Insurance	-0.302	***0.055	-0.522	***0.056	-0.153	***0.015
Number of Children	0.043	***0.010	0.041	***0.010	0.011	***0.003
Age	1.584	***0.204	1.089	***0.208	-0.119	**0.055
Marital Status	-0.128	**0.058	-0.512	***0.057	-0.211	***0.015
Years of Schooling	0.088	0.062	-0.724	***0.063	-0.356	***0.016
Lagged Medical Care	1.898	***0.193	0.574	**0.263	-1.838	***0.080
Lagged Consumption	0.100	***0.015	0.091	***0.015	0.015	***0.004
Widowed	-0.384	***0.064	-0.469	***0.064	-0.129	***0.017
Northeast Region	-0.155	***0.036	-0.129	***0.014		
Western Region	-0.078	**0.036	-0.096	***0.013		
Midwest Region	-0.079	***0.031	0.020	*0.012		
Black	-0.041	0.040	-0.107	***0.143		
Female	0.447	***0.027	0.249	***0.010		
μ_1^m	0.873	***0.010				
μ_2^m	-0.357	***0.012				
ν_t^m	2.362	***0.038				

Note there are three estimates and standard errors for each variable.
All parameter estimates (and standard errors) capture how that variable affects the hazard probability.
The 2nd and 3rd sets of estimates have the variable interacted with γ and γ^2.
γ is a negative valued term that decreases in magnitude in each successive quantile of the distribution of the dependent variable. The estimates for $X\gamma$ and $X\gamma^2$ reflect the effect of given variable on "survival probabilities" changes over the support of the dependent variable.

Table XIV: CDE Parameter Estimates for Per-period Consumption

Variables	**X**		**Xγ**		**Xγ^2**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Constant	0.213	0.207	-6.903	***0.208	-2.867	***0.053
Health	1.647	***0.140	2.647	***0.140	0.709	***0.036
Reference Health	-0.980	***0.173	-1.745	***0.184	-0.463	***0.050
Non-Housing Wealth	-0.152	***0.003	0.305	***0.004	0.131	***0.001
Income	-0.337	***0.018	1.604	***0.028	0.410	***0.010
Insurance	0.567	***0.056	0.589	***0.055	0.119	***0.014
Number of Children	-0.016	*0.009	-0.080	***0.009	-0.028	***0.003
Age	-2.638	***0.183	1.457	***0.198	1.067	***0.054
Marital Status	0.885	***0.065	1.100	***0.064	0.238	***0.016
Years of Schooling	0.051	***0.061	1.973	***0.065	0.627	***0.017
Lagged Medical Care	0.482	**0.220	2.030	***0.270	0.615	***0.077
Lagged Consumption	0.269	***0.007	0.178	***0.009	0.056	***0.003
Widowed	0.298	***0.077	0.680	***0.075	0.199	***0.019
Northeast Region	-0.092	***0.027	-0.014	0.012		
Western Region	-0.086	***0.025	-0.022	**0.011		
Midwest Region	0.038	0.023	0.041	***0.010		
Black	0.511	***0.036	0.178	***0.016		
Female	-0.173	***0.019	-0.089	***0.008		
μ_1^z	-0.042	***0.014				
μ_2^z	-0.050	***0.010				
ν_t^z	0.051	***0.014				

Note there are three estimates and standard errors for each variable.
All parameter estimates (and standard errors) capture how that variable affects the hazard probability.
The 2nd and 3rd sets of estimates have the variable interacted with γ and γ^2.
γ is a negative valued term that decreases in magnitude in each successive quantile of the distribution of the dependent variable. The estimates for $X\gamma$ and $X\gamma^2$ reflect the effect of given variable on "survival probabilities" changes over the support of the dependent variable.

Table XV: CDE Parameter Estimates for Per-period Health Transition

Variables	**X**		**Xγ**		**Xγ^2**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Constant	7.361	***0.663	1.668	**0.746	1.477	***0.213
Health	-11.445	***0.172	-1.024	***0.146	-0.165	***0.036
Age	6.103	***1.939	2.775	2.183	0.090	0.560
Age Squared	0.055	0.146	-0.058	0.162	0.337	0.443
Reference Health	-4.187	***0.595	-0.042	0.515	-0.120	0.120
m_t	3.281	**0.996	-1.389	5.743	-0.947	1.181
z_t	1.086	0.055	-0.253	0.048	-0.098	0.056
$H_t * m_t$	-5.278	**2.431	3.926	6.779	-1.987	1.393
$H_t * z_t$	-0.288	0.255	0.015	0.232	0.018	0.052
Years of Schooling	0.021	***0.007	0.114	***0.008	0.028	***0.002
Years of School*m_t	0.350	-0.259	-0.021	0.237	-0.021	0.055
Northeast Region	0.078	**0.035	0.059	***0.015		
Western Region	-0.175	***0.034	-0.033	0.015		
Midwest Region	0.184	***0.031	0.118	***0.013		
Black	0.686	***0.072	0.268	***0.075		
Female	0.014	0.031	0.138	***0.048		
μ_1^H	-1.216	***0.011				
μ_2^H	1.096	***0.011				
ν_t^H	-0.014	0.015				

Note there are three estimates and standard errors for each variable.
All parameter estimates (and standard errors) capture how that variable affects the hazard probability.
The 2nd and 3rd sets of estimates have the variable interacted with γ and γ^2.
γ is a negative valued term that decreases in magnitude in each successive quantile of the distribution of the dependent variable. The estimates for $X\gamma$ and $X\gamma^2$ reflect the effect of given variable on "survival probabilities" changes over the support of the dependent variable.

Table XVI: CDE Parameter Estimates for Initial Health Expression

Variables	**X**		**Xγ**		**Xγ^2**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Constant	-2.982	***2.331	2.183	2.489	1.618	**0.642
Age	0.516	0.767	0.096	0.081	0.129	0.201
Age Squared	-0.088	0.615	0.134	0.641	-0.551	1.587
Number of Living Parents	0.134	0.145	0.213	0.154	0.055	0.039
Mothers Age	1.039	*0.542	0.952	*0.511	0.159	0.121
Fathers Age	0.929	**0.475	0.971	**0.454	0.207	*0.109
Years of Schooling	0.095	***0.030	0.143	***0.029	0.021	***0.007
Veteran Status	-0.369	*0.209	-0.369	*0.217	-0.091	*0.054
Insured	-0.818	***0.195	-1.169	***0.202	-0.309	***0.051
Married	0.301	0.233	0.163	0.231	-0.034	0.055
Number of Children	0.069	0.046	0.075	0.047	0.021	*0.011
Northeast Region	0.095	0.093	0.079	**0.039		
Western Region	0.092	0.092	0.093	**0.039		
Midwest Region	0.316	***0.083	0.165	***0.036		
Black	0.091	**0.045	0.053	0.067		
Female	-0.072	**0.031	-0.029	0.031		
Physical Requirements	0.272	***0.075	0.182	***0.034		
Strength Requirements	0.157	0.099	0.017	0.043		
Hazard Exposure	-0.697	*0.409	-0.222	0.180		
$\mu_1^{H_1}$	2.896	***0.029				
$\mu_2^{H_1}$	1.238	***0.022				

Note there are three estimates and standard errors for each variable.
All parameter estimates (and standard errors) capture how that variable affects the hazard probability.
The 2nd and 3rd sets of estimates have the variable interacted with γ and γ^2.
γ is a negative valued term that decreases in magnitude in each successive quantile of the distribution of the dependent variable. The estimates for $X\gamma$ and $X\gamma^2$ reflect the effect of given variable on "survival probabilities" changes over the support of the dependent variable.

Table XVII: CDE Parameter Estimates for Initial Medical Care

Variables	**X**		**Xγ**		**Xγ^2**	
	Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
Constant	0.004	3.172	0.112	3.385	-1.480	*0.874
Age	0.174	1.166	-0.033	1.232	-0.377	0.314
Age Squared	-0.096	1.074	0.099	1.120	0.406	2.846
Number of Living Parents	0.215	*0.109	-0.107	**0.044	0.038	0.030
Mothers Age	0.198	0.520	0.373	0.559	0.108	0.144
Fathers Age	0.121	0.481	0.217	0.521	0.067	0.135
Years of Schooling	-0.098	***0.022	-0.140	***0.024	-0.042	***0.006
Veteran Status	0.299	*0.165	0.352	**0.174	0.111	**0.044
Insured	-0.099	0.167	-0.248	0.177	-0.103	**0.046
Married	-0.194	0.162	-0.520	***0.177	-0.176	***0.046
Number of Children	0.027	0.034	0.053	0.037	0.015	*0.009
Northeast Region	-0.047	0.105	-0.002	0.043		
Western Region	0.144	0.093	0.084	**0.038		
Midwest Region	0.121	0.085	0.066	0.048		
Black	0.218	**0.085	0.042	0.120		
Female	-0.126	***0.034	0.109	**0.050		
Physical Requirements	0.167	**0.085	0.033	0.036		
Strength Requirements	0.088	0.111	0.017	0.046		
Hazard Exposure	-0.436	0.455	-0.115	0.189		
$\mu_1^{m_1}$	-0.362	0.037				
$\mu_2^{m_1}$	-0.076	0.036				

Note there are three estimates and standard errors for each variable.
All parameter estimates (and standard errors) capture how that variable affects the hazard probability.
The 2nd and 3rd sets of estimates have the variable interacted with γ and γ^2.
γ is a negative valued term that decreases in magnitude in each successive quantile of the distribution of the dependent variable. The estimates for $X\gamma$ and $X\gamma^2$ reflect the effect of given variable on "survival probabilities" changes over the support of the dependent variable.