THE ECONOMIC JOURNAL

The Economic Journal, **128** (*November*), 2812–2842. Doi: 10.1111/ecoj.12554 © 2017 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

REFERENCE HEALTH AND THE DEMAND FOR MEDICAL CARE*

Matthew C. Harris and Jennifer L. Kohn

We propose that health in prior periods, termed reference health, is theoretically and empirically relevant to the demand for medical care. To address non-normality in the distributions of medical care spending, consumption, and health, we use a conditional density estimator nested in a finite mixture framework. We find that reference health can help explain the variation in spending among individuals with the same contemporaneous health, particularly in the top tail of the spending distribution. We demonstrate that omitting reference health understates the potential cost-savings of healthy aging initiatives by 50%.

In models of individual behaviour, 'health' is a complicated variable. Economists have treated health as a consumption good, an investment good, an outcome of interest and also as a key explanatory variable. Theoretical and empirical models that include health typically include only contemporaneous health, implicitly assuming contemporaneous health captures all information relevant to economic decisions (Grossman, 1972; Viscusi and Evans, 1990; Gilleskie, 1998; Bound *et al.*, 1999; Hall and Jones, 2007; Edwards, 2008; Yang *et al.*, 2009; DiNardi *et al.*, 2010; Khwaja, 2010; Hugonnier *et al.*, 2013; Peijnenberg *et al.*, 2015; Yogo, 2016). However, diverse literatures have successfully incorporated past realisations of key variables to explain economic puzzles ranging from the equity premium puzzle (Constantinides, 1990; Dai and Grischenko, 2014), labour supply (Koszegi and Rabin, 2006, 2009; Crawford and Meng, 2011) and addiction (Becker and Murphy, 1988; Darden, 2017). 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 depending on their health in the past.

In addition to theoretical precedent, certain stylised empirical facts suggest that contemporaneous health alone does not fully capture the impact of health on decision making. For example, when considering the distribution of medical care spending, consumers in the upper tail are of particular interest. Medical care spending accounts for 18% of US GDP, half of which (9% of GDP) is consumed by 5% of the population. Contemporaneous health alone has surprisingly little explanatory power in predicting who will be in the top 5% of medical care consumers. Fewer than 20% of high spenders

* Corresponding author: Matthew C. Harris, Department of Economics and Boyd Center for Business and Economic Research, Haslam College of Business, University of Tennessee, 722A Stokely Management Center, 916 Volunteer Boulevard, Knoxville, TN 37996-0570, USA. Email: mharris@utk.edu.

We are grateful to Robert Patrick for his collaboration on the original theory. We thank our editor, Frederic Vermeulen and two anonymous referees for their helpful comments and guidance. 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.

report poor health, while 7.5% report excellent health (Schoenman, 2012). Moreover, not all those in poor health consume large amounts of medical care. Nearly 40% of medical care spending by individuals in poor health is incurred by the top 5% of this group (Claxton *et al.*, 2014).

We apply elements from the literature on reference dependence to the literature on dynamic models of investment in health to address the economic puzzle of why some individuals, particularly relatively healthy ones, demand high amounts of medical care. We incorporate past realisations of health (which we term reference health) into the utility function in a dynamic model of health, consumption, and the demand for medical care.[1] Consider two individuals with the same contemporaneous level of health. Consistent with the literature on reference dependence, we hypothesise 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 is one mechanism that would justify higher medical care spending.[2] 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 a factor (e.g. investment decisions, labour force participation, smoking, exercise, etc.).

To provide economic intuition and motivate our empirical analysis, we present a theoretical model in the Grossman (1972) tradition and derive an effect of reference health on the demand for medical care. This derivation illustrates the need to model medical care and other consumption jointly to explore empirically the hypothesis that higher reference health increases the demand for medical care. Relevant to our empirical analysis, our theoretical exposition also implies that the effect of reference health on medical care spending is likely to vary over the support of the distribution of medical care spending.

Our econometric approach is motivated by theory and by the aforementioned stylised facts about the distribution of medical care spending. Empirically evaluating the effect of reference health requires accommodating several key features of the model and the data. Because of the non-normality of the distribution of medical care spending and the importance of modelling participation in the top tail of the distribution, it is important to capture not just the conditional mean, but also the conditional probability that an individual is observed in the top quartile, decile or top 5%. Additionally, our theoretical model implies the need to jointly estimate the demands for medical care and consumption and to capture the dynamic effects of these decisions on health in the ensuing period. Finally, our joint estimation procedure must accommodate permanent and time-varying unobservable heterogeneity between the two demand expressions and the evolution of health. We estimate the joint demands for medical care and consumption, using a conditional density estimator (CDE) nested in a finite mixture

[1] We operationalise reference health as the average of the individual's past two realised health states. We show in online Appendix E, using reduced form specifications, that the qualitative implication of including reference health is not sensitive to the functional form of reference health.

[2] While our theoretical model focuses on reference health operating through the utility function, there are several alternative mechanisms by which health in previous periods can affect contemporaneous demand for medical care. We discuss and provide some evidence on these alternative hypotheses in single-equation models in online Appendix E.

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). CDE allows explanatory variables to have different impacts over the distribution of the dependent variable. CDE also enables us to model the probability that individuals are observed in a particular part of the distribution and to easily accommodate the joint estimation of medical care and consumption with requisite concerns for unobserved heterogeneity. We analyse the effect of reference health on the demand for medical care with data from the US Health and Retirement Survey (HRS).

We find strong empirical evidence that reference health directly affects the demand for medical care across the whole distribution of medical care spending. Confirming the value of CDE, we find that the marginal effect of reference health, conditional on contemporaneous health, varies over the support of the distribution. A 10 percentage point increase in reference health increases expected medical care spending by 5.6% in the bottom quartile, but 13% in the top quartile of the distribution. CDE also enables us to simulate the effects of health and reference health on the full distribution of medical care spending. We find a 20 percentage point increase in reference health holding contemporaneous health constant has a similar effect on the distribution as a 10% decline in health and reference health (implying a flat health trajectory). Moreover, a 10 percentage point increase in reference health is associated with a 19.3% increase in the probability that the individual is observed in the top 5% of the medical care spending distribution. Finally, we use our estimates to simulate different paths of medical care spending associated with different health trajectories over a 12 year span. Comparing simulations with and without reference health demonstrates that omitting reference health underestimates by half the potential cost savings from policies such as healthy aging initiatives that would prevent more volatile health trajectories.

Our primary contribution is to demonstrate that reference health should be included in economic models involving health. We are the first, to our knowledge, to incorporate past realisations of health to improve models of medical care spending. We contribute to the literature on the evolution of health by relaxing the Markov assumption that only the most recent observation of health affects transition probabilities (Wouterse *et al.*, 2013; Peijnenberg *et al.*, 2015). We also contribute to the literature on modelling the distribution of medical care spending, confirming the importance of going beyond the conditional mean (French and Jones, 2004; LeCook and Manning, 2009; de Meijer *et al.*, 2013; Jones *et al.*, 2015). Finally, there is a growing finance literature that incorporates medical care spending into dynamic lifecycle models to explain the equity premium and annuity puzzles. Most recently, Peijnenberg *et al.* (2015, p. 1623) showed 'the sensitivity of optimal annuitisation decisions to the exact specification of the medical expense process'. While Peijnenberg *et al.* (2015) call for better data to estimate medical expense risk, we suggest that better models that include reference health can also aid our understanding.

The rest of the article proceeds as follows: Section 1 motivates our empirical specification with a theoretical model. Section 2 describes the data used in estimation including our measures of health, medical care and consumption spending. Section 3 details our estimation strategy, econometric identification, and post-estimation simulation methods. Section 4 describes our results including marginal effects and

simulations on the full distribution of medical care spending, the top 5% of spenders, and spending for different health trajectories. Section 5 concludes. Supplementary information is contained in an online Appendix.

1. Theoretical Motivation

To illustrate one mechanism whereby reference health may affect individual decision making and to motivate our empirical specification we extend the Grossman (1972) model by including reference health as an element of an individual's utility. The model developed here is purely for expository purposes and is not intended to stand on its own.[3] 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 illustrate the effect of reference health on the demand for medical care.[4]

The objective of the individual is to maximise her expected present discounted utility from non-medical consumption, z_t, and health, H_t, conditional on her reference health, R_t and a vector of other all other variables comprising the individual's information set, S_t, including past values of consumption and medical care.[5] To maintain our focus on reference health we treat all variables in S_t as pre-determined in what follows. Empirically, we treat reference health as the average of the past two values of lagged health: $R_t = (H_{t-1} + H_{t-2})/2$.[6] However, to ease our theoretical exposition we maintain a general recursive functional form for reference health as denoted in (1):

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

Higher values of z_t, H_t and R_t reflect higher values of consumption, contemporaneous and reference health respectively. We assume that utility is concave in health and consumption. Consistent with the literatures on habit and reference dependence,

[3] Please see Kohn and Patrick (2008) for a complete exposition of the model in a continuous time optimal control framework.

[4] We make several modelling trade-offs consistent with our focus on reference health and our data. See Hugonnier *et al.* (2013, table 3) for an excellent summary of the major modelling choices in the literature. First, we do not model the choice of insurance, which is clearly endogenous to health status and the demand for medical care, but constrained by employment, income (Medicaid) and in the United States, 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 labour decisions, particularly retirement, but we do include control variables for 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 3 details our econometric strategy.

[5] There are three primary strands of the theoretical literature that allow past values of a variable to affect contemporaneous utility: rational addiction (Becker and Murphy, 1988), habit persistence (Ryder and Heal, 1973; Constantinides, 1990) and reference dependent utility (Kahneman and Tversky, 1979; Koszegi and Rabin, 2006, 2009; Baucells *et al.*, 2011). Under a reference dependence interpretation, if the individual's reference point is defined by her own values in previous periods and relative utility is defined as the difference between contemporaneous and reference levels, then the implications of reference dependence and habit persistence are similar.

[6] We show in online Appendix E, using reduced form specifications, that the qualitative implication of including reference health is not sensitive to the functional form of R_t.

individuals get utility not only from contemporaneous health, H_t, but also from their level of health relative to reference health ($H_t - R_t$). This specification implies that R_t, by itself, has a negative effect on utility. The higher the level of reference health, the less total utility the individual receives from a given level of contemporaneous health.

To complete the model, we specify wealth and health transition equations and endpoint conditions for the health and wealth states:

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

Health investment is a function of medical care, m_t, with diminishing marginal returns, the state of health (reflecting co-morbidities) and the individual's information set, including age, gender, and other control variables, specified as $I(m_t, H_t; S_t)$.[7] Next-period health is also subject to a stochastic shock, δ_t, about which we make no distributional assumptions. The health shock is likely to have two properties that inform our empirical specification. First, the shock 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.[8]

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,[9] minus expenditures on consumption and medical care, $-z_t - p_{mm}t$, with the price of non-medical consumption normalised to 1. Wealth, W_t, must remain above some minimum level that does not preclude debt. Health is bounded from above by a maximum biological level, and if H_t falls below some minimum value of health the individual dies.

We express the constrained optimisation problem recursively:

$$\begin{aligned} V_t(W_t, H_t; R_t, S_t) &= \max_{z_t, m_t} U(z_t, H_t; R_t, S_t) + \beta E V_{t+1}(W_{t+1}, H_{t+1}; R_{t+1}, S_{t+1}) \\ &= \max_{z_t, m_t} U(z_t, H_t; R_t, S_t) + \beta E V_{t+1}\big[(1+r)W_t + Y_t - z_t - p_m m_t, H_t \\ &\quad + I(m_t, H_t; S_t) + \delta_t; f(R_t, H_t), S_{t+1}\big], \end{aligned} \tag{3}$$

where the dynamic constraints are substituted for the state variables in the third line of (3). The first-order conditions for the choice variables, z_t and m_t, are:

[7] The assumption of a health production function with diminishing returns to medical care is consistent with Ehrlich and Chuma (1990) and Galama (2015).

[8] There is a strand of literature that examines uncertainty in the productivity of medical care (Dardanoni and Wagstaff, 1990). Empirically, we cannot separately identify ineffective medical care from a health shock. For the purpose of theoretical exposition, we simplify the stochastic element in health production to one additive shock. In our empirical specification (see subsection 3.4) we address the likely correlation and persistence of the health shock by using discrete factor random effects.

[9] To maintain our focus on reference health, we make income independent of health. Making income a function of health would merely add an additional term to the envelope condition for health and not otherwise change any of the model's implications on the effect of reference health on the demand for medical care.

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

The envelope conditions on the relevant state variables are as follows:

$$\begin{aligned} W_t &: V_t^W = (1+r)\beta E_t V_{t+1}^W, \\ H_t &: V_t^H = U_t^H + \beta E_t\left(V_{t+1}^H I_t^H + V_{t+1}^R f_t^H\right), \\ R_t &: V_t^R = U_t^R + \beta E_t V_{t+1}^R f_t^R. \end{aligned} \tag{5}$$

These first-order and envelope conditions yield a critical insight that motivates our empirical approach to estimate the demands for consumption and medical care jointly. Intuitively, the first-order conditions from (4) show that individuals choose optimal levels of consumption and medical care when the marginal utilities of each equal the discounted marginal value of wealth in the next period. The envelope condition for wealth shows that the discounted marginal utility of wealth in the next period is itself a function of both health and reference health since both are elements of the value function. Thus, any argument that affects the demand for consumption must be included in the demand for medical care and *vice versa*, including reference health. Including reference health and S_t, the individual's information set, we can express optimal z_t and m_t:

$$\begin{aligned} z_t^* &= z(H_t, R_t, W_t, Y_t, S_t), \\ m_t^* &= m(H_t, R_t, W_t, Y_t, S_t). \end{aligned} \tag{6}$$

Finally, we can derive the effect of reference health on the demand for medical care. A detailed derivation is in online Appendix A. Briefly, starting with the first-order condition for medical care from (4), we substitute for V_{t+1}^H by rolling forward the envelope condition for health from (5) by one period. Then we substitute for $\beta E_t V_{t+1}^W$ from the first-order condition for consumption. The resulting equilibrium demand for medical care in (7) shows the marginal benefits of medical care on the left-hand side, including the additional term associated with reference health in the value function, equaling the marginal cost of foregone contemporaneous consumption:[10]

$$E_t\left[U_{t+1}^H + \beta E_t\left(V_{t+2}^H I_{t+1}^H + V_{t+2}^R f_{t+1}^{H_{t+1}}\right)\right] I_t^m = U_t^z p_m. \tag{7}$$

While full comparative dynamics in a three-state optimal control model are not easily computed, we can still obtain accessible insights about the effect of reference health on

[10] Our equilibrium condition exhibits three differences from Grossman's equilibrium condition (1972, p. 229). First, our equilibrium condition includes the effects of health investment on future health production and future reference health. Second, we model the marginal utility from health more generally as U^Hrather than Grossman's more specific health days, G. Third, we do not add back the rate of health depreciation because health changes through additive stochastic shocks rather than a multiplicative rate of depreciation. The effect of these changes is that in equilibrium there are more benefits to investing in health, and the cost does not automatically increase because depreciation is not added back. In the Grossman equilibrium condition, the marginal benefits and marginal costs of health investment move in tandem. In our specification, declines in health create disequilibrium, leading to increased demand for medical care. Our equilibrium condition therefore yields additional insight as to why individuals would spend more on medical care when health declines, and why some in good health would still invest in medical care.

medical care demand that inform our empirical exercise. Since our purpose here is purely expository, we simplify the dynamic model to two periods. In doing so, we appeal to the Principle of Optimality (Caputo, 2005) which proves 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, 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 (7) to only two periods, t and $t+1$, leaves a modified equilibrium condition for medical care demand:

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

We implicitly differentiate (8) with respect to m_t and R_t and rearrange to derive the change in medical care demand resulting from a change in reference health:[11]

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

Equation (9) illustrates how reference health affects the demand for medical care and informs our empirical hypotheses. Reference health can affect the demand for medical care through cross partial effects on both health and consumption, requiring our joint estimation of consumption and medical care. Our hypothesis is that an increase in reference health increases the demand for medical care. A positive effect would be unambiguous if $U^{HR} > 0$ and $U^{Hz} < 0$. The effect of reference health on the demand for medical care could also be positive if the cross-partials have opposite signs or are both positive depending on their relative magnitudes. We make no assumptions about these signs of these cross partials and treat both their signs and magnitudes as empirical questions.[12] Furthermore, the denominator of (9) shows how reference health can explain differences in the distribution of medical care demand among individuals with the same contemporaneous health. Because I_t^{mm} decreases in value as medical care use increases, the marginal effect of reference health is greater at higher levels of medical care spending. If true, reference health can help to explain why some in the top 5% of medical care spending may have relatively good health, and why not all those in poor health are high medical care spenders.

Empirically we evaluate $\partial m^*/\partial R_t$, which we treat as a referendum on our hypothesis that $U^{HR} > 0$. Our estimation method allows us to test for differences in sign, magnitude and significance at different points of the distribution of medical care demand. We make no *a priori* hypothesis on the sign or significance of U^{zR}. There is conflicting empirical evidence in the literature regarding the effect of health on the

[11] See online Appendix A for a step-by-step derivation of (9).

[12] The sign of (9) takes the sign of the numerator because of the preceding negative sign and the negative denominator due to the concavity of the utility and investment functions.

marginal utility of consumption (Viscusi and Evans, 1990; Edwards, 2008; DiNardi *et al.*, 2010; Acemoglu *et al.*, 2013; Finkelstein *et al.*, 2013), and since we are the first to include reference health, we will offer the first empirical insights on the effect of reference health on the marginal utility from consumption. If $U^{zR} \geq 0$ and $\partial m^*/\partial R_t > 0$ then U^{HR} must be positive, consistent with our hypothesis. However, if $U^{zR} < 0$ and/or $\partial m^*/\partial R_t \leq 0$ then reference health will still be relevant to demand, but our empirical test of U^{HR} will be inconclusive depending on relative magnitudes of the two cross-partial effects of reference health.

The purpose of our theoretical motivation is to provide some intuition as to why reference health affects the individual's optimisation problem. Here, we focus on the effect of reference health operating through the utility function. In what follows, we present an econometric strategy and empirical evidence consistent with the utility mechanism of the effect of reference health on the demand for medical care. However, there are several plausible alternative mechanisms. For example, reference health may operate through the health production function by providing additional information about health and/or the trajectory of health decline. Empirically, reference health may capture the effect of the unobservable health shock and/or onset of a new medical condition. In online Appendix E we present empirical evidence using single-equation specifications on the relative importance of these alternatives. While we cannot definitively eliminate these alternative mechanisms, we show that they are unlikely to dominate our proposed mechanism of reference health operating through the utility function. Our theoretical and empirical argument is that reference health can help to explain individual decision making better than models with contemporaneous health alone.

2. Data

We use data 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 to 2010. Several features of the RAND HRS data 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 high-frequency data, e.g. Medical Expenditure Panel Survey or Survey of Income and Program Participation.

Finally, that the HRS is comprised of older individuals yields three additional advantages for our empirical work. First, older individuals are more likely to exhibit differences in their overall health trajectory than younger individuals whose health is more likely to be stable for extended periods. Second, individuals in the HRS are mostly over 50 and have likely completed their education and achieved their peak earnings, thereby mitigating much of the wage production benefits of health. However, health will affect individuals' retirement decisions and thereby earnings from labour. Third, concerns about insurance choices are at least partially mitigated by Medicare.

Our empirical model requires two periods of initial conditions. We therefore restrict the sample to individuals who are observed for at least three periods, yielding a sample of 173,378 observations of 25,827 individuals. We observe individuals for an average of seven waves (14 years). Tables 1 and 2 display summary statistics.

The RAND HRS includes two measures of medical care expenditures: out of pocket (OOP) and total medical care expenditures. We use OOP expenditures for three reasons. First, total expenditures were reported only for the first six waves, cutting our effective sample by more than half. Second, the survey question regarding total expenditure asks respondents to recall the total amount incurred without external validation. As insured individuals are notoriously insulated from their true total costs, these responses were unreliable.[13] Third, while total expenditures are policy relevant, the OOP expense is more relevant to the individual. Even among insured individuals, those over (under) 65 still spend approximately 16% (10%) of disposable income on medical care (Desmond *et al.*, 2007; Banthin and Bernard, 2010). Finally, OOP expenditures reflect our theory in which individuals respond to the price of medical care that they pay.

The RAND HRS includes many discrete categorical variables for level of difficulties with activities of daily living (ADLs), chronic conditions, self-assessed health (SAH), 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. Because 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. Since our focus is on the effect of reference health, we need a continuous, cardinal measure of health to keep the state space manageable.[14] We convert these discrete categorical variables into a single continuous measure of health, using multiple correspondence analysis (Greenacre and Blasius, 2006; Kohn,

Table 1

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

Note. The higher numbers observed for 4 and 7 are due to HRS adding respondents at these waves.

[13] Total expenditures has 16 times the variance of OOP expenditures. The two variables have a correlation coefficient of 0.14.

[14] See online Appendix B for a complete discussion of why and how we create a continuous, cardinal health measure.

Table 2

Summary Statistics

Variable	Mean	SD	Min	Max
Demographic variables				
Female	0.575	0.494	0	1
Black	0.146	0.353	0	1
Hispanic	0.089	0.285	0	1
Other non-white	0.023	0.149	0	1
Health index	0.767	0.172	0	1
Age	67.190	10.526	22	109
Married	0.668	0.470	0	1
Widowed	0.193	0.378	0	1
Number of children	3.175	1.977	0	8
Western region	0.164	0.370	0	1
Midwest region	0.249	0.432	0	1
Northeast region	0.164	0.370	0	1
Number of living parents	0.498	0.680	0	2
Mothers age (or age at death)	75.006	14.864	16	113
Fathers age (or age at death)	71.262	14.327	12	113
Death	0.024	0.155	0	1
Education/human capital				
Highest grade completed	12.0068	3.385	0	17
High school graduate	0.689	0.462	0	1
Attended college (1+ years)	0.386	0.487	0	1
College graduate	0.181	0.385	0	1
Tenure at longest job (years)	20.511	11.826	0	1
Veteran	0.230	0.421	0	1
Strength required (primary occupation)	0.654	0.924	0	7
Physical demand (primary occupation)	1.358	8.465	0	6
Exposure factors (primary occupation)	0.289	0.284	0	3
Financial information				
Insured	0.903	0.295	0	1
Non-housing wealth (100K units)	0.926	2.223	−0.500	15.02
Annual income (100K units, top coded)	0.498	0.542	0	5
Annual out of pocket medical expenses (100K units)	0.029	0.102	0	12.06
Calculated log annual consumption (100K units)	0.792	0.498	0	6.588
Individuals in data set	25,827			
Number of observations	173,378			

2012). Table 3 contains the variables and weights used to construct the health index. These weights are very intuitive: better health whether indicated by SAH, ADLs, chronic conditions or mental and emotional health has a higher weight. Moreover, the weights show economically reasonable diminishing returns to good health for the SAH weights (the difference between excellent and very good is smaller than the difference between poor and fair) and diminishing negative impacts to accumulating impairments.

Figure 1 shows the distribution of the difference between contemporaneous and reference health in our sample. As our sample is comprised of older individuals, it is not surprising 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. However, there are also plenty of instances where individuals in the sample experience improvements in their health relative to recent health history.

Table 3

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

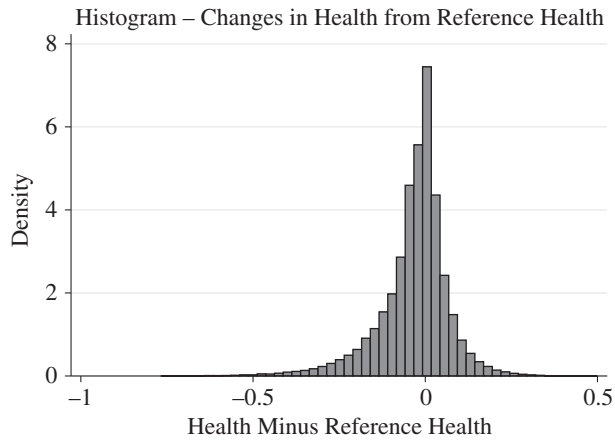

Fig. 1. *Distribution of* $(H_t - R_t)$

Note. Colour figure can be viewed at wileyonlinelibrary.com

Table 4

Mean Change in Log OOP Expenditures by Quintiles of H_t and R_t

	Quintile of reference health				
Quintile of health	1	2	3	4	5
1	0.008	0.013	0.032	0.042	0.061
2	0.002	0.003	0.007	0.009	0.017
3	−0.001	0.001	0.002	0.004	0.008
4	−0.003	0.001	0.002	0.002	0.004
5	–	0.001	0.001	0.000	0.002

Notes. There were only seven observations with a reference health in the first quintile and health in the top quintile. Values are in units of $100,000.

Table 4 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 OOP spending for each quintile combination. For each quintile of health, the mean change in log OOP medical care expenditures is monotonically increasing in reference health. In every row of Table 4, the fifth column (containing individuals in the top quintile of reference health) has the greatest change in OOP medical expenditures, followed by the fourth column, etc. Note that for each quintile of reference health, the mean change in log OOP 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 OOP 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.[15]

The summary statistics reflect commonly known stylised facts about the distributions of health, medical care, and consumption. Health is highly skewed left, while the distributions of medical care expenditures and non-medical consumption are skewed right (see Figure 2). In our sample, the top 5% of medical care consumers account for 46% of all medical care spending, consistent with the stylised fact that the top 5% account for nearly half of medical care spending (Claxton *et al.*, 2014). The non-normality of these distributions underscores the importance of modelling the full distributions, rather than just estimating the conditional mean.

[15] The median person in our sample has non-housing financial wealth of $10,500 and a person in the 90th 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.

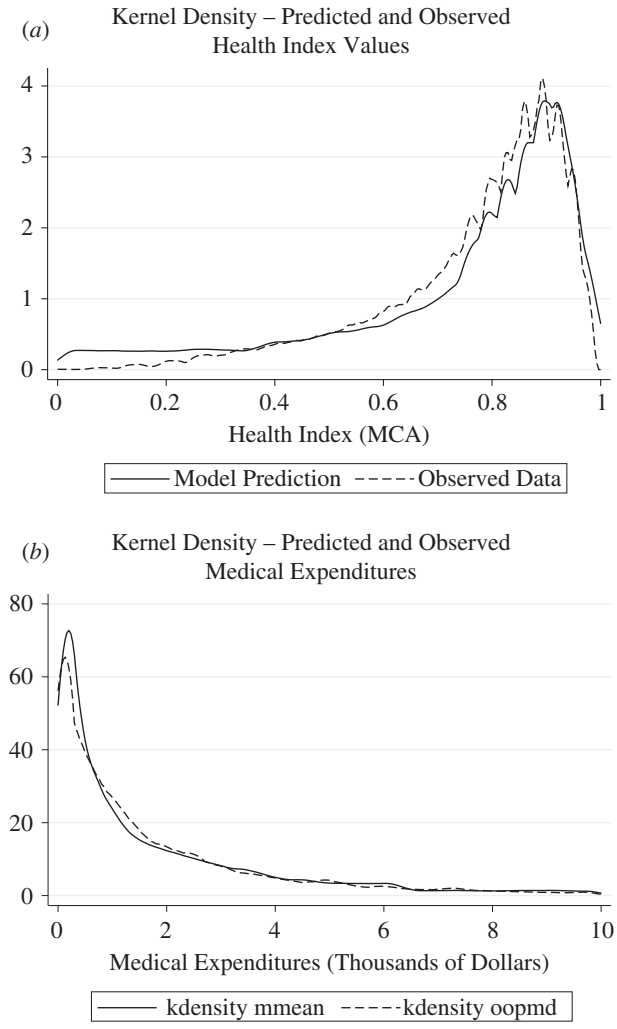

Fig. 2. *Distributions of Predicted and Observed Health and Medical Care*

3. Econometric Methodology

Our econometric approach is motivated by both the theoretical model and by the above stylised facts about the distribution of medical care spending. Empirically evaluating the effect of reference health requires us to accommodate three key features of the model and the data. First, because of the policy relevance of the top tail of the distribution of medical care spending, it is important to capture not just the conditional mean, but also the conditional probability that an individual is observed in the top quartile, decile or top 5% of the distribution. To accomplish this, we require an empirical method which will allow the marginal effects of key variables to vary over the support of the dependent variable. Second, in keeping with the theoretical model we need to jointly estimate the demands for medical care and consumption as well as the dynamic effects of these decisions on health in the ensuing period. Because the

distributions of medical care and consumption are heavily right skewed while the distribution of health is left skewed, we need a method without parametric assumptions on the error terms. Finally, our joint estimation procedure must work in a dynamic model which permits correlated permanent and time-varying unobservable heterogeneity between the two demand expressions and health.

A key paper by Jones *et al.* (2015) presented Monte Carlo evidence on the relative merits of several distributional estimators for modelling medical care expenditures and found that no particular method dominated. As estimating tail probabilities or undertaking joint estimation with unobserved heterogeneity is prohibitive with a quantile estimator, we employ a conditional density estimator (CDE). We selected the method from Gilleskie and Mroz (2004), in which the CDE is estimated using a sequence of logit hazard probabilities. This estimator can accommodate the rich stylised facts of the distributions and also easily accommodate both permanent and time-varying heterogeneity (Heckman and Singer, 1984; Mroz, 1999).

Our discussion proceeds in six subsections. First, we present our estimating equations and link them to our theory. Second, we discuss econometric identification. Third, we provide a brief explanation of the mechanics for our particular choice of CDE. Fourth, we incorporate discrete factor random effects to address the non-independence in the econometric errors. Fifth, we specify the likelihood equation that puts all of these components together, and finally we discuss the specific estimation and simulation methods that produce our results.

3.1. *Estimating Equations*

The empirical joint demand for consumption and medical care is the same as (6) with the addition of the econometric errors, denoted ϵ_t^z and ϵ_t^m:

$$\begin{aligned} z_t^* &= z(H_t, R_t, W_t, Y_t, S_t, \epsilon_t^z), \\ m_t^* &= m(H_t, R_t, W_t, Y_t, S_t, \epsilon_t^m), \end{aligned} \tag{10}$$

where $R_t = (H_{t-1} + H_{t-2})/2$. Recall that S_t is a vector of variables in the individual's information set including m_{t-1}, z_{t-1}, and demographic controls variables including age, gender, race, education, etc. See Table 5 for a list of arguments in each expression.

Our empirical expression for the individual's health transition follows from the theory with two key additions. First, the individual's choices for z_t^* and m_t^* are allowed to affect health in the next period, reflecting our dynamic optimal control model where choices affect the future states. Second, we include reference health to empirically evaluate the extent to which the effect of reference health operates through the production function:

$$H_{t+1} = \alpha(H_t, R_t, m_t, z_t, S_t^H, \epsilon_t^H). \tag{11}$$

3.2. *Econometric Identification*

The econometric identification of this model comes through three sources: exclusion restrictions, timing, and non-linearity in the hazard functions. Table 5 contains the full list of variables included in each expression. We allow m_t and z_t to affect H_{t+1}. We exclude m_{t-1} and z_{t-1} from the health transition equation, assuming that the information from

Table 5

Table of Variables Included in Each CDE Expression

Variable	Per-period medical care	Per-period consumption	Per-period health	Initial medical care	Initial consumption	Initial health	Probability 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
Years of schooling	X	X	X	X	X	X	
Region indicators	X	X	X	X	X	X	X
Number of children	X	X		X	X	X	X
Insured	X	X		X	X	X	
Married	X	X		X	X	X	X
Widowed	X	X		X	X	X	X
Income	X	X		X	X		
Non-housing wealth	X	X		X	X		
Health	X	X	X				X
Reference health	X	X	X				X
Lagged medical care	X	X					
Lagged consumption	X	X					
Medical care			X				X
Consumption			X				X
Health × Consumption			X				
Health × Medical care			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	

the individual's lagged consumption of medical and non-medical goods is captured by H_t, m_t, and z_t. Similarly, we exclude the individual's income, wealth, insurance status, number of kids, and marital state from the health transition expression, assuming these variables affect her demand for medical and non-medical goods, but not her health in the next period. They are therefore excluded from S_t^H but included in S_t.

To appeal to the timing assumption, we must specify endogenous initial conditions for the first two health states and initial demands for medical and non-medical goods:

$$\begin{aligned} H_1 &= H_1^i(S_1^H, \mathbf{X^h}), \\ H_2 &= H_2^i(S_2^H, \mathbf{X^h}, H_1), \\ m_2 &= m_2^i(S_2, \mathbf{X^i}, H_1), \\ z_2 &= z_2^i(S_2, \mathbf{X^i}, H_1), \end{aligned} \tag{12}$$

$\mathbf{X^i}$ and $\mathbf{X^h}$ are sets 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.[16] *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 detailed in online Appendix C.

3.3. *Conditional Density Estimation*

CDE utilises a sequence of conditional logit probabilities 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.[17] For each interval, the kth 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 kth interval, given that it is not observed in intervals 1 through $k-1$, can be expressed:

$$\lambda(k, x) = p[y_{k-1} \leq Y < y_k | x, Y \geq y_{k-1}] = \frac{\int_{y_{k-1}}^{y_k} f(y|x)\mathrm{d}y}{1 - \int_{y_0}^{y_{k-1}} f(y|x)\mathrm{d}y}. \tag{13}$$

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 kth interval is given by:

$$p[y_{k-1} \leq Y < y_k | x] = \lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)]. \tag{14}$$

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 (10)–(11) and for each cell $k \in \{1, \ldots, K\}$, we can form a linear function of our covariates and γ terms:

$$g^j(k, x) = \mathbf{X^j}\beta_1^j + \mathbf{X^j}\gamma_k\beta_2^j + \mathbf{X^j}\gamma_k^2\beta_3^j + \epsilon_t^j \quad \forall j \in \{z, m, H, H_1, H_2, z_2, m_2\}. \tag{15}$$

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

[17] 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 of these distributions.

With some abuse of notation, $\mathbf{X^j}$ includes 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.[18] From (15), we can therefore form the logit probabilities used to form the hazard function:

$$\lambda^j(k, x) = \frac{e^{g^j(k,x)}}{1 + e^{g^j(k,x)}}. \tag{16}$$

and these terms are subsequently combined to form the probabilities in (14).

3.4. *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 utilise 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 consistently invest in more medical care, engage in lower consumption, and enjoy persistently smaller 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

[18] Including higher orders yields greater flexibility in estimating the marginal effects over the distribution at a cost of increasing the number of coefficients to be estimated. As expanding to a cubic polynomial did not improve performance, we have reported the results from the quadratic specification.

permanent components and are designed to capture correlation in the idiosyncratic variation around the overall trends. We therefore decompose the errors in each expression into three components:

$$\epsilon_t^j = \mu^j + v_t^j + e_t^j \quad \forall j \in z, m, H, \tag{17}$$

where μ^j captures the permanent heterogeneity for each expression, v_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.

3.5. *Likelihood Function*

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 as follows:

$$\begin{aligned} L_i(\boldsymbol{\Theta}, \Psi, \Pi) = \sum_{k=1}^{K} \pi_k \Bigg\{ & \prod_{j_{h1}=1}^{J_{h1}} P(H_1 = j_{h1} | \mu_k^{H_1})^{1(H_1 = j_{h1})} \prod_{j_{h2}=1}^{J_{h2}} P(H_2 = j_{h2} | \mu_k^{H_2})^{1(H_2 = j_{h2})} \\ & \times \prod_{j_{m2}=1}^{J_{m2}} P(m_2 = j_{m2} | \mu_k^{m_2})^{1(m_2 = j_{m2})} \prod_{j_{z2}=1}^{J_{z2}} P(z_2 = j_{z2} | \mu_k^{z_2})^{1(z_2 = j_{z2})} \\ & \times \prod_{t=3}^{T_i} \sum_{l=1}^{L} \psi_l \Bigg[\prod_{j_m=1}^{J_m} P(m_t = j_m | \mu_k^m, v_{lt}^m)^{1(m_t = j_m)} \prod_{j_z=1}^{J_z} P(z_t = j_z | \mu_k^z, v_{lt}^z)^{1(z_t = j_z)} \\ & \times \prod_{j_H=1}^{J_H} P(H_t = j_m | \mu_k^H, v_{lt}^H)^{1(H_t = j_H)} \prod_{D=0}^{1} p(death = D | \mu_k^D, v_{lt}^D)^{1(death = D)} \Bigg] \Bigg\}, \end{aligned} \tag{18}$$

where $\boldsymbol{\Theta}$ is the vector of coefficients to be estimated from (10), (11) and (12); $\psi_l \in \Psi$ are the mixing variables for the time-varying heterogeneity, and $\pi_k \in \Pi$ are the mixing variables for the permanent heterogeneity. K and L represent the number of mass points for the distributions of permanent and time-varying heterogeneity, respectively, and t indexes time.[19] 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_{z2}, 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.

[19] We have estimated the model with three mass points in the support of permanent heterogeneity and two points of time-varying heterogeneity. Adding additional mass points does not significantly increase the likelihood function, as vetted with an LR test (p = 0.21).

3.6. *Marginal Effects and Simulation*

Because the estimated coefficients are included as arguments in non-linear hazard functions, they are not directly interpretable without additional simulation. To calculate the marginal effects presented in subsection 4.1, expectations are formed, using the probabilities defined in (14) and the mid-point of each cell as the within-cell expected value. Without loss of generality, denoting the variable of interest y_t and the relevant controls S_t:

$$\mathrm{E}[y_t|S_t,\psi,\pi] = \sum_{k=1}^{K} \overline{y}_t(k|K) \times P[y_{k-1} \leq Y < y_k|S_t]. \quad (19)$$

The marginal effects reported in the next section are calculated analytically by setting the intercepts to produce expectations at the 25th, 50th, and 75th percentiles of the sample distributions of medical care spending, consumption and health. We then use the estimated coefficients from the full model to calculate the change in the expected value from the change in a given variable. Marginal effects for continuous variables are calculated by adding 10% to their values, discrete variables are adjusted by one.

Distributional effects (kernel densities and surface plots in subsections 4.2 and 4.3, and the simulation in subsection 4.4) are calculated by replicating each observation in the data 40 times. We forward simulate the model using the observed values of the exogenous variables and estimated parameters to generate predicted distributions of medical care spending for 100 points in the state space of health and reference health:

$$(H_t = x, R_t = y), \forall x, y \in \{(0.1, 0.1); (0.1, 0.2); \ldots (0.1, 1.0); (0.2, 0.1); \ldots (1.0, 1.0)\}. \quad (20)$$

Values of (H_t, R_t) near the sample mean are used to illustrate the full distributional effects in Figure 3. All such values of (H_t, R_t) are used in forming the surface plots in Figure 4 and corresponding values in Tables 7–8.

Finally, we also compare the predictions of our preferred specification to a model without reference health.[20] For this exercise, we randomly draw 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 coefficients of the model to forward simulate the individual's decisions to consume medical care and non-medical goods.

4. Results

Unlike linear regression where a single number yields an 'average marginal effect', CDE results are not easily encapsulated by a single number in a table. A complete discussion of the full results involves the effects of x on several aspects of the conditional distribution of the outcome of interest. We present these results in four sections. First, our CDE estimator allows us to present marginal effects at three points

[20] 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.

of the distribution of our dependent variables for medical care, consumption and health.[21] Second, using our model estimates, we simulate the full distribution of medical care spending and illustrate the relative effects of contemporaneous and reference health on the distribution. Third we focus on the top tail of the medical spending distribution and show how reference health helps predict high medical care spending. Finally, we conduct a policy-relevant simulation of medical care spending under different health trajectories that demonstrates that excluding reference health will underestimate potential cost savings from healthy aging initiatives that aim to smooth inevitable health declines. Taken together, these results show that reference health is both statistically significant and economically useful in modelling the demand for medical care and better understanding the top spenders.

4.1. *Marginal Effects*

Taking advantage of our CDE estimator, we report three marginal effects: one at the 25th percentile, one at the median, and one at the 75th percentile of the distributions for medical care spending, consumption and health. Each marginal effect is the mean percentage change in the value of the dependent variable, conditional on the dependent variable being observed at the quartile threshold. We focus our discussion on the effects of health and reference health on medical care spending, consumption and the dynamic evolution of health as reported in Table 6 and present complete marginal effects for all covariates in Tables D6–D7 in online Appendix D.

As hypothesised, we find that reference health has a positive marginal effect on the demand for medical care conditional on contemporaneous health and other controls.[22] In other words, at any level of health, individuals whose health was higher in prior periods are predicted to consume more medical care. Moreover, the effect of reference health on the demand for medical care is not constant over the distribution of medical care spending. Again, consistent with our theoretical hypothesis, the effect of reference health is highest in the top quartile of medical care spending (13% *versus* 5.6% in the bottom quartile).[23] Notably, the effect of contemporaneous health also varies over the distribution of medical care spending. The effect of contemporaneous health is negative, which is intuitive in that those in better health need less medical care, but the negative effect is much greater at the top of the distribution.

The next three columns of Table 6 report the marginal effects on non-medical consumption. We find that both contemporaneous and reference health have positive, but small marginal effects on consumption across the distribution up to the top quartile. While the positive marginal effects are consistent with the complementarity of health and consumption, the negative effect of health in the top quartile of consumption is consistent with consumption smoothing among individuals who are dis-saving. Recall our theoretical hypothesis that if $U^{zR} \geq 0$ and $\partial m^*/\partial R_t > 0$ then U^{HR}

[21] Per the discussion in subsection 3.6 on post-estimation methodology, estimated coefficients are included in online Appendix D.

[22] Coefficient estimates in online Appendix D show that reference health is statistically significant with $p < 0.01$ in all features of the medical care and consumption equations.

[23] For example, a 10 percentage point increase in reference health will increase medical care spending of an individual who was *ex ante* expected to spend in the top quartile by 13%.

Table 6

Marginal Effects of Health and Reference Health

	Medical care			Consumption		
Variables	Bottom quartile (%)	Inter quartile (%)	Top quartile (%)	Bottom quartile (%)	Inter quartile (%)	Top quartile (%)
Panel (*a*): Marginal effects of health and reference health on m_t and c_t						
Health	−6.1	−13.3	−19.8	0.5	0.3	−0.2
Reference health	5.6	8.2	13.0	0.6	0.5	0.0

	Bottom quartile (%)	Inter quartile (%)	Top quartile (%)
Panel (*b*): Marginal effects of health and reference health on H_{t+1}			
Health	13.3	8.9	5.6
Reference health	9.1	6.2	3.9

Notes. Marginal effects are calculated analytically using the midpoint of each cell to form expectations. Discrete variables are 'bumped' from zero to one. Continuous variables are 'bumped' up by 10%. The percentage change is calculated as:

$$[E(Y|x+\Delta x, F(Y)\approx j) - E(Y|x, F(Y)\approx j)]/E(Y|x, F(Y)\approx j),$$

where $j = 0.25, 0.5, 0.75$. $F(Y) \approx j$ implies that the dependent variable takes it's 25th, 50th or 75th percentile value before changing the variable of interest. For reference these values are \$230, \$980, and for \$2,620 for m_t, \$7,900, \$25,209, and \$71,212 for consumption, and 0.694, 0.807, and 0.895 for H_t.

must be positive. Thus, these results support our hypothesis that the cross-partial effect of reference health on the marginal utility from health is positive: the greater past health, the greater the marginal utility from any increase in contemporaneous health.

Finally, we find that both contemporaneous and reference health have positive marginal effects on health next period. This result indicates health is highly persistent: individuals who have been in historically better health will have better health in the future. We offer two interpretations for this result. From a health production perspective, reference health may correct some measurement error in contemporaneous health by providing additional information about the individual's true latent health state. However, if what we term 'reference health' only provides additional information about contemporaneous health, then we would expect individuals with higher reference health to consume less medical care, which is contrary to what we find in panel (*a*).[24,25] Because we find that reference health has a positive effect on both

[24] For example, take two individuals with contemporaneous health values of 0.7. One individual has a reference health value of 0.7, the other of 0.9. Our results are consistent with the person with the higher reference health being in better contemporaneous health. The person with the higher reference health, by virtue of being in better health, will face lower expected marginal utility from consuming medical care. Therefore, individuals in greater reference health should consume less medical care unless reference health also affects the demand for medical care.

[25] Similarly, reference health may provide information about the individual's health trajectory. An individual experiencing a decline in health from reference levels may feel compelled to consume more medical care lest their downward trajectory continue. 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 because this individual is consuming more medical care. Our econometric methods address the endogeneity between the demand for health and the demand for medical care.

health next period and the demand for medical care, it is plausible that reference health operates through both the health production and utility mechanisms.

4.2. *Reference Health and the Distribution of Medical Care Spending*

We motivated the use of CDE by emphasising the need to model the distribution of medical care spending, not just the conditional mean. Figure 3 uses the simulations (see subsection 3.6 for post-estimation simulation methodology) from a few key points in the health/reference health state space to illustrate the distributional effects of reference health and the relative importance of reference health and contemporaneous health in shaping the distribution. In each panel of Figure 3, the solid line represents the medical care spending distribution for a base case where both contemporaneous and reference health are near our sample mean: $(H_t, R_t) = (0.8, 0.8)$. In panel (a), we hold contemporaneous health constant and shift reference health by ± 0.2, holding and all other x fixed. In this Figure the base case is in the middle, the dashed line reflects improving health $(H_t, R_t) = (0.8, 0.6)$, and the dotted line reflects declining health (0.8, 1.0). This Figure isolates the effect of reference health on the distribution of medical care spending and illustrates that higher levels of reference health shift the whole spending distribution, including the mode, to the right. Thus, reference health can explain why individuals with the same contemporaneous health can nonetheless demand different amounts of medical care.

In panel (b), we isolate the effect of contemporaneous health on medical care spending by shifting both contemporaneous and reference health in tandem, thereby eliminating any change in health.[26] Panel (b) shows that a decline in health shifts the distribution to the right.

Finally, in panel (c), we combine the effects of health and reference health. The solid line remains our base case while the dotted line is (0.7, 0.7) from panel (b) and the dashed line is the (0.8, 1.0) from panel (a). The second two scenarios yield very similar spending distributions. Thus, panel (c) demonstrates that a 10 percentage point decrease in the individual's health state and a 20 percentage point increase in reference health have very similar effects on the overall distribution of medical care spending.

4.3. *Evidence on the Top* 5 %

Medical care spending is so skewed that the distributions in Figure 3 do not offer any real visibility into the top 5%, which accounts for nearly half of all medical care

[26] Failing to adjust reference health with contemporaneous health leads to misleading results as the individual's utility is affected by both contemporaneous health and health relative to reference health. Suppose instead of changing (H_t, R_t) from (0.8, 0.8) to (0.7, 0.7), we decrease contemporaneous health from 0.8 to 0.7, leaving reference health fixed at 0.8. Individuals will now consume more medical care for two reasons: the marginal utility from improved health has increased due to decreased contemporaneous health, and individuals experience additional disutility from being in a worse health state than their reference point. By changing the individual's health state from (0.8, 0.8) to (0.7, 0.7), we remove the confounding effect of the difference between health and reference health. This makes for a better 'apples to apples' comparison between the effects of contemporaneous and reference health.

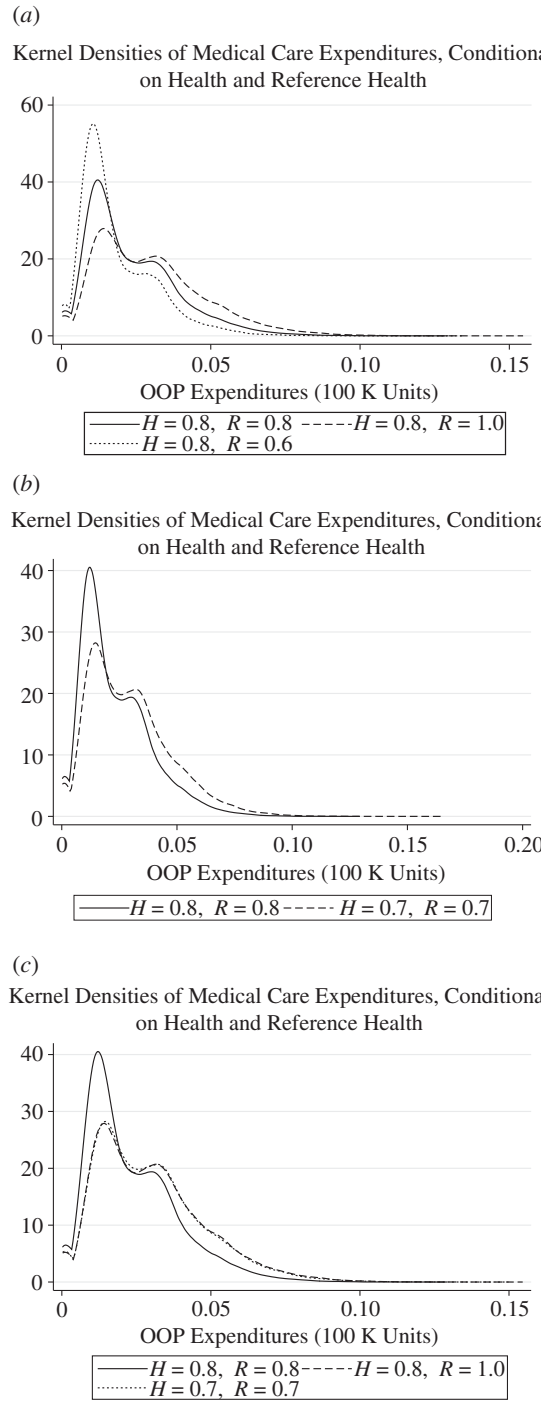

Fig. 3. *Distributional Effects of Health and Reference Health. (a) Distributional Effects of Reference Health. (b) Distributional Effects of Health, Restricting $H_t = R_t$. (c) Comparing the Effects of Health and Reference Health*

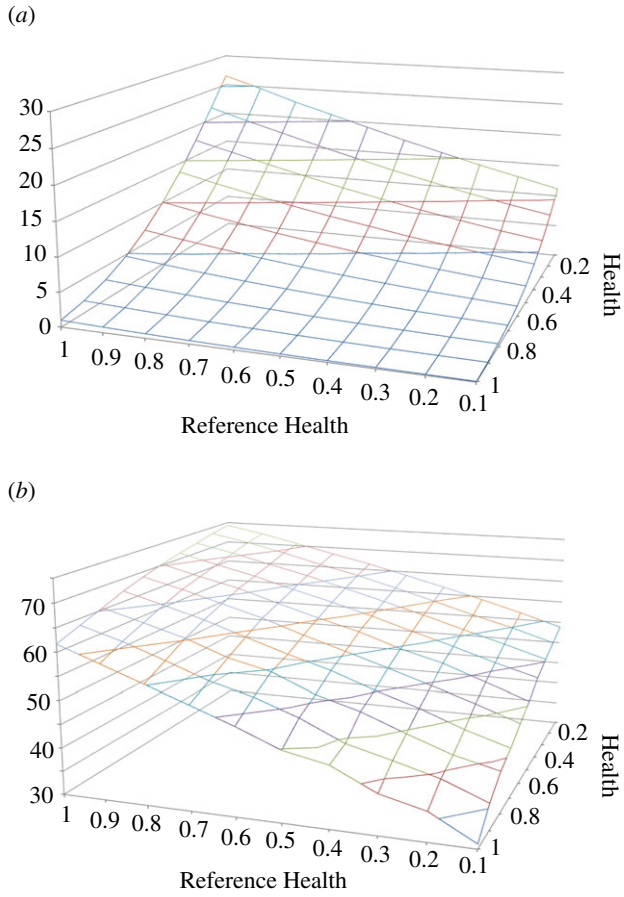

Fig. 4. *Surface Plots – % Probability of Being in the Top 5% of Medical Care Consumers, Conditional on Health and Reference Health. (a) P(Top 5%|**X**, H, R). (b) P(Top 5%|Top 10%, **X**, H, R)*
Note. Colour figure can be viewed at wileyonlinelibrary.com

spending, both in our data and in the population. Figure 4 depicts probabilities that an individual is in the top 5% of medical care spending for each (H_t, R_t) combination defined in (20). Reference health is on the front axis, contemporaneous health on the right receding axis, and the probability of being in the top 5% on the vertical axis.[27] Because the medical care spending distribution is skewed even in the top tail, we further condition probabilities of being in the top 5% on already being in the top decile in panel (*b*). If reference health was irrelevant to predicting top 5% medical care spending, these surface plots would be flat with respect to reference health. Rather, these panels show how higher reference health increases the probability of high medical care spending at every level of contemporaneous health.

[27] Numerical values underlying each panel are available in Tables 7–8.

Table 7

Percentage Probability of Being Observed in the Top 5% of Medical Care Consumers, by Health and Reference Health (Values for Figure 4(a))

	Reference health									
Health	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
1	1.00	0.80	0.64	0.50	0.39	0.30	0.23	0.18	0.14	0.10
0.9	1.78	1.49	1.22	0.98	0.80	0.63	0.49	0.38	0.30	0.23
0.8	2.96	2.53	2.12	1.75	1.47	1.16	0.94	0.76	0.59	0.47
0.7	4.74	4.08	3.47	2.93	2.46	2.05	1.68	1.35	1.10	0.87
0.6	7.13	6.22	5.38	4.61	3.92	3.32	2.78	2.30	1.92	1.55
0.5	10.07	8.92	7.85	6.82	5.93	5.09	4.36	3.70	3.08	2.57
0.4	13.58	12.16	10.88	9.65	8.49	7.48	6.43	5.56	4.74	4.04
0.3	17.59	16.05	14.49	13.01	11.60	10.31	9.08	7.96	6.88	5.96
0.2	22.00	20.19	18.51	16.79	15.19	13.61	12.19	10.78	9.53	8.39
0.1	26.53	24.68	22.68	20.94	19.17	17.41	15.76	14.15	12.65	11.17

Table 8

Percentage Probability of Being Observed in the Top 5% of Medical Care Consumers, Conditional on Being in the Top 10%, Health and Reference Health (Values for Figure 4(b))

	Reference health									
Health	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
1	61.68	58.45	55.18	51.91	48.58	44.83	42.96	38.45	36.16	31.02
0.9	63.58	60.89	57.70	54.34	51.91	48.59	44.36	41.95	39.38	35.64
0.8	65.37	62.85	60.27	56.90	54.74	50.85	48.20	44.59	41.23	38.96
0.7	67.32	64.82	62.27	59.21	56.25	53.56	50.20	47.25	43.70	40.77
0.6	68.85	66.59	63.99	61.45	58.34	55.53	52.61	49.63	46.59	43.25
0.5	70.24	67.88	65.74	62.93	60.25	57.84	54.71	52.26	48.88	45.51
0.4	71.37	69.15	66.99	64.45	62.06	59.49	56.63	53.88	50.94	48.30
0.3	72.37	70.46	68.24	66.08	63.60	61.09	58.51	55.77	53.08	50.30
0.2	73.44	71.67	69.62	67.23	65.13	62.63	60.04	57.33	54.90	52.04
0.1	74.30	72.58	70.52	68.44	66.38	64.10	61.74	59.28	56.64	53.85

Panel (*a*) shows that individuals in good health have small probabilities of being in the top 5% of medical care spenders. As health declines and reference health increases, the probability of high spending increases. For example, from the Tables underlying the surface plots, 7–8, an individual near our sample mean of (0.8, 0.8) has a 2.12% probability of being in the top 5%. However, if that person had a 10 percentage point higher reference health of 0.9, then her probability would increase by 20% to 2.53%. For comparison, the pure effect of a decline in contemporaneous health would be moving from (0.8, 0.8) to (0.7, 0.7) or from 2.12 to 2.93, a 38% change in the probability of being in the top 5%. Thus, the effect of a change in reference health is almost 50% of the magnitude of the effect of the change in contemporaneous health. If individuals are already in the top decile of spending, reference health is as important as health in determining the probability of being observed in the top 5%.

4.4. *Simulation of Medical Care Spending for Different Health Trajectories*

A central 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. Thus, policy initiatives such as those to promote healthy aging are 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 suggest that omitting reference health from models of medical care spending will underestimate the cost savings. This will lead to under-investment in healthy aging initiatives that would result in decreased aggregate lifecycle spending.

To explore our hypothesis on healthy aging, 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 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. We then sum the expenses over the simulation period. The spending estimates for each period and trajectory are reported in Table 9.

These simulations illustrate several key features of the reference health model. First, in period one when there is no effect of reference health (by construction), there is little 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 capture variation in medical care spending among individuals in similar health states. The model without reference health provides similar estimates for all periods and all trajectories when $H_t = 0.45(m_t \in [0.0422 - 0.04441])$. By contrast, the model with reference health

Table 9

Simulated Average Medical Care Expenditures with and without Reference Health Under Different Health Trajectories

	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 reference health	0.0143	0.0146	0.0151	0.0707	0.0591	0.0469	0.2207
m_t without reference health	0.0148	0.0153	0.0157	0.0422	0.0435	0.044	0.1755
Moderate decline							
H_t	0.95	0.95	0.70	0.70	0.45	0.45	
m_t with reference health	0.0143	0.0147	0.0339	0.0307	0.058	0.0524	0.2043
m_t without reference health	0.0148	0.0151	0.0265	0.0271	0.0437	0.0441	0.1713
Gradual decline							
H_t	0.95	0.85	0.75	0.65	0.55	0.45	
m_t with reference health	0.0143	0.0202	0.0273	0.0346	0.0441	0.0538	0.1946
m_t without reference health	0.0148	0.0186	0.0237	0.0294	0.0364	0.0439	0.1668

Note. Values are in units of \$100,000.

exhibits considerable variation in medical care spending, even among individuals in poor health ($m_t \in [0.0469 - 0.0707]$). This result reflects the stylised fact 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.

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 4.9% over the simulation period. By contrast, the model with reference health predicts cost savings of 11.8% (0.195 compared to 0.221). Thus, our results indicate that modelling the savings from healthy ageing initiatives without reference health will understate the impact by more than half. Population ageing trends make this result highly policy relevant.

5. 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 modelled. Intuitively it is reasonable to expect that the value of health, like the value of wealth modelled in other literatures, is relative to past realisations of health rather than absolute. If the value an individual attributes to contemporaneous health depends on an individual's reference health, then reference health would be a significant input to any decision that also includes health as a factor.

Our theoretical argument and discussion of empirical results is centred on the idea that reference health can explain the demand for medical care by increasing the marginal utility from health. Our theoretical argument mirrors that from the prospect theory, habit formation and rational addiction literatures. Our theory illustrates how the cross partial utility of health and reference health can contribute to higher demand for medical care. Furthermore, the theory suggests that the effect of reference health would be greater at higher levels of medical care spending. However, we acknowledge that ruling out all other potential mechanisms that explain our empirical result is beyond our current scope. Future research could further explore other potential mechanisms whereby reference health affects the demand for medical care.

Empirically, we demonstrate that reference health is both statistically significant and economically relevant to modelling the full distribution of medical care spending. Our econometric methods address the substantial challenges of estimating a dynamic model with skewed variables of interest and multiple sources of unobservable heterogeneity. We nest the estimation of joint demands in a finite mixture framework that allows for both permanent and time varying unobservable heterogeneity. We use CDE to estimate the marginal effects of the covariates, including reference health, at different points of the distribution of our dependent variables. Validating our use of

CDE, we find that the marginal effects of reference health vary over the distributions of medical care spending, consumption spending, and health.

We find that the marginal effect of reference health on medical care spending is positive, consistent with our theoretical hypothesis. The greater the reference health the more medical care spending at all parts of the spending distribution. By simulating the distribution of medical care spending with different combinations of contemporaneous and reference health, we illustrate that reference health shifts both the mode and the mass of the distribution towards higher spending. The shift in the medical care spending distribution associated with a 20 percentage point increase in reference health is comparable to the shift associated with a 10 percentage decrease in contemporaneous health. We also find that higher reference health predicts a higher probability of an individual being in the top 5% of the spending distribution at every level of contemporaneous health. A 10 percentage point increase in reference health increases the probability of being in the top 5% by approximately 19.3%. We find that the predictive power of reference health, relative to contemporaneous health, is strongest at the top end of the distribution. Conditional on being in the top decile of medical care spending, the incremental effects of a 10 percentage point increase in reference health and a 10 percentage point decrease in contemporaneous health on the probability of top 5% spending are nearly equivalent.

The policy implication of the relevance of reference health is that while poor health matters, the path to poor health matters too. We illustrate the impact of this finding by simulating medical care spending for three different health trajectories with and without using reference health. Omitting reference health underestimates the potential cost-savings of healthy aging initiatives by over 50%.

Our findings suggest several interesting avenues for future work. First, our marginal effects show variations across the distributions of medical care spending, consumption and health. In particular, the varying effects of contemporaneous and reference health on the distribution of consumption may help reconcile conflicting findings in the literature about whether health and consumption are complements or substitutes; they may be either depending on health dynamics and the distribution of health and consumption. In addition, we found varying marginal effects of age, income, insurance, marital status, race, and education on medical care spending. Exploring the distribution of medical care spending associated with changes in these covariates may expose additional policy-relevant insights into the health/wealth gradient, the age *versus* time-to-death argument, and additional levers to reduce spending. Another future avenue is to model reference health with respect to a social rather than individual measure of health. Modelling a social reference point is in line with a large clinical and economic literature on social network effects on obesity and risky behaviours such as smoking. A social reference point for health suggests a game theoretic model where an individual's choice is influenced by, and in turn influences, the choices of others. A social reference point for health may also help to explain differences in medical care demand across regions and countries. Researchers should also continue to dig deeper into how to measure health and reference health with respect to specific medical conditions including chronic conditions such as asthma or cancer. Work exploring changes in specific health indicators can shed more light on the different mechanisms other than utility whereby reference health may empirically

affect the demand for medical care including those explored in online Appendix E. Finally, the recent work on the annuity puzzle and asset allocations in retirement may benefit from modelling medical care expenditures using reference health. More generally, our theoretical and empirical results suggest that economic models that include health should also include reference health to reflect that the marginal utility of health, like wealth, is not absolute, but relative to reference values.

University of Tennessee
Drew University

Accepted: 26 July 2017

Additional Supporting Information may be found in the online version of this article:

Appendix A. Derivation of the Effect of Reference Health on the Demand for Medical Care.
Appendix B. Formation of the Health Index.
Appendix C. Formation of Occupational Demands.
Appendix D. Coefficient Estimates and Other Marginal Effects from CDE.
Appendix E. Alternative Mechanisms for Reference Health.
Data S1.

References

Acemoglu, D., Finkelstein, A. and Notowidigdo, M. (2013). 'Income and health spending: evidence from oil price shocks', *Review of Economics and Statistics*, vol. 95(4), pp. 1079–95.
Banthin, J. and Bernard, D.M. (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*, vol. 296 (22), pp. 2712–9.
Baucells, M., Weber, M. and Welfens, F. (2011). 'Reference-point formation and updating', *Management Science*, vol. 57(3), pp. 506–19.
Becker, G.S. and Murphy, K.M. (1988). 'A theory of rational addiction', *Journal of Political Economy*, vol. 96(4), pp. 675–700.
Bound, J., Schoenbaum, M., Stinebrickner, T. and Waidmann, T. (1999). 'The dynamic effects of health on the labor force transitions of older workers', *Labour Economics*, vol. 6(2), pp. 179–202.
Cameron, A. and Johansson, P. (1997). 'Count data regression using series expansions: with applications', *Journal of Applied Econometrics*, vol. 12(3), pp. 203–23.
Cameron, A. and Trivedi, P. (1986). 'Econometric models based on count data: comparisons and applications of some estimators and tests', *Journal of Applied Econometrics*, vol. 1(1), pp. 29–53.
Caputo, M.R. (2005). *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*, Cambridge: Cambridge University Press.
Claxton, G., Kamal, R. and Cox, C. (2014). 'How health spending patterns vary by demographics in the US', Technical Report, Kaiser Family Foundation, available at: http://www.healthsystemtracker.org/2014/12/how-health-spending-patterns-vary-by-demographics-in-the-u-s/ (last accessed: 3 November 2017).
Cohn, S. and Yu, W. (2012). 'The concentration and persistence in the level of health expenditures over time: estimates for the US population 2008–2009', AHRQ Statistical Brief No. 354.
Constantinides, G.M. (1990). 'Habit formation: a resolution of the equity premium', *Journal of Political Economy*, vol. 98(3), pp. 519–43.
Crawford, V.P. and Meng, J. (2011). 'New York city cab drivers' labor supply revisited: reference-dependent preferences with rational-expectations targets for hours and income', *American Economic Review*, vol. 101 (5), pp. 1912–32.
Dai, Q. and Grischenko, O.V. (2014). 'An empirical investigation of consumption-based asset pricing', *Quarterly Journal of Finance*, vol. 4(1), pp. 1–34.
Dardanoni, V. and Wagstaff, A. (1990). 'Uncertainty and the demand for medical care', *Journal of Health Economics*, vol. 9(1), pp. 23–38.

Darden, M.E. (2017). 'Smoking, expectations, and health: a dynamic stochastic model of lifetime smoking behavior', *Journal of Political Economy*, vol. 125(5), pp. 1465–1522.
de Meijer, C., O'Donnell, O., Koopmanschap, M. and van Doorslaer, E. (2013). 'Health expenditure growth: looking beyond the average through decomposition of the distribution', *Journal of Health Economics*, vol. 32(1), pp. 88–105.
Deb, P. and Trivedi, P. (1997). 'Demand for medical care by the elderly: a finite mixture approach', *Journal of Applied Econometrics*, vol. 12(3), pp. 313–56.
Desmond, K., Rice, T., Cubanski, J. and Neuman, P. (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., French, E. and Jones, J. (2010). 'Why do the elderly save? The role of medical expenses', *Journal of Political Economy*, vol. 112(2), pp. 379–444.
Edwards, R. (2008). 'Health risk and portfolio choice', *Journal of Business and Economic Statistics*, vol. 26(4), pp. 472–85.
Ehrlich, I. and Chuma, H. (1990). 'A model of the demand for longevity and the value of life extensions', *Journal of Political Economy*, vol. 98(4), pp. 761–82.
Finkelstein, A., Luttmer, E. and Notowidigdo, M. (2013). 'What good is wealth without health? The effect of health on the marginal utility of consumption', *Journal of the European Economic Association*, vol. 11(1), pp. 221–58.
French, E. and Jones, J.B. (2004). 'On the distribution and dynamics of health care costs', *Journal of Applied Econometrics*, vol. 19(6), pp. 705–21.
Galama, T. (2015). 'A contribution to health capital theory', CESR-Schaeffer Working Paper No. 2015-004.
Gilleskie, D.B. (1998). 'A dynamic stochastic model of medical care use and work absence', *Econometrica*, vol. 66(1), pp. 1–46.
Gilleskie, D.B. and Mroz, T.A. (2004). 'A flexible approach for estimating the effects of covariates on health expenditures', *Journal of Health Economics*, vol. 23(2), pp. 319–418.
Greenacre, M. and Blasius, J. (2006). *Multiple Correspondence Analysis and Related Methods*, London: Chapman & Hall/CRC Taylor & Francis Group.
Grossman, M. (1972). 'On the concept of health capital and the demand for health capital', *Journal of Political Economy*, vol. 80(2), pp. 223–55.
Gurmu, S. (1997). 'Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization', *Journal of Applied Econometrics*, vol. 12(1), pp. 225–42.
Hall, R. and Jones, C. (2007). 'The value of life and the rise in health spending', *Quarterly Journal of Economics*, vol. 122(1), pp. 39–72.
Heckman, J. and Singer, B. (1984). 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica*, vol. 52(2), pp. 271–320.
Hugonnier, J., Pelgrin, F. and StAmour, P. (2013). 'Health and (other) asset holdings', *Review of Economic Studies*, vol. 80(2), pp. 663–710.
Jones, A., Lomas, J. and Rice, N. (2015). 'Healthcare cost regressions: going beyond the mean to estimate the full distribution', *Health Economics*, vol. 24(9), pp. 1192–212.
Kahneman, D. and Tversky, A. (1979). 'Prospect theory: an analysis of decision under risk', *Econometrica*, vol. 47(2), pp. 263–92.
Khwaja, A. (2010). 'Estimating willingness to pay for Medicare using a dynamic life-cycle model of demand for health insurance', *Journal of Econometrics*, vol. 156(1), pp. 130–47.
Kohn, J.L. (2012). 'What is health: a multiple correspondence health index', *Eastern Economic Journal*, vol. 38(2), pp. 223–55.
Kohn, J.L. and Liu, J. (2013). 'The dynamics of medical care use in the British household panel survey', *Health Economics*, vol. 22(6), pp. 687–710.
Kohn, J.L. and Patrick, R.H. (2008). 'Health and wealth: a dynamic demand for medical care', Working Paper, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract-id=994723 (last accessed: 3 November 2017).
Koszegi, B. and Rabin, M. (2006). 'A model of reference-dependent preferences', *Quarterly Journal of Economics*, vol. 121(4), pp. 1133–65.
Koszegi, B. and Rabin, M. (2009). 'Reference-dependent consumption plans', *American Economic Review*, vol. 99(3), pp. 909–36.
LeCook, B. and Manning, W.G. (2009). 'Measuring racial/ethnic disparities across the distribution of health care expenditures', *Health Services Research*, vol. 44(5), pp. 1603–21.
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*, vol. 92(2), pp. 233–74.
Peijnenberg, K., Nijman, T. and Werker, B.J. (2015). 'Health cost risk: a potential solution to the annuity problem', Economic Journal, vol. 127(607), pp. 1589–625.
Pohlmeier, W. and Uhlrich, V. (1995). 'An econometric model of the two-part decision making process in the demand for health care', *Journal of Human Resources*, vol. 30(2), pp. 339–61.

Ryder, H.E. and Heal, G.M. (1973). 'Optimal growth with intertemporally dependent preferences', *Review of Economic Studies*, vol. 40(1), pp. 1–31.
Schoenman, J. (2012). 'The concentration of health care spending', Brief, National Institute for Health Care Management, available at: http://www.nihcm.org/pdf/DataBrief3%20Final.pdf (last accessed: 3 November 2017).
Shen, C. (2013). 'Determinants of health care decisions: insurance, utilization, and expenditures', *Review of Economics and Statistics*, vol. 95(1), pp. 142–53.
Viscusi, W.K. and Evans, W.N. (1990). 'Utility functions that depend on health status: estimates and economic implications', *American Economic Review*, vol. 80(3), pp. 353–74.
Wouterse, B., Huisman, M., Meijboom, B.R., Deeg, D.J. and Polder, J.J. (2013). 'Modeling the relationship between health and health care expenditures using a latent markov model', *Journal of Health Economics*, vol. 32(2), pp. 423–39.
Yang, Z., Gilleskie, D.B. and Norton, E.C. (2009). 'Health insurance, medical care, and health outcomes', *Journal of Human Resources*, vol. 44(1), pp. 47–114.
Yogo, M. (2016). 'Portfolio choice in retirement: health risk and the demand for annuities, housing, and risky assets', *Journal of Monetary Economics*, vol. 80(C), pp. 17–34.