NBER WORKING PAPER SERIES

THE LIFETIME MEDICAL SPENDING OF RETIREES

John Bailey Jones
Mariacristina De Nardi
Eric French
Rory McGee
Justin Kirschner

Working Paper 24599
http://www.nber.org/papers/w24599

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018

De Nardi and French gratefully acknowledge support from Norface Grant (TRISP 462-16-120). French gratefully acknowledges support from the Economic and Social Research Council (Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (RES-544-28-50001) and from Inequality and the insurance value of transfers across the life cycle (ES/P001831/1)). The views expressed in this paper are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago, the Federal Reserve Bank of Richmond, the IFS, the CEPR, or the National Bureau of Economic Research.

The Lifetime Medical Spending of Retirees
John Bailey Jones, Mariacristina De Nardi, Eric French, Rory McGee, and Justin Kirschner
NBER Working Paper No. 24599
May 2018
JEL No. D1,D14,E02,E2,H31

ABSTRACT

Using dynamic models of health, mortality, and out-of-pocket medical spending (both inclusive and net of Medicaid payments), we estimate the distribution of lifetime medical spending that retired U.S. households face over the remainder of their lives. We find that at age 70, households will on average incur $122,000 in medical spending, including Medicaid payments, over their remaining lives. At the top tail, 5 percent of households will incur more than $300,000, and 1 percent of households will incur over $600,000 in medical spending inclusive of Medicaid. The level and the dispersion of this spending diminish only slowly with age. Although permanent income, initial health, and initial marital status have large effects on this spending, much of the dispersion in lifetime spending is due to events realized later in life. Medicaid covers the majority of the lifetime costs of the poorest households and significantly reduces their risk.

John Bailey Jones
Federal Reserve Bank of Richmond
P.O. Box 27622
Richmond, VA 23261
john.jones@rich.frb.org

Mariacristina De Nardi
Federal Reserve Bank of Chicago
230 South LaSalle St.
Chicago, IL 60604
and University College London
and also NBER
denardim@nber.org

Eric French
University College London, CEPR, and IFS
30 Gordon Street
London, WC1H 0AX
United Kingdom
Eric.French.econ@gmail.com

Rory McGee
Department of Economics
Drayton House
30 Gordon St
London WC1H 0AX
United Kingdom
rory.mcgee.13@ucl.ac.uk

Justin Kirschner
Federal Reserve Bank of Richmond
P.O. Box 27622
Richmond, Virginia 23261
justin.kirschner@rich.frb.org

1 Introduction

Despite nearly universal enrollment in the Medicare program, most elderly Americans still face the risk of catastrophic health care expenses. There are many gaps in Medicare coverage: for example, Medicare does not pay for long hospital and nursing home stays, and requires co-payments for many medical goods and services. Medical spending is thus a major financial concern among elderly households. In a recent survey, more affluent individuals were worried about rising health care costs than about any other financial issue (Merrill Lynch Wealth Management, 2012).

Several papers (De Nardi et al. 2010, Kopecky and Koreshkova 2014, Ameriks et al. (2015)) show that health care costs that rise with age and income explain much of the U.S. elderly's saving behavior.[1] Differences in medical spending risk are also important in explaining cross country differences in the consumption (Banks et al., 2016) and savings decisions (Nakajima and Telyukova, 2018) of elderly households. More generally, the literature on the macroeconomic implications of health and medical spending is growing rapidly. Recent studies have considered important questions such as: bankruptcy (Livshits et al., 2007); the adequacy of savings at retirement (Skinner 2007, Scholz et al. 2006); annuitization (Pashchenko 2013, Lockwood 2012, Reichling and Smetters 2015); portfolio choice (Love, 2009); optimal taxation of health (Boerma and McGrattan, 2018); health insurance reform (Pashchenko and Porapakkarm 2013, Jung and Tran 2016, Conesa et al. 2017); and portfolio choice (Hugonnier et al., 2012).

All of the aforementioned studies rely on accurate measures of medical risk and medical spending. But even though there is a large literature documenting annual medical spending at older ages, there has been relatively little work documenting the distribution of cumulative lifetime spending. Yet it is in many ways lifetime totals, rather than spending in any given year, that are most important to saving decisions and household financial well-being. Households care not only about the risk of catastrophic expenses in a single year, but also about the risk of moderate and persistent expenses that accumulate into catastrophic lifetime costs.

In this paper we estimate the distribution of lifetime medical spending for retired households whose heads are 70 or older. Our focus is out-of-pocket spending, the payments made by households themselves. High out-of-pocket expenses, however, can leave

[1] Additional mechanisms proposed to explain the "elderly savings puzzle," or the slow decumulation of assets in old age, include bequests (De Nardi 2004 and Lockwood 2012) and the desire of older individuals to remain in their current homes (Nakajima and Telyukova, 2012). See De Nardi et al. (2016b) for a review.

households financially indigent and reliant on Medicaid, the means-tested public insurance program. Our spending estimates therefore include payments made by Medicaid, to fully capture all of the medical spending risk that households potentially face. Our estimates thus measure the medical spending risk that wealthier households would face, and the medical spending risk that less wealthy households would face were Medicaid not available (absent any other changes in their insurance).

Our main dataset is the Health and Retirement Survey (HRS) which has high quality information on out-of-pocket medical spending over the period 1995 to 2014. Because the HRS does not have Medicaid payment data, we impute Medicaid payments using the the Medicare Current Beneficiary Survery (MCBS). Our data allow us to estimate dynamic models of health, mortality, and out-of-pocket medical spending. Medical spending depends on age, household composition, health, and idiosyncratic shocks. Simulating our estimated models allows us to construct household histories, calculate discounted sums, and ultimately to compute the distribution of lifetime medical spending.

This paper uses the estimated model of De Nardi et al. (2018), which builds on earlier analyses of the HRS data by French and Jones (2004) and De Nardi et al. (2010, 2016a). French and Jones (2004) show that medical spending shocks are well described by the sum of a persistent AR(1) process and a white noise shock.[2] They also find that the innovations to this process can be modelled with a normal distribution that has been adjusted to capture the risk of catastrophic health care costs. Simulating this model, they find that in any given year 0.1% of households receive a health cost shock with a present value of at least $125,000 (in 1998 dollars). That paper abstracts away from much of the variability in costs coming from demographics or observable measures of health. De Nardi et al. (2010, 2016a) extend the spending model to account for health and lifetime earnings, but consider only singles and do not control for end-of-life events (see Poterba et al. (2017) on the importance of these events). The model used here addresses both shortcomings by including couples and singles and accounting for the additional medical expenditures incurred at the end of life.

Closely related papers include Fahle et al. (2016), who document the HRS medical spending data in some detail, and Hurd et al. (2017), who use the HRS to calculate the lifetime incidence and costs of nursing home services. Alemayehu and Warner (2004) construct a measure of lifetime spending by combining data from the MCBS and the Medical Expenditure Panel Survey with detailed data for patients in Michigan. However, their estimates only distinguish gender, current age, and age of death, and abstract from

[2]Feenberg and Skinner (1994) find a similar result. See also Hirth et al. (2015).

health and marital status, among other factors.

Of particular note is Webb and Zhivan (2010), who use the HRS to estimate the distribution of lifetime expenses at ages 65 and above. Our paper is complementary to theirs, but differs along two dimensions. The first is methodology. While both papers rely on simulation, our approach is to combine a 3-state model of health (good, bad or nursing home) with a two-component idiosyncratic shock, and to control for socioeconomic status with a measure of permanent income (PI). In contrast, Webb and Zhivan (2010) estimate a rich model of stochastic morbidity and mortality with multiple health indicators and assume that medical expenditures are a function of these health conditions, along with a collection of socioeconomic indicators.[3] In their framework, all of the variation in medical spending is due to variation in these controls; there are no residual shocks. In contrast, in our framework the idiosyncratic shocks capture any spending variation not attributable to age, PI, health, or household composition. The second major difference between our exercise and Webb and Zhivan's (2010) is the spending measure. As discussed above, the HRS data exclude expenses covered by Medicaid, the means-tested public insurance program, which otherwise might have been paid out of pocket. Webb and Zhivan (2010) address this issue by excluding households that receive Medicaid. But all else equal, Medicaid beneficiaries tend to have higher medical expenses, in part because households that face overwhelming medical expenses are more likely to qualify for Medicaid (De Nardi et al. (2016a,c)). Our approach is to impute the missing Medicaid expenditures, using data from the MCBS, and work with the sum of out-of-pocket and Medicaid expenditures. To put our results in context, we also analyze the HRS out-of-pocket spending measure. Comparing the two measures reveals the extent to which Medicaid reduces out-of-pocket expenditures.

We find that lifetime medical spending during retirement is high and uncertain. At age 70, households will on average incur \$122,000 in medical spending, including Medicaid payments, over their remaining lives. At the top tail, 5 percent of households will incur more than \$300,000 and 1 percent of households will incur over \$600,000 in medical spending inclusive of Medicaid. The level and the dispersion of remaining lifetime spending diminishes only slowly with age. The reason for this is that as they age, surviving individuals on average have fewer remaining years of life, but are also more likely to live to extremely old age when medical spending is very high. Although PI, initial health, and initial marital status have large effects on this spending, much of the dispersion in lifetime spending is due to events realized in later years.

[3]Using data from Catalonia, Carreras et al. (2013) perform a similar analysis.

We find that Medicaid lowers average lifetime expenditures by 20 percent. It covers the majority of the medical costs of the poorest households and significantly reduces their risk. Medicaid also reduces the level and volatility of medical spending for high-income households, but to a much smaller extent.

The rest of the paper is organized as follows. In section 2 we discuss some key features of the datasets that we use in our analysis, the HRS and the MCBS, and describe how we construct our measure of medical spending. In section 3 we introduce our model and describe our simulation methodology. We discuss our results in section 4 and conclude in section 5.

2 Data

The medical spending models used here were developed and estimated as inputs for the structural savings model used in De Nardi et al. (2018). Our description of these models and the underlying data thus borrows heavily from the text of that project.

2.1 The HRS

We use data from the Asset and Health Dynamics Among the Oldest Old (AHEAD) cohorts of the HRS. The AHEAD is a sample of non-institutionalized individuals aged 70 or older in 1993. These individuals were interviewed in late 1993/early 1994, and again in 1995/96, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, and 2016. We use data for 10 waves, from 1995/96 to 2014. We exclude data from the 1994 wave because medical expenses are underreported (Rohwedder et al. (2006)), and exclude data from the 2016 wave because they are preliminary.

We only consider retired households, defined as those earning less than $3,000 in every wave. Because our demographic model allows for household composition changes only through death, we drop households who get married or divorced, or report other marital transitions not consistent with the model. Consistency with the demographic model also leads us to drop households who: have large differences in ages; are same-sex couples; or have no information on the spouse. This leaves us with 4,324 households, of whom 1,249 are initially couples and 3,075 are singles.

Households are followed until both members die; attrition for other reasons is low. When the respondent for a household dies, in the next wave an "exit" interview with a knowledgeable party – usually another family member – is conducted. This allows the

HRS to collect data on end-of-life medical conditions and expenditures (including burial costs). Fahle et al. (2016) compare the medical spending data from the "core" and exit interviews in some detail.

The HRS has a variety of health indicators. We assign individuals to the nursing home state if they were in a nursing home at least 120 days since the last interview or if they spent at least 60 days in a nursing home before the next scheduled interview and died before that scheduled interview. We assign the remaining individuals a health status of "good" if their self-reported health is excellent, very good or good and a health status of "bad" if their self-reported health is fair or poor.

The HRS collects data on all out-of-pocket medical expenses, including private insurance premia and nursing home care. The HRS medical spending measure is backward-looking: medical spending at any wave is measured as total out-of-pocket expenditures over the preceding two years. It is thus not immediately obvious whether medical spending reported in any given wave should be expressed as a function of medical conditions reported in that wave, or those reported in the prior wave. Our empirical spending model includes indicators from both sets of dates. French et al. (2017) compare out-of-pocket medical spending data from the HRS, MCBS, and MEPS. They find that the HRS data match up well with data from the MCBS. They also find that the HRS matches up well with the MEPS for items that MEPS covers, but that the HRS is more comprehensive than the MEPS in terms of the items covered.

To control for socioeconomic status, we construct a measure of permanent income (PI). We first find each household's "non-asset" income, a pension measure that includes Social Security benefits, defined benefit pension benefits, veterans benefits and annuities. Because there is a roughly monotonic relationship between lifetime earnings and these pension variables, post-retirement non-asset income is a good measure of lifetime permanent income. We then use fixed effects regression to convert non-asset income, which depends on age and household composition as well as lifetime earnings, to a scalar measure comparable across all households. In particular, we assume that log income for household i at age t follows

$$\ln y_{it} = \alpha_i + \kappa(t, f_{it}) + \omega_{it}, \tag{1}$$

where: α_i is a household-specific effect; $\kappa(t, f_{it})$ is a flexible function of age and family structure f_{it} (i.e., couple, single man, or single woman); and ω_{it} represents measurement error. The percentile ranks of the estimated fixed effects, $\widehat{\alpha}_i$, form our measure of permanent income, $\widehat{I}_i$.

2.2 The MCBS

While the HRS contains reasonably accurate measures of out-of-pocket medical spending, it does not contain Medicaid payments. To circumvent this issue, we use data from the 1996-2010 waves of the MCBS. The MCBS is a nationally representative survey of Medicare beneficiaries. Survey responses are matched to Medicare records, and medical expenditure data are created through a reconciliation process that combines survey information with Medicare administrative files. MCBS respondents are interviewed up to 12 times over a 4-year period, resulting in medical spending panels that last up to three years. We use the same sample selection rules for the MCBS that we use for the HRS data.

The MCBS data include information on marital status, health, health care spending and household income. One drawback of the MCBS is that it does not have information on the medical spending or health of the spouse.

2.3 Our Medical Spending Measure

Because the HRS medical spending data exclude expenses covered by Medicaid, which otherwise might have been paid out of pocket, they are censored. If the incidence of Medicaid were random, we could simply drop Medicaid recipients from our sample. However, this is not the case, because Medicaid beneficiaries tend to have higher medical expenses, in part because households that face overwhelming medical expenses are more likely to qualify for Medicaid (De Nardi et al. (2016a,c)). Our approach is to use MCBS data to impute the missing Medicaid expenditures in the HRS, and to then sum observed out-of-pocket and imputed Medicaid expenditures into a single cost measure. In addition to removing the censoring, our measure allows us to assess the spending risk that older households would face in the absence of Medicaid. Knowing this risk is key to assessing the effects of Medicaid itself.

We proceed in two steps. First, we use the MCBS data to regress Medicaid payments for Medicaid recipients on a set of observable variables found in both datasets. This regression has an a R^2 statistic of 0.67, suggesting that our predictions are fairly accurate. Second, we impute Medicaid payments in the AHEAD data using a conditional mean-matching procedure, a procedure very similar to hot-decking. We combine the regression coefficients with the observables in the HRS to predict Medicaid payments, then add to each predicted value a residual drawn from an MCBS household with a similar value of predicted medical spending. We describe our approach in more detail in Appendix A.

Although our principal spending measure is the sum of out-of-pocket and Medicaid payments, we also analyze out-of-pocket spending by itself. The extent to which out-of-pocket spending is lower and/or less volatile than combined spending directly reflects the extent to which Medicaid shields households from medical expenses.

3 The Model

Our model of lifetime medical spending consists of two parts. The first is a Markov Chain model of health and mortality. The second part is the model of medical expenditure flows, where medical spending over any given interval depends on health, family structure, and the realizations of two idiosyncratic shocks.

3.1 Health and Mortality

Let hs_{it}^{h} and hs_{it}^{w} denote the health of, respectively, the husband h and the wife w in household i at age t. Each person's health status, hs^{g}, has four possible values: dead; in a nursing home; in bad health; or in good health. We assume that the transition probabilities for an individual's health depend on his or her current health, age, household composition, permanent income I, and gender $g \in \{h, w\}$. It follows that the elements of the health transition matrix are given by

$$\pi_{i,j,k}(t, f_{it}, I_i, g) = \Pr\left(hs_{i,t+2}^{g} = k \,\middle|\, hs_{i,t}^{g} = j; t, f_{i,t}, I_i, g\right), \tag{2}$$

with the transitions covering a two-year interval, as the HRS interviews every other year.[4] We estimate health/mortality transition probabilities by fitting the transitions observed in the HRS to a multinomial logit model.[5]

Table 1 shows the life expectancies implied by our demographic model for those still alive at age 70. The first panel of Table 1 shows the life expectancies for singles under different configurations of gender, PI percentile, and age-70 health. The healthy live longer than the sick, the rich (higher PI) live longer than the poor, and women live longer

[4] As discussed in De Nardi et al. (2016a), one can fit annual models of health and medical spending to the HRS data. The process becomes significantly more involved, however, especially when accounting for the dynamics of two-person households.

[5] We do not control for cohort effects. Instead, our estimates are a combination of period (cross-sectional) and cohort probabilities. While our HRS sample covers 18 years, it is still too short to track a single cohort over its entire post-retirement lifespan. This may lead us to underestimate the lifespans expected by younger cohorts as they age. Nevertheless, lifespans have increased only modestly over the sample period. Accounting for cohort effects would have at most a modest effect on our estimates.

than men. For example, a single man at the 10th PI percentile in a nursing home expects to live only 3.0 more years, while a single woman at the 90th percentile in good health expects to live 15.4 more years. The second panel of Table 1 shows the same results for married men and women. Married people live longer, unless they are in a nursing home, in which case the differences are small.[6] The third panel of Table 1 shows life expectancies for married *households*, that is, the average length of time that at least one member of the household is still alive or, equivalently, the life expectancies for the oldest survivors. While wives generally outlive husbands, a non-trivial fraction of the oldest survivors are men, and the life expectancy for a married household is roughly two years longer than that of a married woman.

		Men			Women	
Income Percentile	Nursing Home	Bad Health	Good Health	Nursing Home	Bad Health	Good Health
Single Individuals						
10	3.03	6.92	8.68	4.07	11.29	13.18
50	3.02	7.78	10.29	4.05	12.29	14.86
90	2.91	8.11	10.94	3.80	12.51	15.37
Married Individuals						
10	2.73	7.83	9.82	3.95	12.10	14.05
50	2.77	9.39	12.18	3.99	13.74	16.27
90	2.74	10.39	13.50	3.88	14.59	17.28
Married Households (Oldest Survivor)						
10				4.51	13.94	15.93
50				4.61	15.91	18.40
90				4.50	16.85	19.41

Table 1: Life expectancy in years, conditional on reaching age 70

Another key statistic for our analysis is the probability that a 70-year-old will spend significant time (more than 120 days) in a nursing home before he or she dies. Nursing home incidence differs relatively modestly across the PI distribution. Although high-income people are less likely to be in a nursing home at any given age, they live longer,

[6] The results for couples reported in Table 1 are based on the assumption that the two spouses have the same health at age 70. While our model allows an individual's health transition probabilities to depend on his or her marital status, they do not depend on the spouse's health. Spousal health affects the life expectancy calculations only in that healthy spouses live longer.

and older individuals are much more likely to be in a nursing home. In contrast, the effects of gender are pronounced, as are the effects of marital status for men. While 37% of single women and 36% of married women alive at age 70 will enter a nursing home before they die, the corresponding quantities for single and married men are 26% and 19%, respectively.[7] The differences between men and women are largely driven by the differences in life expectancy and marital status. Because women tend to live longer than men, they are more likely to live long enough to enter a nursing home. Moreover, although being married reduces the probability of entering a nursing home, wives tend to outlive their husbands. Women who are married at age 70 tend to be widows for several years, at which point they face the higher probability of entering a nursing home faced by women who are single at age 70. It is not surprising that the two groups face similar nursing home risk. In contrast, husbands usually die before their wives, so that men married at age 70 rarely face the high risk of transitioning into a nursing home faced by their single counterparts. Individuals initially in good health are 2 to 3 percentage points more likely to spend time in a nursing home than those initially in bad health, as nursing home risk is higher at older ages, and those initially in good health live longer.

Because all households in the HRS are initially non-institutionalized, our estimates understate the fraction of individuals in nursing homes at any age. Our simulations begin with the second wave of the AHEAD cohort, at which point roughly 3% of men and 1% of women in the simulations had entered nursing homes. However, the HRS does a good job of tracking individuals as they enter in nursing homes. French and Jones (2004) show that by 2000 the HRS sample matches very well the aggregate statistics on the share of the elderly population in a nursing home. We also understate the number of nursing home visits because we exclude short-term visits: as Friedberg et al. (2014) and Hurd et al. (2017) document, many nursing home stays last only a few weeks and are associated with lower expenses. We focus only on the longer and more expensive stays faced by households.

3.2 Medical Spending

Our preferred medical spending measure is the sum of expenditures paid out-of-pocket plus those paid by Medicaid. Let m_{it} denote the expenses incurred between ages t and $t + 2$. We observe the household's health at the beginning and the end of this interval,

[7] These figures depend on the distribution of PI and initial health across men and women. We construct these distributions with bootstrap draws from the second wave of the AHEAD, using households whose heads were between 70 and 72 in the first wave.

that is, at the time of the interview conducted at age t and at the time of the interview conducted at age $t+2$. Accordingly, we assume that medical expenses depend upon a household's PI, its family structure at both t and $t+2$, the health status of its members at both dates, and the idiosyncratic component ψ_t:

$$\begin{aligned}\ln m_{it} = {} & m(I_i, t+2, hs^h_{i,t}, hs^w_{i,t}, hs^h_{i,t+2}, hs^w_{i,t+2}, f_{i,t}, f_{i,t+2}) \\ & + \sigma(I_i, t+2, hs^h_{i,t}, hs^w_{i,t}, hs^h_{i,t+2}, hs^w_{i,t+2}, f_{i,t}, f_{i,t+2}) \times \psi_{i,t+2}.\end{aligned} \tag{3}$$

Including both family structure indicators allows us to account for the jump in medical spending that occurs in the period a family member dies. Likewise, including health indicators from both periods allows us to distinguish persistent health episodes from transitory ones. Finally, we include cohort dummies in the regression.[8]

We estimate equation (3) in two steps. In the first step we estimate $m(\cdot)$ using a fixed effects estimator. Fixed effects regression can only identify the role of time-varying factors, such as age, household structure, and health. Factors that are not time-varying, such as PI and the cohort dummies, are not identified separately from the estimated fixed effects. To address this problem, we first regress log medical spending on all the time-varying factors using a fixed effects estimator. We then take the residuals from this regression, inclusive of the fixed effects, and regress them on the time-invariant factors (a quadratic in permanent income and cohort dummies). In the simulations we make predictions using the dummy coefficient for the cohort aged 72-76 in 1996. Thus the level of $m(\cdot)$ is set to be consistent with the outcomes of this youngest cohort.

A key feature of our spending model is that the variance as well as the mean of medical spending depends on demographic and socioeconomic factors. To find $\widehat{\sigma^2(\cdot)}$, in the second step of estimating (3) we square the residuals (the ψs) from the first step and regress them on demographic and socioeconomic variables.

An accurate estimate of the lifetime medical expenditure distribution requires an accurate model of the intertemporal correlation of the idiosyncratic shock $\psi_{i,t}$. Following Feenberg and Skinner (1994) and French and Jones (2004), we assume that $\psi_{i,t}$ can be

[8]In particular we regress log medical spending on a fourth order age polynomial, indicators for single man (interacted with an age quadratic), single woman (interacted with an age polynomial), both the contemporaneous and lagged values of indicators for {man in bad health, man in a nursing home, woman in bad health, woman in a nursing home}, whether the man died (interacted with age and permanent income), whether the woman died (interacted with age, and permanent income), a quadratic in permanent income, and cohort dummies.

decomposed as

$$\psi_{i,t} = \zeta_{i,t} + \xi_{i,t}, \quad \xi_{i,t} \sim N(0, \sigma_\xi^2), \tag{4}$$

$$\zeta_{i,t} = \rho_m \zeta_{i,t-2} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_\epsilon^2), \tag{5}$$

where $\xi_{i,t}$ and $\epsilon_{i,t}$ are serially and mutually independent. We normalize the variance of $\psi_{i,t}$ to one, so that σ_ξ^2 can be interpreted as the fraction of idiosyncratic variance due to transitory shocks. We estimate the parameters of equations (4) and (5) using a standard error components method. Although the estimation procedure makes no assumptions on the the distribution of the error terms $\psi_{i,t}$, we assume normality in the simulations. French and Jones (2004) shows that if the data are are carefully constructed, normality captures well the far right tail of the medical spending distribution.[9]

Approximately 40% of the cross-sectional variation in log medical spending is explained by the observables, which are quite persistent. Of the remaining cross-sectional variation, 40% comes from the persistent shock ζ and 60% from the transitory shock ξ. In keeping with the results in Feenberg and Skinner (1994), French and Jones (2004) and De Nardi et al. (2010), we estimate substantial persistence in the persistent component, with $\rho_m = 0.85$.

3.3 Quantitative Approach

After estimating our model, we assess its implications through a series of Monte Carlo exercises. The simulations begin at age 72 (reflecting medical spending between 70 and 72) and end at age 102.[10] Each simulated household receives a bootstrap draw of PI, initial health, and initial marital status from the HRS data used to estimate the health and spending models.[11] The household also receives initial values of ζ and ξ drawn from their unconditional distributions. We then use our Markov Chain model of health and mortality (equation (2)) to simulate demographic histories for each household, and

[9] To help us match the distribution of medical spending, we bottom code medical spending at 10% of average medical spending. French and Jones (2004) also bottom code the data to match the far right tail of medical spending. Because we include Medicare B payments in our medical spending measure, which most elderly households pay, for the vast majority of households these bottom coding decisions are not important.

[10] In couples, wives are assumed to be 3 years younger than husbands – the data average – and are thus initially 69. Single women are assumed to be 72.

[11] Although the simulations begin at age 72, we take bootstrap draws from the set of people aged 72 to 74 in 1996. This gives us a larger pool of households to draw from, which should in turn improve the accuracy of our exercise. We set the initial values of lagged (age-70) health and marital status equal to their age-72 values, allowing us to calculate medical expenditures made between the ages of 70 and 72.

give each household a sequence of idiosyncratic shocks consistent with equations (4) and (5). Combining these inputs through equation (3) yields medical spending histories. We generate 1 million such histories and calculate summary statistics at each age.

4 Results

4.1 Unconditional Spending Distributions

Figure 1 shows our model's implications for the cross sectional distribution of our preferred medical spending measure, the sum of costs paid either out-of-pocket or by Medicaid.

Figure 1a, in the upper left corner, summarizes the health care expenditures of surviving households. Mean and median expenditures are shown, along with the 90th, 95th and 99th percentiles. The results are dated by the beginning of the spending interval: the numbers for age 72 describe the medical expenses incurred between ages 72 and 74 by people alive at both dates. Expenditures are expressed in annual terms. The medical expenses of surviving households rise rapidly with age. For example, mean medical spending rises from $5,100 per year at age 70 to $29,700 at age 100. The upper tail rises even more rapidly, with the 95th percentile increasing from $13,400 to $111,200.[12]

Figure 1b shows end-of-life costs, which include burial expenses; the results for age 72 describe the expenses incurred by households who die between ages 72 and 74.[13] On average, end-of-life medical expenses exceed those of survivors. Mean end-of-life expenses range from $11,000 at age 72 to $34,000 at age 100.

Figure 1c plots our main variable of interest, lifetime expenditures. At each age, we calculate the present discounted value of remaining medical expenditures from that age forward, using an annual real discount rate of 3 percent. These lifetime totals are considerable. At age 70, households will on average incur over $122,000 of medical expenditures over the remainder of their lives. The top 5 and 1 percent of spenders will incur spending in excess of $330,000 and $640,000, respectively. One might expect the lifetime totals to fall rapidly as households age and near the ends of their lives. This is not the case. A household alive at age 90 will on average spend more than $113,000 before they die. The

[12] In general, our estimated model matches well the distribution of medical spending found in the raw data. However, the model overstates the 99th percentile of the medical spending distribution after age 90. Given the low probability of having medical spending in the 99th percentile, along with the low probability of living much past age 90, this discrepancy should have only a modest impact on our estimated lifetime spending distribution.

[13] We define a dead household as one that has no members. Couples who become singles are classified as survivors.

95th percentile of remaining lifetime spending is higher at age 90 than at age 70. The slow decline of lifetime costs is due mostly to the tendency of medical costs to rise with age. Households that live to older ages have shorter remaining lives but higher annual expenditure rates.

A number of papers have considered whether medical expenses rise with age generically or mostly because older people are more likely to incur end-of-life expenses: see the discussion in De Nardi et al. (2016c). In our spending model both forces are present. The top row of Figure 1 shows that there is considerable age growth in the medical expenses of both survivors and the newly deceased. Nonetheless, except for the 99th percentile at the oldest ages, the end-of-life expenses shown in panel 1b are larger than the expenses faced by survivors of the same age (panel 1a).

Figure 1d presents the annuitized spending associated with these lifetime totals. For each household, we convert lifetime expenses into the constant spending flow that would, over that household's realized lifespan, have the same present value. The average annuity payment rises from \$9,100 at age 70 to \$31,800 at age 100. Comparing panels 1a and 1d shows that mean annuitized spending at age 70 is almost double mean current spending. This reflects the rapid growth in medical spending that occurs as households age. The 95th and 99th percentiles of the annuitized medical spending distribution are also higher than the corresponding percentiles of the current spending distribution. One might think that those who have high medical spending in the present will usually have lower medical spending in the future, leading annuitized spending, which is essentially an average, to be less dispersed than current spending. The wide variation in annuitized spending found in panel 1d thus shows that medical spending is persistent, and that those with high spending in the present are likely to have high spending in the future.

4.2 Lifetime Medical Spending Determinants and Medical Spending Risk

The graphs presented in Figure 1 show that the medical costs of older households are high, rising with age and widely dispersed. A significant portion of this variation, however, is due to factors that are known to the household (PI, health, marital status, the persistent shock ζ). The spending distributions that individual households actually face, conditional on what they know at any point in time, can be quite different.

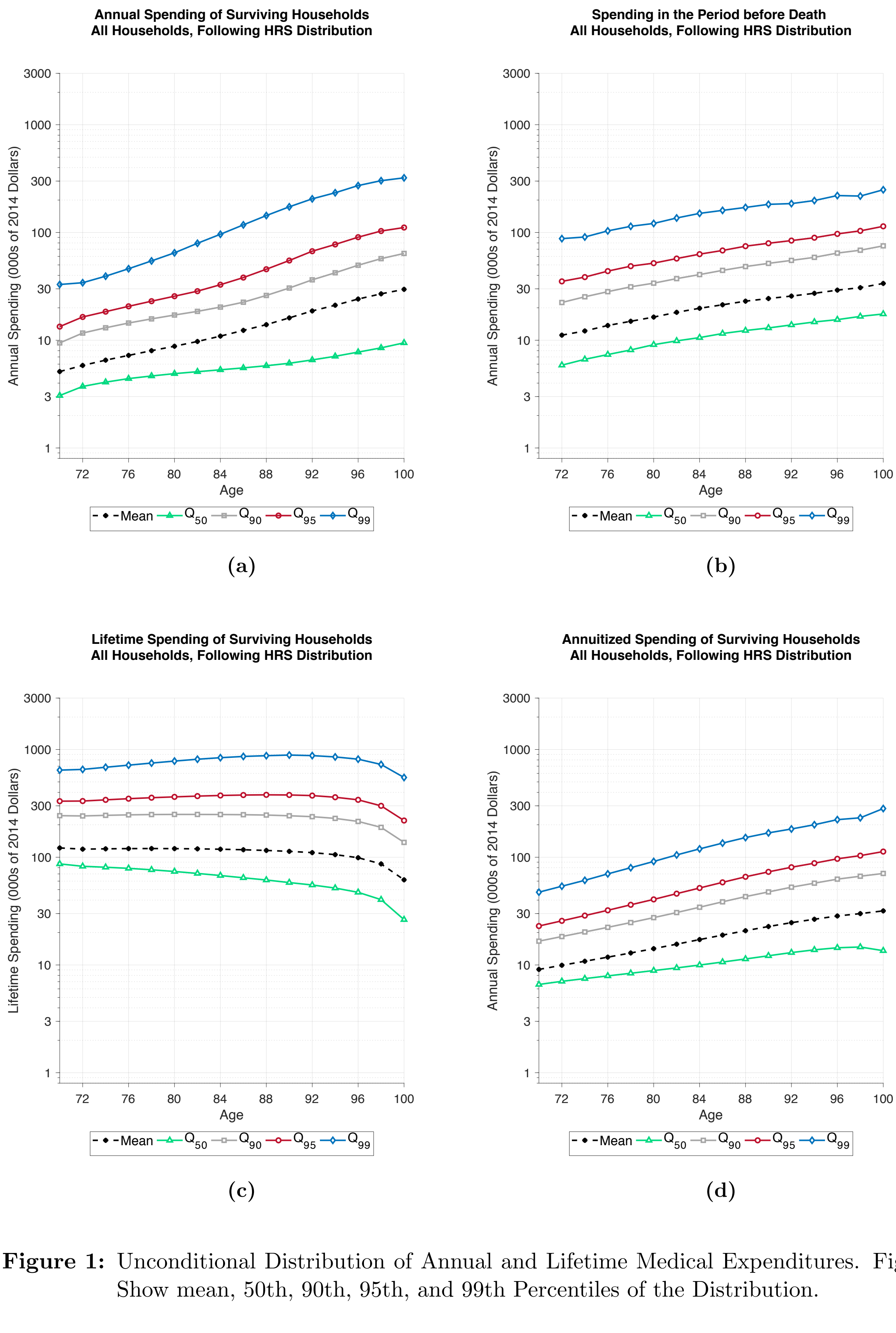
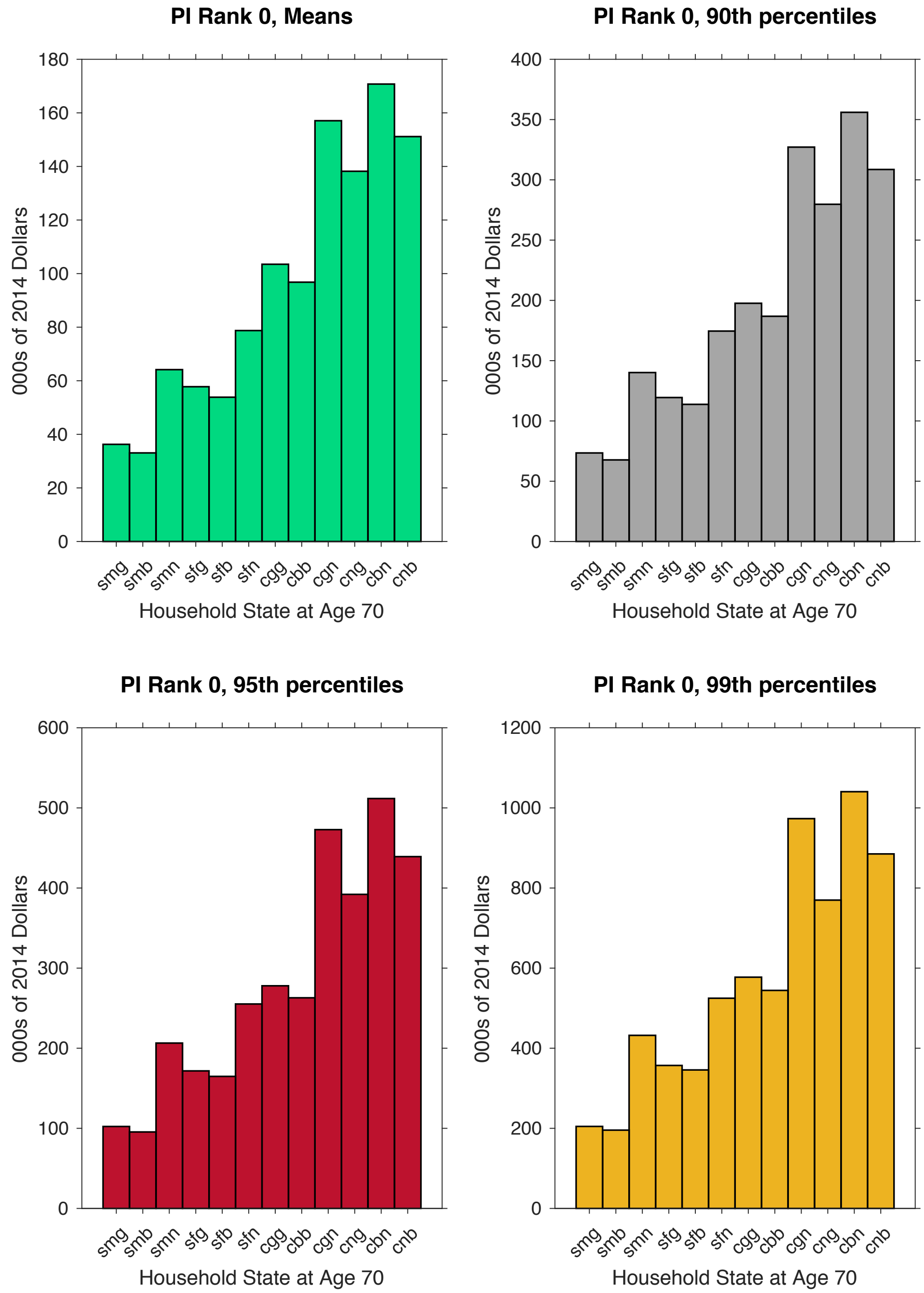
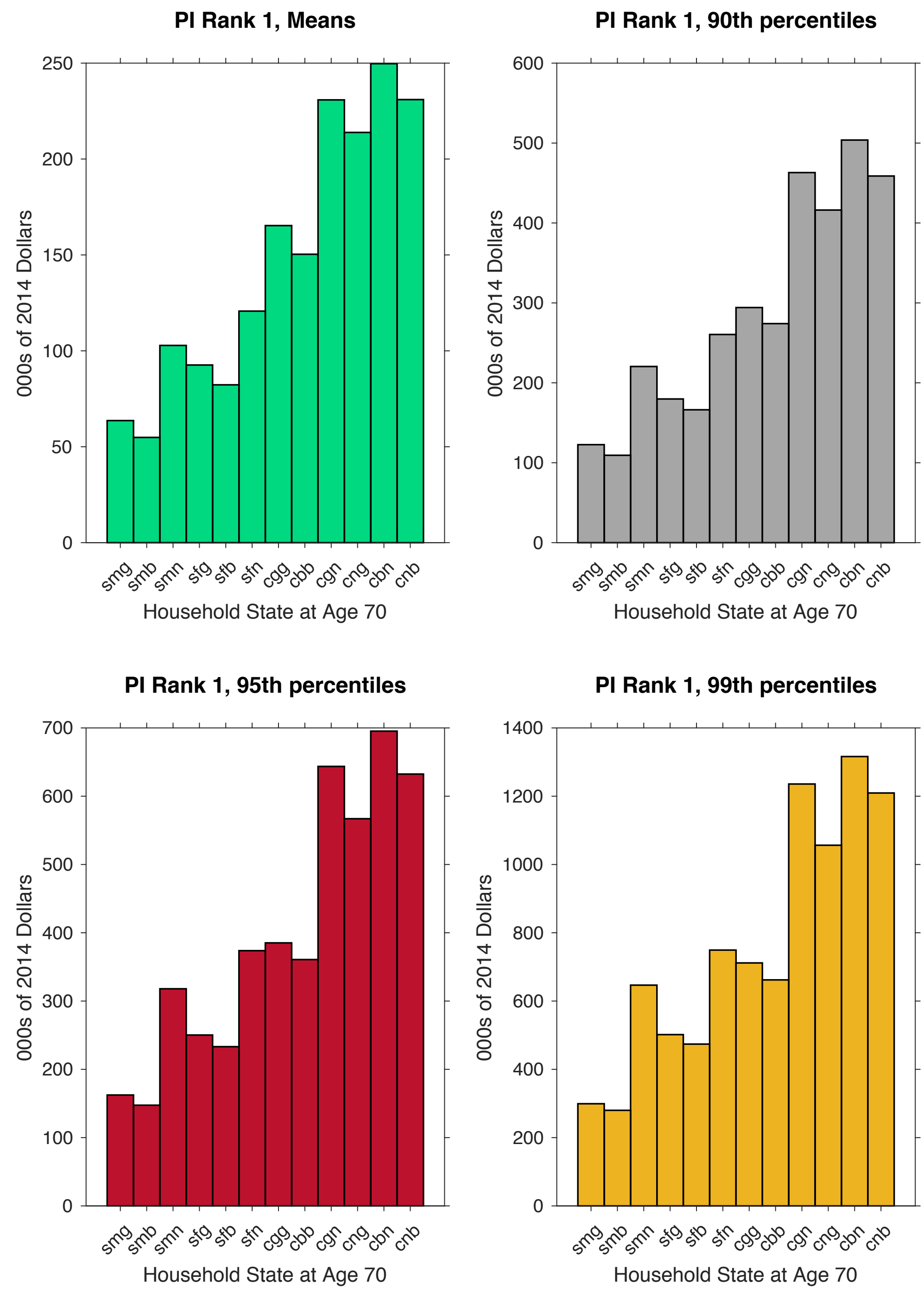


Figure 1: Unconditional Distribution of Annual and Lifetime Medical Expenditures. Figures Show mean, 50th, 90th, 95th, and 99th Percentiles of the Distribution.



smg:	single male, good health	smb:	single male, bad health	smn:	single male, nursing home
sfg:	single female, good health	sfb:	single female, bad health	sfn:	single female, nursing home
cgg:	couple, both good health	cbb:	couple, both bad health	cgn:	male in good health, female in nursing home
cng:	male in nursing home, female in good health	cbn:	male in bad health, female in nursing home	cnb:	male in nursing home, female in bad health

Figure 2: Distributions of Lifetime Medical Expenditures by Initial Health and Household Structure, PI = 0



smg:	single male, good health	smb:	single male, bad health	smn:	single male, nursing home
sfg:	single female, good health	sfb:	single female, bad health	sfn:	single female, nursing home
cgg:	couple, both good health	cbb:	couple, both bad health	cgn:	male in good health, female in nursing home
cng:	male in nursing home, female in good health	cbn:	male in bad health, female in nursing home	cnb:	male in nursing home, female in bad health

Figure 3: Distributions of Lifetime Medical Expenditures by Initial Health and Household Structure, PI = 1

Figures 2 and 3 compare the mean, 90th, 95th and 99th percentiles of lifetime medical spending at age 70 for different values of PI and initial health and marital status. Figure 2 shows results for households at the very bottom of the income distribution (PI = 0). Lifetime spending varies greatly across the distribution of initial health and marital status. Some trends are apparent:

1. Women have higher lifetime medical expenditures than men.
2. People initially in good health have higher lifetime expenditures than those initially in bad health. This reflects their longer life expectancies, combined with the tendency of medical costs to rise with age.
3. Households initially in nursing homes have the highest lifetime expenditures, in spite of their high mortality. Nursing home care is expensive, and most people – more than 70 percent of men and 60 percent of women – outside a nursing home at age 70 never have an extended nursing home visit.

Figure 3 shows results for households at the very top of the income distribution (PI = 1). Households at the top of the income distribution spend considerably more than those at the bottom, often well in excess of 50 percent more. By way of example, consider 70-year-old couples where both members are initially in good health. With a PI rank of 0 these couples would on average spend $104,000 over their remaining lives. With a PI rank of 1 they would spend over $165,000. Households with higher income may have higher lifetime expenditures because they live longer, or because they have higher expenses at any given age. Figure 4, which compares the same two groups in more detail, shows that both effects are present. The top two panels of this figure compare annual expenditures for surviving individuals. While high income households typically spend more each year, at earlier ages and higher percentiles the opposite is often true.

Figures 2-3 show that a significant part of the dispersion in retiree medical spending can be attributed to health and demographic factors known at the very beginning of retirement. On the other hand, Figure 4 shows that spending remains dispersed even after conditioning on these factors. For example, the gaps between the conditional means and 99th percentiles of lifetime spending shown in Figures 4c and 4d are of roughly the same size as the unconditional gap shown in Figure 1c.

Another potentially predictable source of spending variation is the persistent idiosyncratic component of medical spending, ζ. The importance of the initial idiosyncratic shocks can be seen in Figure 5. The two panels in the left-hand column of this figure are

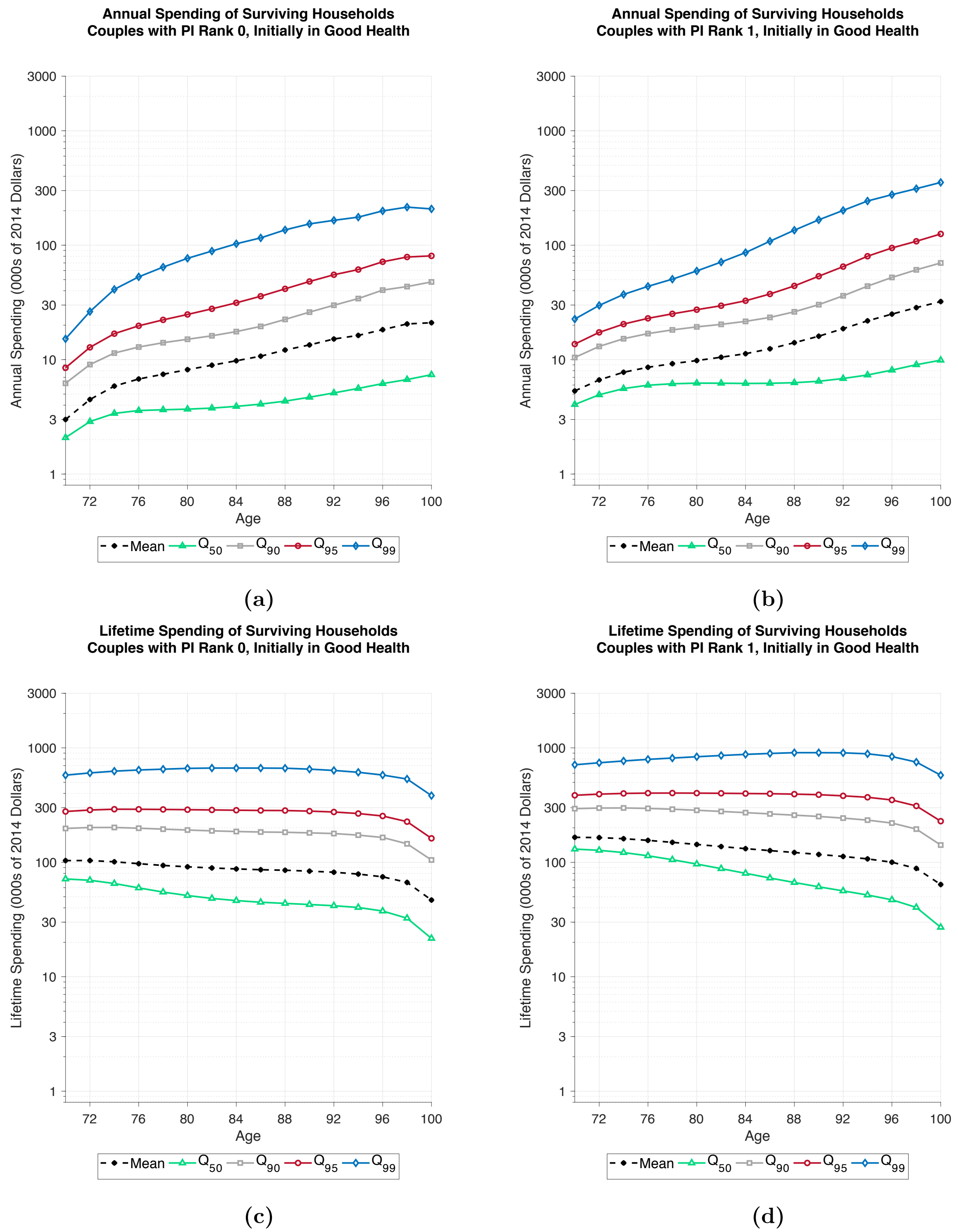


Figure 4: Annual and Lifetime Medical Expenses of Couples in Initial Good Health, with PI Ranks of 0 (left column) and 1 (right column)

directly comparable to the corresponding columns in Figure 4; the only difference is that the results in the new graphs are generated using a permanent income rank of 0.5. The two panels in the right-hand column differ from those on the left in that all the simulated histories begin with $\zeta = \xi = 0$. This can be seen in Figure 5b, where the distribution of annual expenses is initially degenerate. Comparing the panels in the top row shows that the effects of the shocks last for several years. On the other hand, the bottom panels show that eliminating the initial spending shocks has a relatively small effect on the dispersion of lifetime expenditures. Knowing the initial idiosyncratic shocks removes little risk.

There are three reasons why the effects of the initial shocks wear off. First, as we document above, a significant portion of the variation in annual medical spending is driven by the health status of the household. Second, the transitory component ξ accounts for a large fraction of the residual variation, and imposing an initial condition has no effect on future transitory shocks. Finally, the effect of the initial realization of the persistent component ζ declines with age, as the persistence parameter ρ_m is less than 1.

The previous results notwithstanding, a large part of the elderly's medical spending uncertainty is due to the idiosyncratic shocks. Recall that only about 40% of the cross-sectional variation in log medical spending is explained by the observables. Likewise, if we remove all of the idiosyncratic shocks in our simulations, so that the only uncertainty is health and household structure, the unconditional variation of lifetime medical spending is only a fraction of its original value.

4.3 Out-of-Pocket Medical Spending

Our baseline measure of medical spending is the sum of payments made out-of-pocket and Medicaid. A number of recent papers have argued that Medicaid significantly reduces the out-of-pocket spending risk faced by older households. Brown and Finkelstein (2008) conclude that Medicaid crowds out private long-term care insurance for about two-thirds of the wealth distribution. De Nardi et al. (2016a) find that most single retirees, including those at the top of the income distribution, value Medicaid at more than its actuarial cost. While both of these papers model Medicaid formally, as part of the budget set in a dynamic structural model, it is also useful to assess the program in a less structured way. In particular, repeating our Monte Carlo exercises with the HRS out-of-pocket measure, which excludes Medicaid, allows us to compare the pre- and post-Medicaid distribution of medical spending.

Figure 6 compares unconditional distributions. The two panels in the left-hand column of this figure show results for our baseline spending measure; the panels in the right-hand

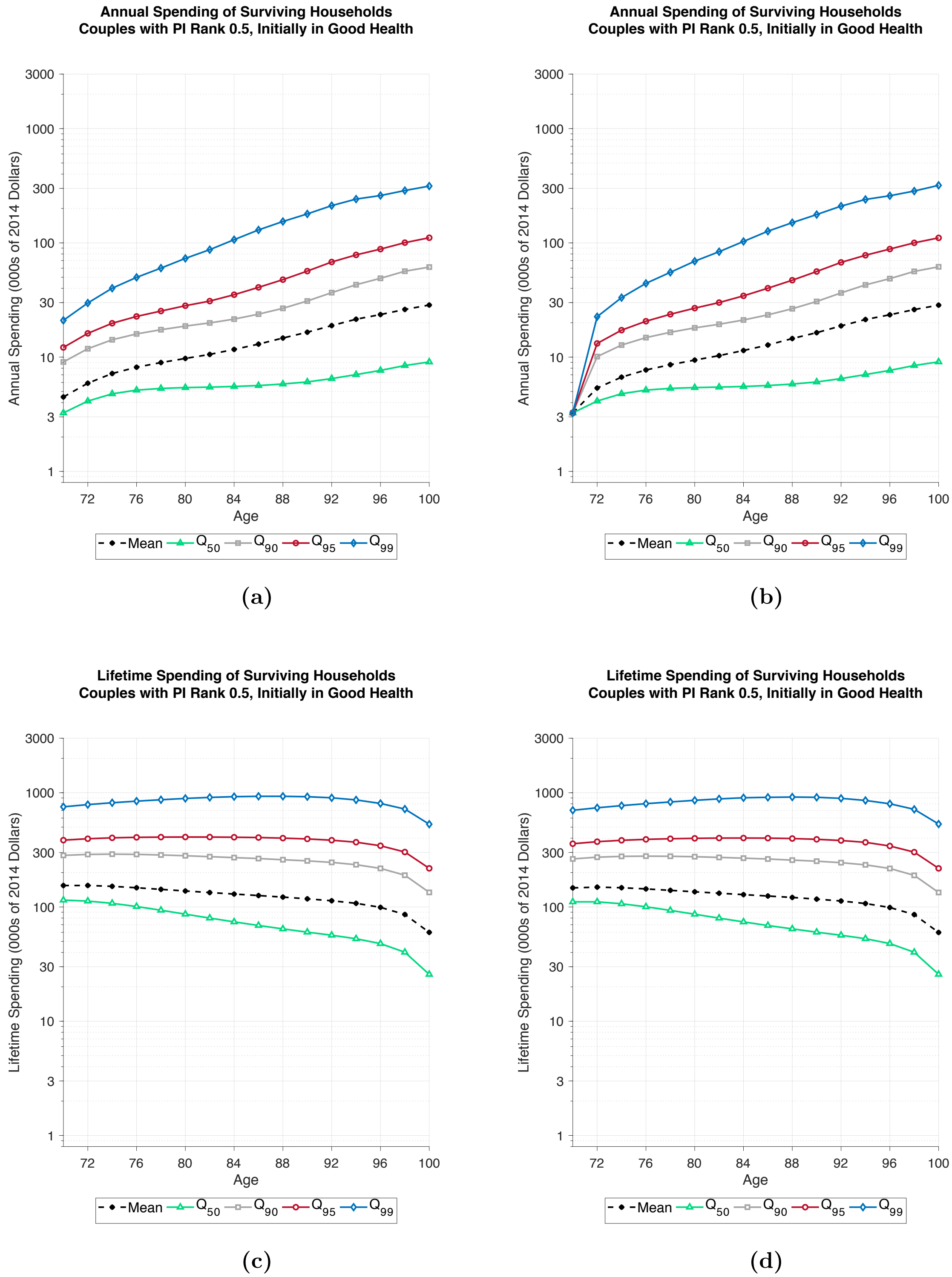


Figure 5: Annual and Lifetime Medical Expenses of Couples in Initial Good Health, with (left column) and without (right column) Initial Idiosyncratic Spending Shocks

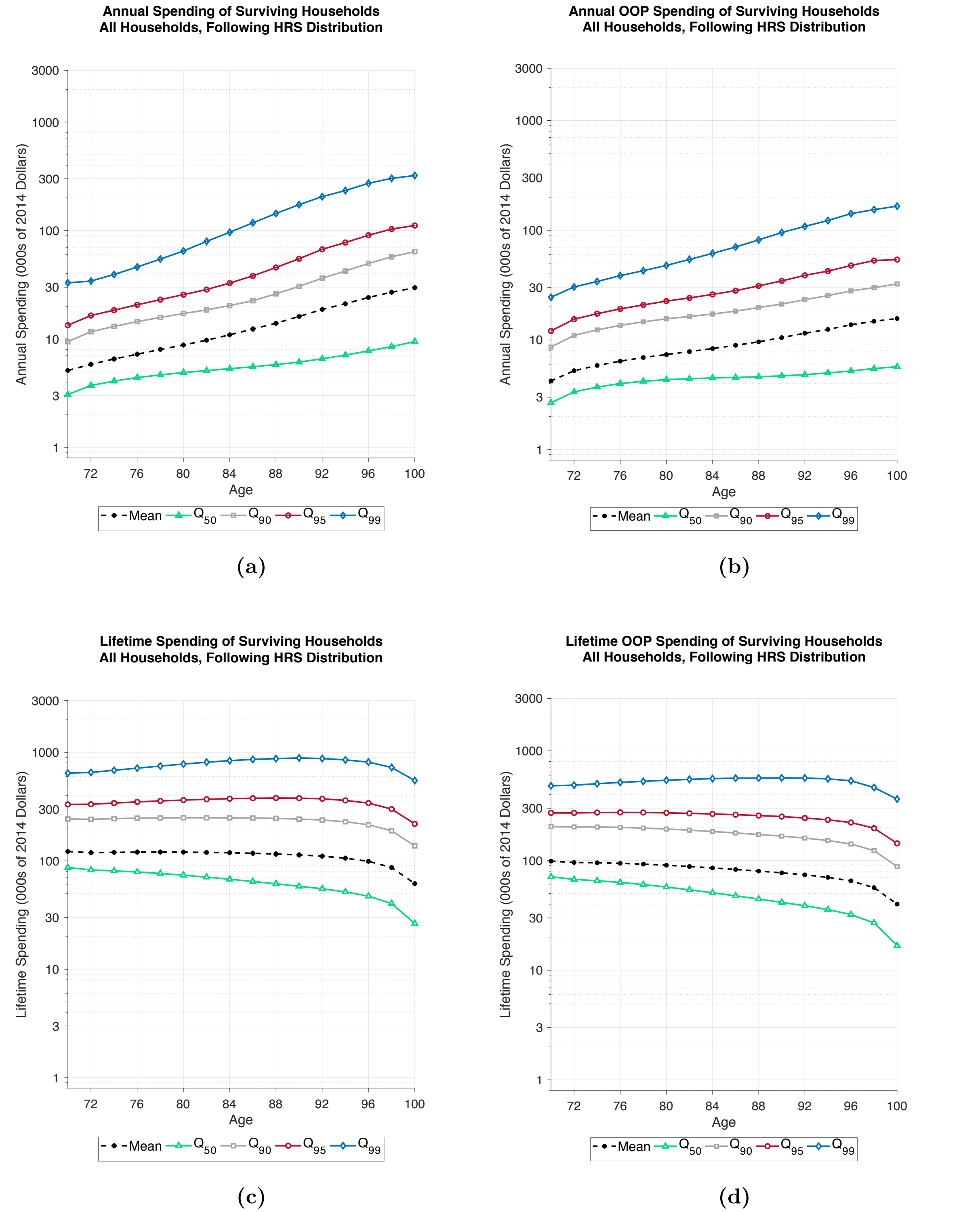
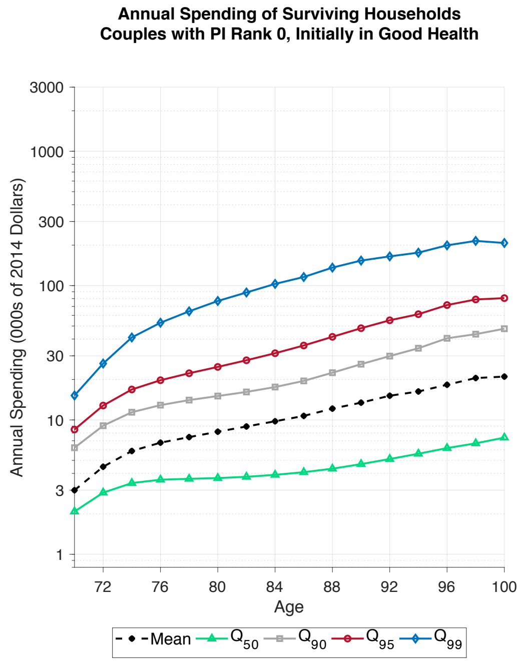


Figure 6: Unconditional Distribution of Annual and Lifetime Medical Expenditures, with (Left Panels) and without (Right Panels) Medicaid Payments

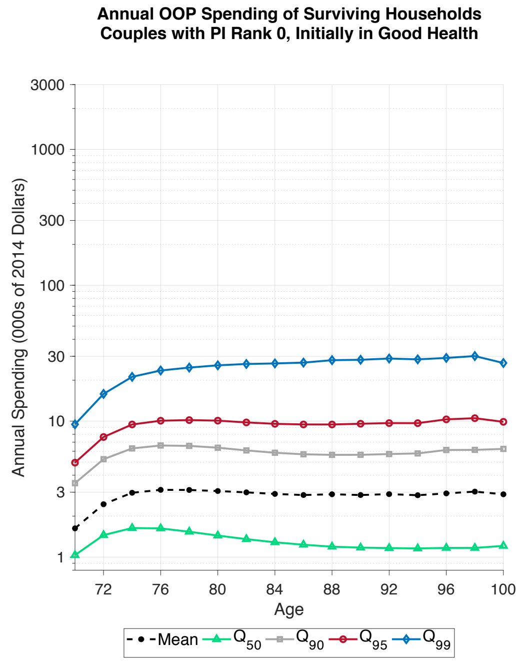
column show results for out-of-pocket spending alone. The first row of Figure 6 compares annual expenditures for survivors. At age 70 mean out-of-pocket expenditures ($4,200) are about 18 percent less than mean combined expenditures ($5,100). In other words, Medicaid covers about 18 percent of the total for 70 year olds. However, at older ages and higher percentiles, out-of-pocket expenditures are considerably lower. The second row of Figure 6 shows lifetime expenditures. At age 70 mean lifetime out-of-pocket expenses are about 20 percent lower than mean combined expenditures. This difference may seem small given the differences in the first row, but end-of-life expenditures (not shown) are fairly similar across the two spending measures.

Because Medicaid is means-tested, it is most prevalent at the bottom of income distribution. To show this more clearly, Figure 7 compares the annual spending of surviving households at different points of the PI distribution. Consistent with Figures 4 and 5, we look at couples where both spouses were initially in good health. The top row of Figure 7, which compares the two spending measures for households at the bottom of the PI distribution, shows that Medicaid picks up a large share of these households' medical expenditures. At age 70, mean out-of-pocket expenditures are about 45 percent lower than mean combined expenditures, meaning that Medicaid constitutes about 45 percent of the total. The share of costs covered by Medicaid rises rapidly with age, however, to around 85 percent. The bottom row of Figure 7 repeats the comparison for the top of the PI distribution. Not surprisingly, Medicaid covers a much smaller fractions of these households' expenditures.

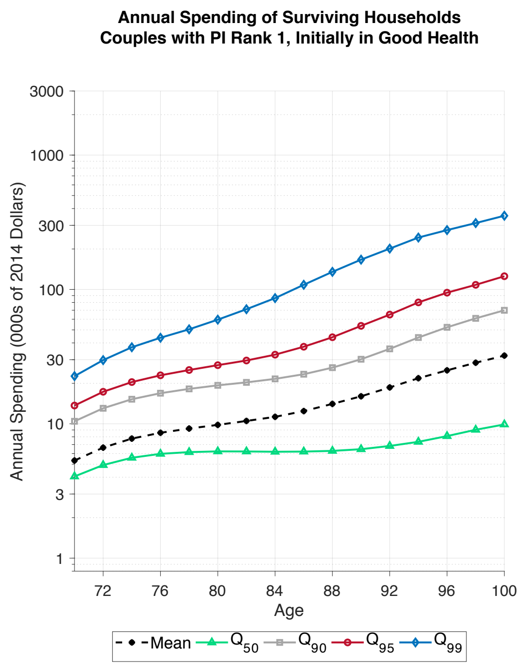
Figure 8 compares lifetime spending totals. The top row of this figure shows that at the bottom of the income distribution, Medicaid covers 57 percent of lifetime costs as of age 70. At older ages and higher percentiles it covers even more. The bottom row shows results for households at the top of the income distribution. Medicaid covers 21 percent of lifetime costs at age 70, with the fraction rising to nearly 30 percent at age 100. While most high-income households do not receive Medicaid, those that do qualify under the Medically Needy provision, which assists households whose financial resources have been exhausted by medical expenses. Such households tend to have high medical expenses and tend to receive large Medicaid benefits (De Nardi et al., 2016a).



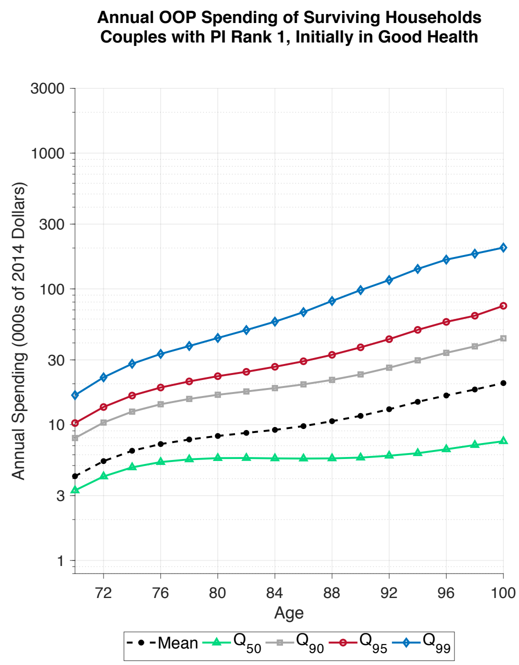
(a)



(b)

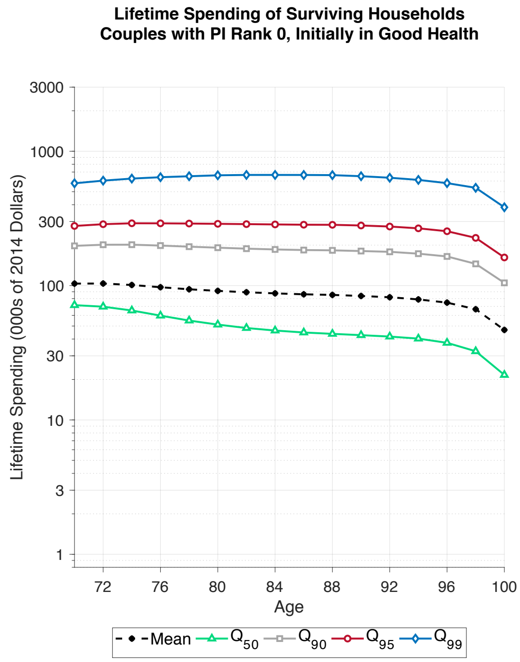


(c)

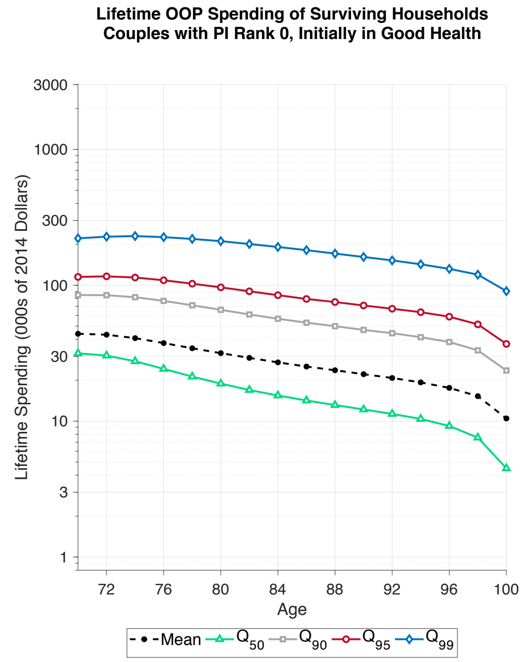


(d)

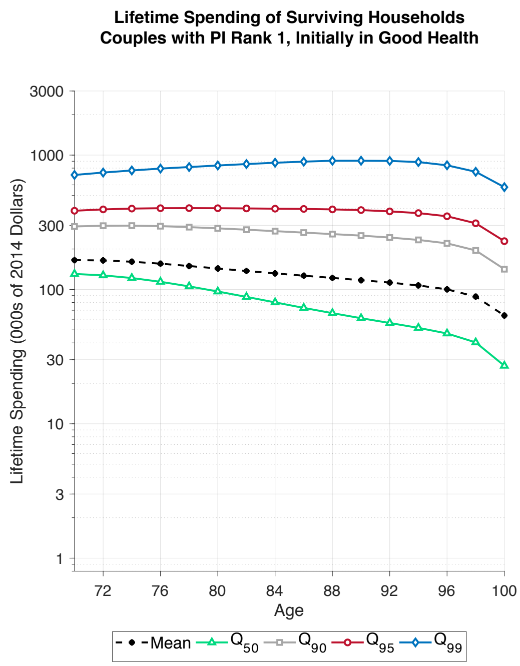
Figure 7: Annual Medical Expenses of Couples in Initial Good Health, with (Left Panels) and without (Right Panels) Medicaid Payments



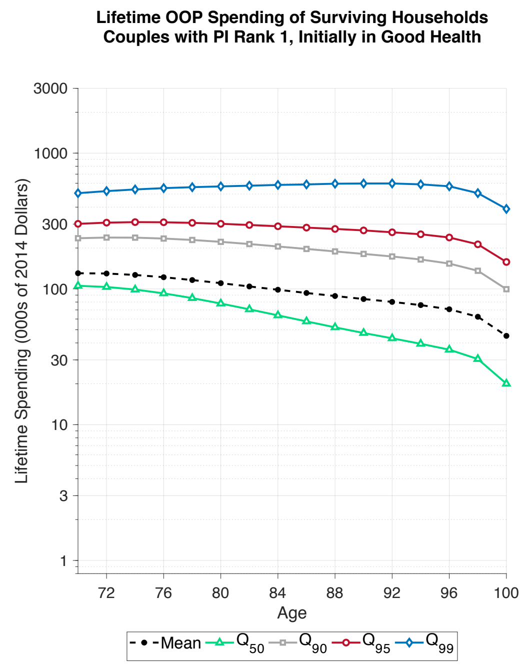
(a)



(b)



(c)



(d)

Figure 8: Lifetime Medical Expenses of Couples in Initial Good Health, with and without Medicaid Payments

5 Discussion and Conclusions

In this paper we use the health and spending models developed in De Nardi et al. (2018) to simulate the distribution of lifetime medical expenditures as of age 70, adding to the handful of studies on this topic. We also assess the importance of Medicaid in reducing lifetime medical spending risk. The simulations show that lifetime medical spending is high and uncertain, and that the level and the dispersion of this spending diminish only slowly with age. Although PI, initial health and initial marital status have large and predictable effects, much of the dispersion in lifetime spending is due to events realized at older ages. The poorest households have the majority of their medical costs covered by Medicaid, which significantly reduces their risk as well. Medicaid also reduces the level and volatility of medical spending for high-income households, albeit to a much smaller degree.

The paper closest to ours is Webb and Zhivan (2010), which we discussed in our introduction. Webb and Zhivan (2010) also find that lifetime out-of-pocket medical spending is high and widely dispersed, and that the level and conditional dispersion of this spending diminish only slowly as households age. The levels of their estimated expenses, however, are even higher than ours, even though their spending measure excludes Medicaid. For instance, they find that 70-year-old couples with high school degrees and no chronic diseases will on average spend about \$300,000 over their remaining lives, and 5 percent of these households will spend well over \$600,000.[14] We find that 70-year-old couples with a PI rank of 0.5 and good initial health will on average spend about \$150,000 over their remaining lives and that 5 percent of these households will spend in excess of \$380,000.

One likely reason why Webb and Zhivan (2010) find higher medical spending is that they estimate their model using a pooled cross-section regression. They then correct for cohort bias – the fact that, for instance, the medical spending of a 90-year-old observed in 1996 is likely to be lower than the medical spending a 70-year-old observed in 1996 would face when she turned 90 in 2016 – by allowing their simulated medical expenses to grow over time, independent of age, at historical rates. In contrast, we estimate our spending model using a fixed effects regression with no time controls, so that our age effects measure the year-to-year spending growth that households realized over the sample period. Because medical spending growth has been fairly slow in recent years – and out-of-pocket spending was reduced by the introduction of Medicare Part D in 2006 – Webb and Zhivan's (2010) assumed spending growth likely exceeds recent experience.

[14]We inflate their results from 2009 to 2014 dollars – roughly 10 percent – using the CPI.

We conclude by pointing out some caveats to our analysis. We assume, as do most other empirical papers, that medical spending is exogenous, while in reality it is a choice variable. Although the demand for some medical goods and services is extremely inelastic, the demand for others might be quite elastic. Nursing home care, for example, is a bundle of medical and non-medical commodities, and the latter can vary greatly in quality, with the choice between a single and a shared room being just one example. It is also worth noting that our analysis excludes payments made by Medicare and private insurers. Medicare substantially reduces out-of-pocket medical expenses throughout the retiree population (Barcellos and Jacobson, 2015). While the combination of out-of-pocket and Medicaid expenditures considered here may be sufficient for some analyses, such as studies of household saving, other analyses require that all health costs be accounted for. Extending our exercise to include all medical expenditures would be useful, and we leave it to future research.

References

Alemayehu, Berhanu and Kenneth E. Warner, "The Lifetime Distribution of Health Care Costs," *Health Services Research*, 2004, *39* (3), 627–642.

Ameriks, John, Joseph S Briggs, Andrew Caplin, Matthew D Shapiro, and Christopher Tonetti, "Long-Term-Care Utility and Late-in-Life Saving," Technical Report, National Bureau of Economic Research 2015.

Banks, James, Richard Blundell, Peter Levell, and James Smith, "Life-cycle consumption patterns at older ages in the US and the UK: can medical expenditures explain the difference?," Sep 2016. Mimeo.

Barcellos, Silvia Helena and Mireille Jacobson, "The Effects of Medicare on Medical Expenditure Risk and Financial Strain," *American Economic Journal: Economic Policy*, November 2015, *7* (4), 41–70.

Boerma, Job and Ellen McGrattan, "Health Capital Taxation," 2018. Mimeo.

Brown, Jeff and Amy Finkelstein, "The Interaction of Public and Private Insurance: Medicaid and the Long Term Care Insurance Market," *American Economic Review*, 2008, *98* (5), 837–880.

Carreras, Marc, Pere Ibern, Jordi Coderch, Inma Sánchez, and Jose M. Inoriza, "Estimating Lifetime Healthcare Costs with Morbidity Data," *BMC Health Services Research*, 2013, *13* (1), 440.

Conesa, Juan Carlos, Daniela Costa, Parisa Kamali, Timothy J Kehoe, Vegard M Nygard, Gajendran Raveendranathan, and Akshar Saxena, "Macroeconomic effects of Medicare," *The Journal of the Economics of Ageing*, 2017.

De Nardi, Mariacristina, "Wealth Inequality and Intergenerational Links," *Review of Economic Studies*, 2004, *71* (3), 743–768.

_ , Eric French, and John Bailey Jones, "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 2010, *118* (1), 39–75.

_ , _ , and _ , "Medicaid Insurance in Old Age," *American Economic Review*, 2016, *106* (11), 3480–3520.

_ , _ , and _ , "Savings After Retirement: A Survey," *Annual Review of Economics*, 2016, *8* (1), 177–204.

_ , _ , _ , and Jeremy McCauley, "Medical Spending on the US Elderly," *Fiscal Studies*, 2016, *37* (3-4), 717–747.

_ , _ , _ , and Rory McGee, "Couples' and Singles Savings After Retirement," 2018. Work in progress.

Fahle, Sean, Kathleen McGarry, and Jonathan Skinner, "Out-of-Pocket Medical Expenditures in the United States: Evidence from the Health and Retirement Study," *Fiscal Studies*, 2016, *37* (3-4), 785–819.

Feenberg, Daniel and Jonathan Skinner, "The Risk and Duration of Catastrophic Health Care Expenditures," *Review of Economics and Statistics*, 1994, *76*, 633–647.

French, Eric and John Bailey Jones, "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Econometrics*, 2004, *19* (4), 705–721.

_ , _ , and Jeremy McCauley, "The Accuracy of Economic Measurement in the Health and Retirement Study," *Forum for Health Economics and Policy*, 2017, *20* (2).

Friedberg, Leora, Wenliang Hou, Wei Sun, Anthony Webb, and Zhenyu Li, "New Evidence on the Risk of Requiring Long-term Care," November 2014. Center for Retirement Research at Boston College Working Paper 2014-12.

Hirth, Richard A., Teresa B. Gibson, Helen Levy, Jeffrey Smith, Sebastian Calnico, and Anup Das, "New Evidence on the Persistence of Health Spending," *Medical Care Research and Review*, 2015, *72* (3), 277–297.

Hugonnier, Julien, Florian Pelgrin, and Pascal St-Amour, "Health and (other) asset holdings," *Review of Economic Studies*, 2012, *80* (2), 663–710.

Hurd, Michael D., Pierre-Carl Michaud, and Susann Rohwedder, "Distribution of Lifetime Nursing Home use and of Out-of-Pocket Spending," *Proceedings of the National Academy of Sciences*, 2017, *114* (37), 277–297.

Jung, Juergen and Chung Tran, "Market Inefficiency, Insurance Mandate and Welfare: U.S. Health Care Reform 2010," *Review of Economic Dynamics*, 2016, *20*, 132–159.

Kopecky, Karen and Tatyana Koreshkova, "The Impact of Medical and Nursing Home Expenses and Social Insurance Policies on Savings and Inequality," *American Economic Journal: Macroeconomics*, July 2014, *6* (3), 29–72.

Livshits, Igor, James MacGee, and Michele Tertilt, "Consumer bankruptcy: A fresh start," *American Economic Review*, 2007, *97* (1), 402–418.

Lockwood, Lee, "Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks," 2012. Mimeo.

Love, David A, "The effects of marital status and children on savings and portfolio choice," *The Review of Financial Studies*, 2009, *23* (1), 385–432.

Merrill Lynch Wealth Management, "Merrill Lynch Affluent Insights Survey Fact Sheet," February 2012.

Nakajima, Makoto and Irina A. Telyukova, "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden," Apr 2018. Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute.

_ **and Irina Telyukova**, "Home Equity in Retirement," 2012. Mimeo.

Pashchenko, Svetlana, "Accounting for non-annuitization," *Journal of Public Economics*, 2013, *98*, 53–67.

— **and Ponpoje Porapakkarm**, "Quantitative analysis of health insurance reform: Separating regulation from redistribution," *Review of Economic Dynamics*, 2013, *16* (3), 383–404.

Poterba, James M., Steven F. Venti, and David A. Wise, "The Asset Cost of Poor Health," *Journal of the Economics of Ageing*, 2017, *9*, 172–184.

Reichling, Felix and Kent Smetters, "Optimal Annuitization with Stochastic Mortality Probabilities," *American Economic Review*, 2015, *105* (11), 3273–3320.

Rohwedder, Susann, Steven J. Haider, and Michael Hurd, "Increases in Wealth among the Elderly in the Early 1990s: How Much is Due to Survey Design?," *Review of Income and Wealth*, 2006, *52* (4), 509–524.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun, "Are Americans Saving Optimally for Retirement?," *Journal of Political Economy*, 2006, *114*, 607–643.

Skinner, Jonathan, "Are you sure you're saving enough for retirement?," *Journal of Economic Perspectives*, 2007, *21* (3), 59–80.

Webb, Anthony and Natalia A. Zhivan, "How Much is Enough?: The Distribution of Lifetime Health Care Costs," February 2010. Center for Retirement Research at Boston College Working Paper 2010-1.

Appendix A: Imputing Medicaid Expenditures

Let i index individuals in the HRS. Define oop_{it} as out-of-pocket medical expenses, Med_{it} as Medicaid payments, and m_{it} as the sum of out-of-pocket and Medicaid payments that we wish to plug in the model. To impute Med_{it}, which is missing in the HRS, we follow David et al. (1986) and French and Jones (2011) and use a predictive mean-matching regression approach. There are two steps to our procedure. First, we use the MCBS data to regress Medicaid payments (for Medicaid recipients) on observable variables that exist in both datasets. Second, we impute Medicaid payments in the HRS data using a conditional mean-matching procedure, a procedure very similar to hotdecking.

First Step Estimation Procedure

Let j index individuals in the MCBS. For the subsample of the MCBS with a positive Medicaid indicator (i.e., a Medicaid recipient), we regress the variable of interest, Med_{jt}, on the vector of observable variables z_{jt}, yielding $Med_{jt} = z_{jt}\beta + \varepsilon_{jt}$. We include in z_{jt} nursing home status, the number of nights spent in a nursing home, a fourth-order age polynomial, total household income, marital status, self-reported health, race, visiting a medical practitioner (doctor, hospital or dentist), out-of-pocket medical spending, education and death of an individual. Because the measure of medical spending in the HRS is medical spending over two years, we take two-year averages of the MCBS data to be consistent with the structure of the HRS. The regression of Med_{jt} on z_{jt} yields a R^2 statistic of 0.67, suggesting that our predictions are accurate.

Next, for every observation in the MCBS subsample we calculate the predicted value $\widehat{Med}_{jt} = z_{jt}\widehat{\beta}$ and the residual $\hat{\varepsilon}_{jt} = Med_{jt} - \widehat{Med}_{jt}$. We then sort the observations into deciles by predicted values, $\{\widehat{Med}_{jt}\}_{j,t}$, and keep track of all values of $\hat{\varepsilon}_{jt}$ within each decile.

Second Step Estimation Procedure

For every observation in the HRS sample with a positive Medicaid indicator, we impute $\widehat{Med}_{it} = z_{it}\widehat{\beta}$, using the values of $\widehat{\beta}$ estimated from the MCBS. Then we impute ε_{it} for each observation of this subsample by finding a random observation in the MCBS with a value of $\widehat{Med}_{jt}$ in the same decile as $\widehat{Med}_{it}$, and setting $\hat{\varepsilon}_{it} = \hat{\varepsilon}_{jt}$. The imputed value of Med_{it} is $\widehat{Med}_{it} + \hat{\varepsilon}_{it}$.

As David et al. (1986) point out, our imputation approach is equivalent to hot-decking when the "z" variables are discretized and include a full set of interactions. The advantages of our approach over hot-decking are two-fold. First, many of the "z" variables are continuous. Second, to improve goodness of fit we use a large number of "z" variables. Because hot-decking uses a full set of interactions, this would result in a large number of hot-decking cells relative to our sample size.