SI APPENDIX

The Distribution of Lifetime Nursing Home Use and of Out-of-Pocket Spending

Michael D. Hurd, Pierre-Carl Michaud, Susann Rohwedder

Contents	**Page**

1. **Data and non-parametric method**

The Health and Retirement Study (HRS) was a community-based probability sample at its initial wave in 1992. Additional birth cohorts were added in 1993, 1998, 2004 and 2010, using the same sampling approach. The HRS has surveyed biennially about 20,000 persons, including those who moved to a nursing home in subsequent waves. In the non-parametric method we use longitudinal observations on four multiyear birth cohorts in this paper: (1) Those born in the years 1919-1923, who were ages 75-79 in 1998. We use observations from 1998 through 2010, at which time this cohort would be ages 87-91, subject to survival. We choose this as our central cohort, because it is observed during those ages when the risk of nursing home entry rises steeply. Data from this cohort need to be adjusted for left censoring, and for survivors to 2010 the data need to be adjusted for right censoring. (2) Those born in the years 1931-1936 who were ages 56-61 in the initial 1992 HRS interview. We use their data on nursing home use and out-of-pocket spending as they age to 74-79 in 2010. Those who survive to 2010 are used to correct for the left-censoring of data on individuals from our central cohort. The data on nursing home use and out-of-pocket spending of those who die before 2010 are used to complete the lifetime use of an entire cohort that reaches ages 57-61. (4) Those born in the years 1912-1918, who were ages 80 – 86 in 1998, and (4) those born in the years 1905-1911, who were ages 87 – 93 in 1998. We use observations from 1998 through 2010 for these latter two cohorts, at which time survivors would be ages 92 – 98 and 99-105, respectively. Data on nursing home use and out-of-pocket spending by these cohorts were used to correct for right censoring of our central cohort.

2. **Representation of nursing home population in the HRS**

Because the HRS started out as a community-based probability sample, it had no nursing home residents at baseline. The HRS first interviewed those ages 51-61 in 1992 (birth cohort 1931-1941). An older cohort, age 70 and over, was first interviewed in 1993 ("AHEAD cohort", birth years 1923 and earlier). Because nursing home residence is not rare at older ages, basing estimates of nursing home use on data from the initial wave and following wave would lead to under-estimation of use. However, because the HRS follows individuals when they move to nursing homes, after several waves the HRS represented the nursing home population. Figure S1 shows the fraction in nursing-home residence by age for each wave of the AHEAD cohorts. Because the AHEAD wave 1 sample is drawn from the community, nursing-home stay prevalence upon entering the study is zero at all ages. By wave 2, substantial numbers are living in a nursing home, for example, 7.1% of those 86-87 years of age. Nonetheless the curve for wave 2 mostly lies below the curves for later waves suggesting that after two years, nursing homes still had residents that were not represented in the initial AHEAD wave and its follow-up wave. By wave 3 (1998) or five years after the baseline wave, the prevalence of residing in a nursing home was about the same as in later waves, leading us to surmise that by then, at least as far as prevalence is concerned, the AHEAD cohort was representative of the entire population, not just the community-dwelling population. For the AHEAD cohorts, we begin our data analysis in 1998 at wave 3. But because nursing home residence is very rare at younger ages (less than one percent among those in their 50s), we use data from the baseline (1992) wave of HRS respondents ages 56-61.

3. **Unit nonresponse and loss to follow-up**

Table S1 shows the vital status in HRS 2010 of the four birth cohorts used in this paper. For our central cohort (born between 1919 and 1923) we had completed surveys (including exit interviews) on 92.1% of the sample. An additional 3.8% were alive in 2010 but not interviewed in that wave. We imputed their nursing home use from the last observed wave before 2010. We imputed nursing home use that would be recovered in the exit interview for the 2.7% of the sample lacking an exit interview. We deleted from the sample those whose vital status was unknown (1.5%). For the middle AHEAD and older AHEAD cohorts (born between 1912-1918 and 1905-1911) we had complete interviews on 94.0% and 95.6% of the sample, respectively. Because these two cohorts are only used to correct for right censoring of the central cohort and because the rate of missing interviews is minimal, we deleted the observations with missing interviews. We treated the youngest cohort in a similar way because it is only used to correct for left censoring.

4. Use of nursing homes and spending by four cohorts

Table S2 has data on nursing home use and spending by individuals from four cohorts directly accumulated from raw data, not adjusted for left or right censoring. Panel (a) shows the experience of the youngest birth cohort, 1931- 1936. Between 1992 and 2010, 29% had died. Averaged over both the deceased and survivors, 15% had some nursing home use, and the average accumulated number of nights (averaged over all, not just users) was 46. About 2/3 of the nights were completely covered by insurance. Discounted out-of-pocket spending was $1,078 on average. This is discounted to age 57 using a 3% real discount rate to account for the time-value of money and to permit a comparison with wealth at an early age when there has been little nursing home use. The fraction that died is substantially higher than the fraction with nursing home use, implying that substantial numbers die without a nursing home stay. We use this cohort to correct for left-censoring.

The cohort shown in panel (b) is our core cohort. Conditional on reaching 75-79 in 1998, 48% of those in our core cohort had some nursing home use between 1998 and 2010. Thus, even ignoring any right censoring, the lifetime risk of nursing home is substantially greater than the widely cited lifetime risk of 37%. By 2010, 35% were still alive and would experience further nursing home use. Accumulated out-of-pocket spending discounted to age 57 averaged $5,253.

The next two cohorts shown in panels (c) and (d) consist of the survivors to ages 80-86 (Middle AHEAD) and to age 87-93 (Older AHEAD). They experienced considerably higher nursing home use and out-of-pocket spending because of their greater age. According to a year 2000 life table, there was a 30% chance of surviving from age 57 to 87, which implies a fairly large risk of surviving to experience the high levels of nursing home use and spending shown in the table. We use observations for these cohorts to correct for right-censoring.

5. Validation of self-reported data on nursing home use and associated out-of-pocket expenditures

We compared average nursing home use and out-of-pocket spending as reported in the HRS with external data. We take 2002 as the comparison year because it is in the middle of our observation period.

Nursing home use

In the 2002 HRS, the average number of nights in a nursing home in the population 55 or older were 19.3 for 2 years or 9.7 per year. We used the following external data for an alternative calculation of average nights spent in a nursing home in the population age 55 or older: total spending on nursing homes: 94.5 billion[1]; population age 55 or older: 59.6 million[2]; average cost per night in a nursing home: $157.[3] We calculate the average number of nights per person per year in 2002 in the population age 55 or older to be 10.1 nights (=94.5/59.6/157/1000). Note that the total spent on nursing homes will include some spending by people younger than 55. The resulting error is expected to be very small, but if corrected it would lead to a slight reduction in the estimated average number of nights, bringing it closer to the estimate based on HRS data.

Out-of-pocket spending on nursing home stays

According to HRS 2002 data, average out-of-pocket spending for nursing home care among those 55 or older was $820 for two years or $410 per year. According to the U.S. Census, there were 59.6 million people 55 or older in 2002, so that total out-of-pocket spending would be $24.5 billion. Actual total out-of-pocket spending for nursing homes in 2002 was $27.9B (CMS National Health Expenditures) for the entire population; the total for those 55 or older would be slightly smaller. Thus our estimates are close to national aggregates.

6. Imputation for left and right censoring and early mortality

Individuals from our central cohort were initially ages 75-79 in 1998. By 2010 some had died, and for those individuals we have data on completed nursing home use. The survivors, however, may have additional nursing home use; that is, their nursing home use is right censored. We impute remaining nursing home use by matching by sex and single years of age a donor to a recipient. For example, a 90 year-old survivor to 2010 (the recipient) will be matched to a 90 year-old donor of the same sex. The donor is chosen from the pool of 90 year-olds from the 1998 wave of the AHEAD cohort, that is, those who were 90 in 1998. By 2010 the donor would have been 102. (We ignore any right censoring of the nursing home use by the donor beyond age 102.) Because we are interested in any lifetime nursing home use, and in the distribution of out-of-pocket spending, we match recipients and donors on the basis of observed out-of-pocket spending on nursing home use in the HRS 2010 interview, which records spending between the 2008 and 2010 interview. Specifically, we choose the donor that has the same percentile ranking in the distribution of out-of-pocket spending among those in the donor pool as the percentile ranking of the recipient among all 90 year-old recipients of the same sex. We then assign the rest-of-lifetime (from age 91 on) nursing home use and associated out-of-pocket spending of the donor to the recipient. We follow this procedure for each survivor from our central cohort, changing the donor pool according to the age and sex of the recipient. For respondents who are still alive in 2010, but did not provide an HRS interview in 2010 we perform the same splicing procedure using their last available interview from an earlier wave. For example, for a respondent from our central cohort who was age 77 at baseline in 1998 and age 87 when last interviewed in 2008 yet was still alive in 2010, we match this observation to a donor of the same age and sex and same percentile rank of out-of-pocket spending observed in 1998.

For left censoring (nursing home use prior to our observation period) we follow a similar procedure. The recipient is someone who survived to 75-79 in 1998. To account for nursing home use and the associated out-of-pocket spending prior to 1998 we choose from the donor pool of individuals of the same sex and single year of age in 2010. These individuals were 57-61 in 1992. We choose the donor that has the same percentile ranking of out-of-pocket spending in 2010 among all donors of the same age and sex as the percentile ranking of out-of-pocket of the recipient among all recipients. The nursing home use and out-of-pocket spending of the donor prior to 2008 is assigned to the recipient.

"Early mortality" is death prior to the initial ages of our central cohort, ages 75-79. Because nursing home use increases with age, failure to account for those who die early would overestimate lifetime nursing home use. We use data on HRS respondents who were initially ages 57-61 in 1992. The survivors were used as donors to account for left censoring as described above. The completed lifetime nursing home use and out-of-pocket spending by the decedents as observed in the HRS interviews from 1992 until death (including exit interviews) were used to account for early mortality. We weighted each individual by life table survival probabilities to account for the fact that HRS cohorts are not the same size. By this procedure we generated the missing part of our central cohort, that is, those who died prior to ages 75-79.

Robustness of matching method

We performed an experiment in the data to check the accuracy of our matching method. We selected a cohort where we observe completed nursing home use and then deleted some of their data as if they were not observed in some last years of their lives. Then we matched them to donors to impute the now-missing data, following the same procedures as in the paper. We call this "splicing." We then compared the actual data with the spliced data.

We selected those who were aged 75-79 in 1998 and who died by 2010. Thus we observe completed rest-of-lifetime nursing home use and out-of-pocket spending from those initial ages. The splicing procedure specifies an overlap wave where both the recipients and donors are alive so that we can match individuals on the level of out-of-pocket spending. We use HRS 2004 (wave 7) as the overlap wave so that we are matching on out-of-pocket spending incurred between HRS 2002 and HRS 2004.

Among those initially 75-79 there are two groups: those who survived to 2004 (survivors) and those who died by 2004. The survivors were 81-85 in 2004. By 2010 they had all died by our selection; they would have been 87-91. Then we artificially set survivors data to missing in 2006, 2008 and 2010 and we replaced up to three waves (6 years) of actual data with three waves of spliced data.

The donor pool is those aged 81-85 in HRS 2000 (wave 5) who died by 2006. It is necessary to use the same mortality criteria because of the increased intensity of nursing home use in the months prior to death. Thus the recipients and donors were the same ages at the overlap year and they survived on average the same number of years after the overlap year.

There were 1,329 persons aged 75-79 in HRS 1998 who had died by 2010, and so had completed rest-of-lifetime nursing home use and out-of-pocket spending from ages 75-79 (Table S3). About 56% had a nursing home stay and average out-of-pocket spending on those stays was $5,733 (including zeros). Of the 1,329, 649 were alive in 2004 and were subject to splicing in this experiment. Comparing the "actual" and "spliced" data for those alive in 2004 shows that average nursing home out-of-pocket spending, the frequency of any nursing home use and the average age at death were almost unchanged by the splicing. However, splicing increased the number of nursing home nights and the number of nursing home nights completely covered by insurance by 64 and 51 respectively. At the population level (All) the increase was 31 and 25 nights respectively. Thus almost all the increase was in covered nights so there was minimal effect on spending out-of-pocket. The explanation for the close match of actual and spliced out-of-pocket spending and the less close match of the number of nights lies in the fact that the splicing algorithm used out-of-pocket spending as a matching criterion, while number of nights was not used for matching purposes.

A graphical representation of the actual distribution of out-of-pocket spending of the population (1329 observations) is in Figure S2 along with the distribution when data on the 649 survivors to 2004 have been spliced. The distributions are almost identical, with the most important deviation beginning at about the 97th percentile where the actual data are below the spliced data. For example, the 98th percentile is $74k in the spliced data and $69k in the actual data. However, at the extreme (top 0.1%) the actual are again greater than the spliced. Thus the spliced data do not completely replicate the values at the very top of the distribution. All in all, however, it appears to us that the spliced data convey very well the main results from the actual data.

7. Parametric model

Overview

The goals of the parametric model are to provide an alternative to the nonparametric estimates of nursing home use and out-of-pocket spending, to use the model in forecasting future use and spending by altering initial conditions, and to take advantage of the flexibility of the model to find how nursing home use and spending vary by personal characteristics. In addition, we are interested in comparing model-based estimates of the type used in prior studies with the nonparametric estimates using the same data to assess how well such models replicate the observed process over a long period of time. The overall strategy is to estimate panel data models of the probability of having any nursing home use and of mortality between waves of the HRS, and the expected number of nights in a nursing home and expected out-of-pocket spending. We use data from HRS waves from 1998-2010 on persons aged 50-100. Having estimated these models, we then simulate multiple times at the individual level from a representative sample of 50-55 year-olds the trajectories of nursing home use and out-of-pocket spending from age at initial interview until death. With sufficient repetitions of the simulations we calculate average lifetime nursing home use, out-of-pocket spending on nursing home use, and the distributions of use and spending.

Models

We model the probability of any nursing home use between the present wave (wave *t*) and the immediately following wave (wave *t+1*) as reported in the following wave. Because of persistence in nursing home use, we employ Markov models, and because mortality is associated with elevated nursing home use, we distinguish between nursing home use by those who survive to the following wave and nursing home use by those who die between the current wave and the following wave. Thus we specify the probabilities of four outcomes in wave t+1:

1) No nursing home use between the current wave and the following wave and survived to the following wave;
2) Some nursing home use between the current wave and the following wave and survived to the following wave;
3) No nursing home use between the current wave and the following wave and died prior to the following wave;
4) Some nursing home use between the current wave and the following wave and died prior to the following wave.

In the last two cases the nursing home use is reported in the exit interviews.

Each of these probabilities depends on observed characteristics such as age, sex, and education, on whether the respondent reported nursing home use in wave *t,* and on whether the respondent reported nursing home use in wave *t*-1. Thus the probability of nursing home use between the present wave and the following wave will depend on nursing home use in each of the (approximately) two-year periods preceding the present wave, that is, from *t-2* to *t-1* and from *t-1* to *t.* To account for the amount of nursing home use and out-of-pocket spending, we specify and estimate econometric models of the number of nights spent in a nursing home over a two-year period; whether any of the nights resulted in out-of-pocket costs; and the amount of out-of-pocket costs conditional on having positive out-of-pocket costs. The specifications follow.

Probability of having any nursing home use between t and t+1.

Let $i = 1,...,N$ denote respondents and $t = 1,...,T$ denote the wave during which an interview takes place. Each wave takes place approximately every two years. We use reports of any nursing home stays in the previous two years and reports of mortality to construct a combined status variable, d_{it}, which can take four values, as listed above. Obviously, states 3 and 4 are absorbing. Hence, we model the four outcomes to depend on each of the two states 1 and 2 in the previous wave. To permit additional state dependence, we use second order Markov models, accounting for nursing home use in each of the two preceding waves.

We define the probability of entering state $j = 1,...,4$ at $t+1$ given a current state $k = 1,2$ at t, a vector of socio-demographic characteristics x_i, and age a_{it} using a multinomial logit :

$$Pr(d_{it+1}=j|x_i,a_{it},d_{it}=k,d_{n,it-1})=\frac{\exp(x_i\gamma_{1,jk}+\gamma_{a,jk}(a_{it})+\gamma_{d,jk}d_{n,it-1})}{\sum_{j'}\exp(x_i\gamma_{1,j'k}+\gamma_{a,j'k}(a_{it})+\gamma_{d,j'k}d_{n,it-1})}$$

where $d_{n,it-1}=I(d_{it-1}=2)$ is an indicator variable for whether the respondent reported nursing home use at the previous interview (that is, between t-2 and t-1) and permits the estimation of additional persistence effects. We do not impose parametric restrictions on the functions $\gamma_{a,jk}$ and instead use categorical variables indicating 5-year age groups from age 50 to 100. After obtaining estimates of the parameters by maximum likelihood, we interpolate linearly the age functions at single years of age intervals. We extrapolate for ages between 100 and 111 (maximum age in the simulations).

Number of nights spent in a nursing home between wave t and t+1

We use reports of the number of nights spent in a nursing home between waves. The models are estimated over individuals who report at least some nursing home nights between the present and following wave. They take the form

$$\log v_{it}=x_i\beta_{jk}+\beta_{a,jk}(a_{it})+\epsilon_{it}$$

Where ε_{it} is assumed normally distributed with mean 0 and variance $\sigma_{\varepsilon,jk}$. Again, we assume the age functions are given by a set of age indicators (5-year age groups). We use interpolation for intervening years.

We estimate separate models of the log of the number of nights v_{it} between waves

a) for individuals who reported no nursing home use since the previous wave (between t-1 and t) and
    1) survived to the next wave (t+1) or
    2) died between waves; and

b) for individuals who reported some nursing home use since the previous wave (between t-1 and t) and
    3) survived to the next wave (t+1) or
    4) died between waves.

Probability of any out-of-pocket costs

For each stay in a nursing home, we have information on whether the respondent had any out-of-pocket costs or whether costs were entirely covered by insurance. We use those reports to construct an indicator variable, $p_{it}=1$ if costs were positive and zero if not. We model the probability of having positive costs as a function of characteristics, age, and the number of nights spent in a nursing home using a logit formulation:

$$\Pr(p_{it}=1|x_i,a_{i,t},\log v_{it})=\frac{1}{1+\exp(-(\psi_a a_{i,t}+\psi_x x_{i,t}+\psi_v\log v_{it}))}$$

Out-of-pocket costs amount

Conditional on costs being positive, we model the log of out-of-pocket costs using a linear regression as a function of age, characteristics, and a fourth-order polynomial in the log of the number of nights in a nursing home:

$$\log oop_{it} = x_i \alpha_x + \alpha_a(a_{it}) + \sum_{j=1}^{4} \alpha_{v,j} (\log v_{it})^j + \mu_{it}$$

where μ_{it} is an error term normally distributed with variance σ_μ.

8. Parametric model: data and model estimates

Figure S3 shows the probabilities of any nursing home use and vital status at wave $t+1$, conditional on reporting no nursing home use at the preceding wave (that is, use between waves $t-1$ and t). At younger ages the probability of being alive and having no nursing home use is close to 1.0, but that steadily declines with age. The probabilities of all other outcomes increase with age. Figure S4 also shows the probabilities of nursing home use and vital status at wave $t+1$, but now conditional on reporting some nursing home use at the preceding wave (use between waves $t-1$ and t). At younger ages the probability of surviving and having no nursing home use is about 60%, and until age 80 this is the most likely outcome. The probability of dying without any nursing home use is approximately constant with age at about 5%. The probability of dying and using a nursing home increases sharply with age, reaching about 50% by age 90.

Comparison of Figures S3 and S4 shows the importance of conditioning on use between the previous waves: for example, the probability of being alive at wave $t+1$ and having some nursing home use between waves *t* and $t+1$ is close to zero at age 57 among those who had no use between waves $t-1$ and t, whereas the probability is about 0.27 among those who had use between waves $t-1$ and t.

Table S4 has estimates of the multinomial logit model of nursing home use and vital status among those not reporting any nursing home use between the previous wave (wave $t-1$) and the present wave (wave t). The reference group is those who survive to the following wave and do not report any nursing home use between the current (wave t) and following wave (wave $t+1$). The probability of reporting nursing home use between the current wave and the following wave increases strongly with age as does the probability of dying between the waves. Persistent heterogeneity in nursing home use is shown by the significant coefficient on "lagged NH use" which indicates a report of nursing home use between waves $t-2$ and $t-1$. Relative to females, males have a greater risk of dying, whether or not they use a nursing home, and they have reduced chances of surviving and not having nursing home use. Lower education is consistently associated with higher nursing home use and higher mortality. Being married at age 50 also protects against entering a nursing home. Of course, the natural channel for this association is that one spouse may be able to provide help to the other who needs it. Those married at age 50 are less likely to die. Less nursing home use is associated with having more children although the effect is significant in just two instances. An indicator for having a daughter is not significant although the point estimate acts to reduce nursing home use. Those who smoked have greater nursing home use and greater mortality risk.

Table S5 has estimates of a similar model, but conditional on reporting nursing home use between the previous and present waves. The reference group is those who survive to the following wave and report nursing home use between the current and following waves. Now the patterns of the age coefficients vary according to the outcome: age reduces the chances of survival and no nursing home use, whereas it increases the chances of mortality and nursing home use. Persistence in nursing home use is indicated by the coefficient on "in NH last wave" where someone who reported nursing home use between waves $t-2$ and $t-1$ is more likely to report nursing home use between the current and following waves.

For the intensity of nursing home use, we estimate separate models according to whether the respondent had any nursing home use between the previous wave and the present wave. Table S6 reports estimation results for the log of the number of nights spent in nursing homes, conditional on no nursing home use between the previous and present waves. Because the intensity of use depends on whether a respondent survived to the following wave, we estimate separate models conditional on vital status at wave $t+1$.

We find that for both mortality outcomes, log nights increases with age and being male, whereas having a college education and having daughters reduce the number of nights spent in a nursing home.

Table S7 has similar results, but now conditioned on having some nursing home use between the previous and present waves. The number of significant coefficients is substantially less than in Table S6, possibly reflecting the smaller sample size. The number of nights generally increases with age. Having a college degree and having daughters both decrease the number of nights spent in a nursing home. Being non-white increases the number of nights.

Table S8 has estimates of a logit model of the probability that out-of-pocket spending, conditional on having nursing home use, is positive. The probability increases with age, likely because of the increasing fraction of nursing home nights that are due to long-term care not covered by Medicare. The coefficient on "log nights in NH" is 0.320, which indicates that as the number of nights increases, the likelihood that any out-of-pocket spending is incurred increases but at a declining rate, likely reflecting Medicaid's becoming the payer. The likelihood of incurring out-of-pocket costs is higher for the better educated.

Table S9 has the estimates of the regression model for the amount of out-of-pocket spending conditional on having any spending. Out-of-pocket spending increases with age. It is higher for the more educated, for whites, and for those married at age 50, which reflects greater economic status for those groups and hence a lower likelihood that Medicaid pays.

Simulation and confidence intervals

The simulations begin at ages 50-55. To gain sufficient data on initial conditions we draw with replacement 10,000 respondents from waves 1998-2010 of the HRS. We then simulate their history until death (or the maximum age of 111) by drawing from the models above.

We repeat this process 500 times to compute the statistics reported in Tables 3 and 4 of the paper. Let $\hat{\theta}$ be the vector of all parameters estimated above. The covariance matrix of these estimates is computed from the asymptotic formula $V[\hat{\theta}]$. At each replication r, we take a draw $\tilde{\theta}_r = \hat{\theta} + L(\hat{\theta})\eta_r$ where *L* is the Cholesky decomposition of V and η_r is a vector of standard normal draws. We use $\tilde{\theta}_r$

when simulating at replication *r*. This incorporates the uncertainty from the estimation of the parametric models. We compute the confidence intervals from taking the 2.5 and 97.5 percentiles of the distribution of simulated statistics.

Figures

FigureS1
Fraction residing in nursing home at interview (AHEAD cohorts born in 1923 or earlier)

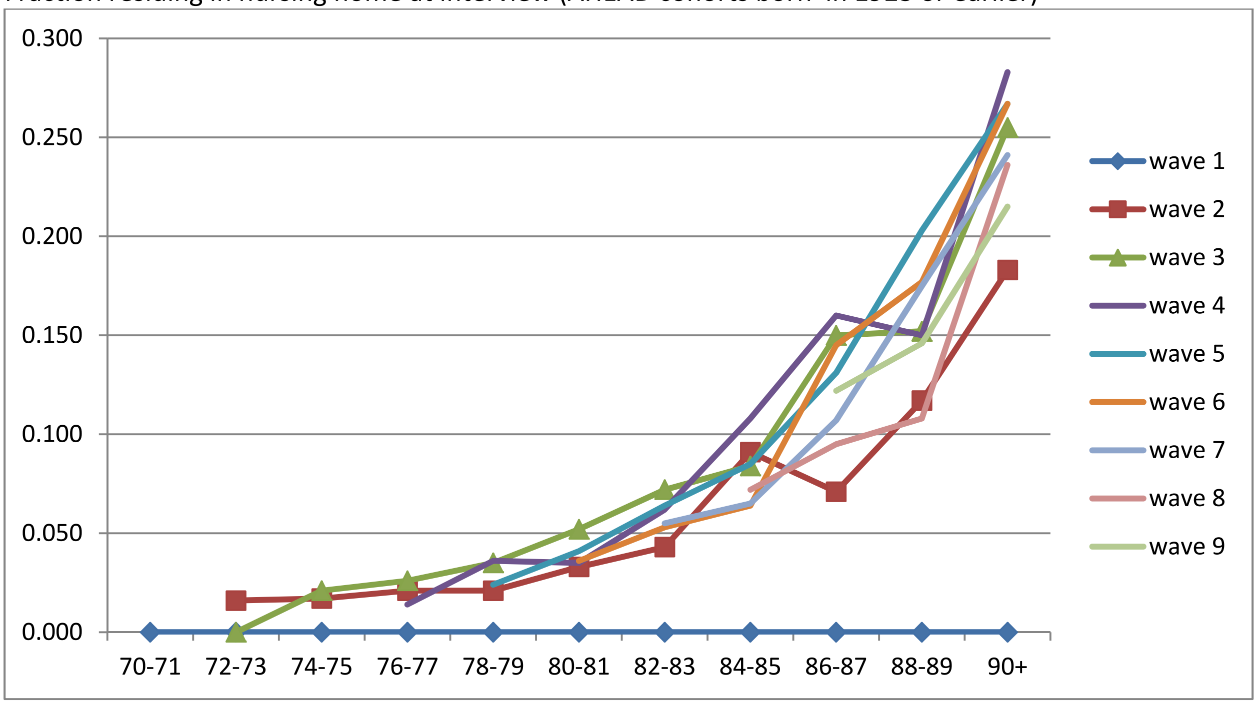
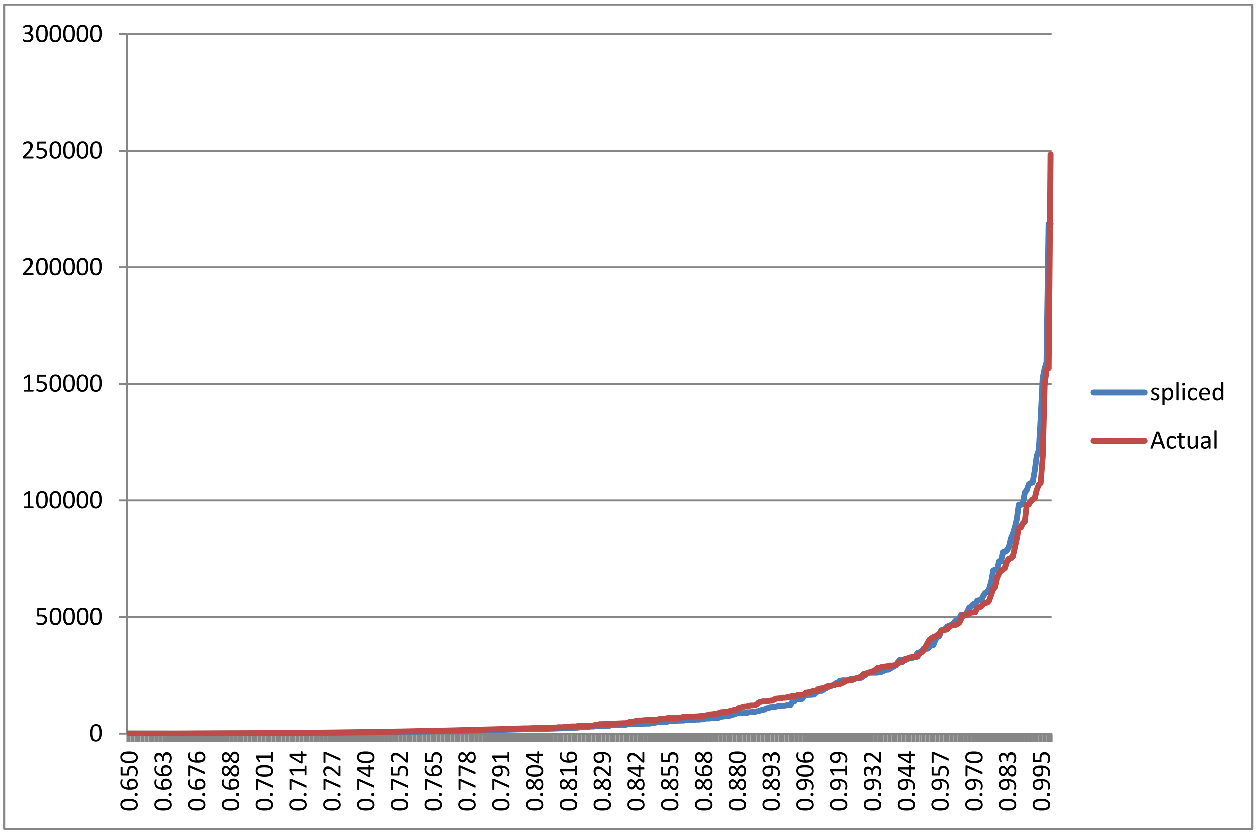
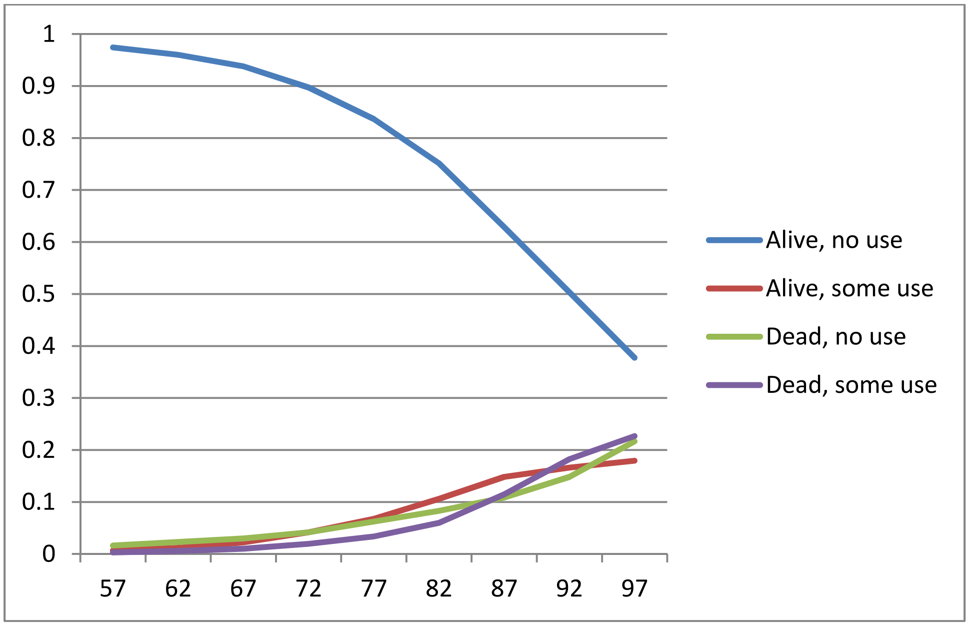


Figure S2. Cumulative distributions of actual and spliced out-of-pocket spending from 1998 to end of life by those 75-79 in 1998 and dead in 2010.



Note: maximum deleted for readability: $340k actual and $219k spliced

Figure S3: Probabilities of nursing home use and vital status, conditional on reporting no nursing home use at previous interview.

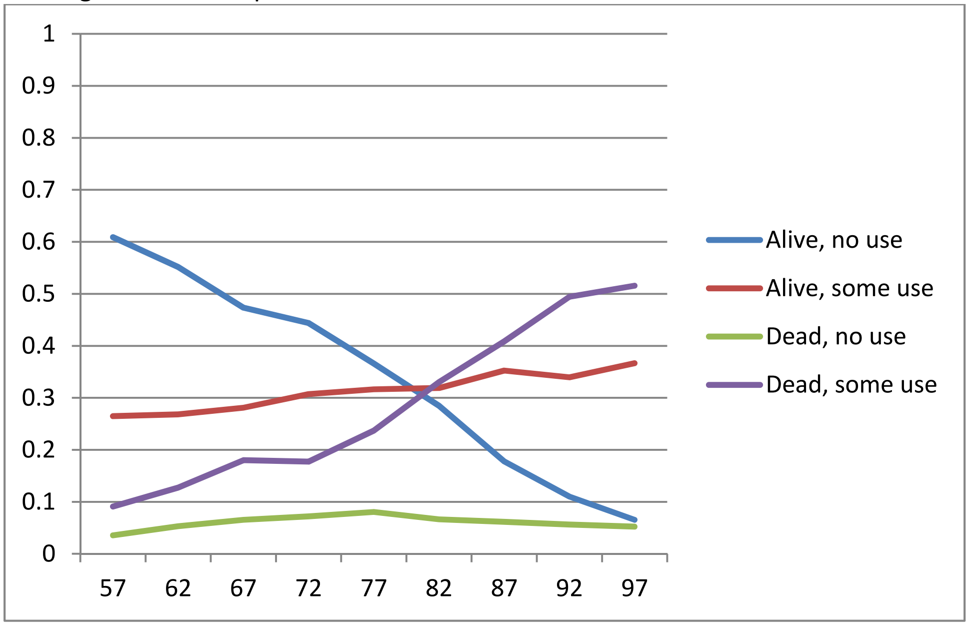


Notes:

"Nursing home use" means having one or more nights in a nursing home between interviews. Vital status pertains to the second interview. All were alive at the first interview.

Probabilities estimated over the sample of respondents who reported no nursing home use between the first of those interviews and the preceding interview.

Figure S4. Probabilities of whether any nursing home use and vital status conditional on reported nursing home use at previous interview.



Notes:

"nursing home use" means having one or more nights in a nursing home between interviews. Vital status pertains to the second interview. All were alive at the first interview.

Probabilities estimated over the sample of respondents who reported some nursing home use between the first of those interviews and the preceding interview.

Tables

Table S1. Vital status in 2010: number of observations and missing observation status.

	HRS cohort (used to correct for left-censoring)		Young AHEAD (central cohort)		Middle AHEAD (used to correct for right-censoring)		Older AHEAD (used to correct for right-censoring)	
	N	Percent	N	Percent	N	Percent	N	Percent
died, exit interview	1,317	27.4	1,370	62.7	1,531	81.2	680	93.9
died, no exit interview	129	2.7	58	2.7	39	2.1	17	2.4
alive in 2010, interviewed in 2010	2,556	53.1	641	29.4	242	12.8	12	1.7
alive in 2010, not interviewed in 2010	358	7.4	83	3.8	42	2.2	6	0.8
vital status unknown	453	9.4	32	1.5	32	1.7	9	1.2
Total	4,813	100.0	2,184	100.0	1,886	100.0	724	100.0

Notes:
"HRS cohort" (birth years 1931-1936) are those initially age 56-61 in 1992. Followed until 2010.
"Young AHEAD" (birth years 1919-1923) are those initially age 75-79 in 1998. Followed until 2010.
"Middle AHEAD" (birth years 1912-1918) are those initially age 80-86 in 1998. Followed until 2010.
"Old AHEAD" (birth years 1905-1911) are those initially age 87-93 in 1998. Followed until 2010.

Table S2. Nursing home use and associated out-of-pocket costs, Health and Retirement Study, descriptive statistics[a]

	N	Fraction Died[b]	Fraction with any NH[c]	Mean NH nights[d]	Fraction nights covered completely[e]	Discounted mean OOP spending[f]
a) Age in 1992 (birth cohorts 1931-1936): HRS Cohort						
56	883	0.231	0.104	39.7	0.703	731
57	887	0.242	0.127	33.1	0.675	770
58	833	0.278	0.174	49.0	0.654	1,249
59	798	0.311	0.168	39.8	0.660	536
60	836	0.345	0.198	60.1	0.617	2,075
61	576	0.319	0.150	58.9	0.649	1,086
All	4813	0.285	0.153	46.1	0.655	1,078
b) Age in 1998 (birth cohorts 1919-1923): Young AHEAD, Central Cohort						
75	460	0.560	0.437	169.9	0.501	5,965
76	451	0.619	0.421	160.6	0.505	4,054
77	469	0.644	0.479	194.1	0.535	4,930
78	421	0.687	0.516	187.8	0.512	4,921
79	383	0.754	0.583	223.6	0.537	6,555
All	2,184	0.648	0.483	185.9	0.519	5,253
c) Age in 1998 (birth cohorts 1912-1918): Middle AHEAD						
80	341	0.732	0.599	338.6	0.564	6,894
81	285	0.820	0.600	303.2	0.547	9,871
82	271	0.791	0.584	362.0	0.563	7,175
83	275	0.870	0.717	380.2	0.586	7,846
84	268	0.876	0.653	365.3	0.602	7,191
85	261	0.918	0.700	444.0	0.486	12,379
86	185	0.904	0.698	379.1	0.497	7,076
All	1,886	0.833	0.643	362.3	0.553	8,305
d) Age in 1998 (birth cohorts 1905-1911): Older AHEAD						
87	161	0.963	0.671	511.2	53.4	8,910
88	157	0.966	0.703	343.6	48.2	8,915
89	115	0.937	0.674	512.8	39.6	11,743
90	107	1.000	0.769	573.8	45.8	13,993
91	79	0.985	0.790	514.5	49.9	13,561
92	57	1.000	0.713	435.6	47.2	9,443
93	48	0.984	0.833	907.1	44.4	22,141
All	724	0.972	0.721	506.6	47.4	11,596

[a] Because the HRS field period is about 10 months, some individuals from a particular birth cohort may be of different ages at interview. Our selection is on age which can cause a discrepancy between the number of single years of age and the number of birth cohorts.

[b] "Fraction died" is the fraction that died between 1992 and 2010 (panel a) or between 1998 and 2010 (panels b, c and d).

[c] "Fraction with any NH" is the fraction that had at least one nursing home stay.

[d] "Mean NH nights" is the average number of nights spent in a nursing home, averaged over all individuals, not just those with a stay.
[e] "Fraction nights completely covered" is the fraction of nursing home nights completely paid for by insurance.
[f] "Discounted mean OOP spending" is the average of accumulated out-of-pocket spending for nursing home stays, discounted at 3% real to age 57, and expressed in 2013 dollars.

Table S3. Nursing home use and out-of-pocket spending from 1998 to end of life by those 75-79 in 1998 and dead in 2010.

	Alive in 2004		All	
	Actual	Spliced	Actual	Spliced
Average nursing home OOP spending	8,195	8,149	5,733	5,711
Any nursing home stay	0.649	0.675	0.561	0.573
Number of NH nights	250.2	314.3	211.5	242.8
Number of NH nights completely covered by insurance	101.6	153.1	96.3	121.4
Average age at death	86.5	86.3	83.4	83.4
N	649	649	1329	1329

Notes: Spending adjusted to 2013$ and discounted to age 57 at 3%.

Table S4. Multinomial logit estimates of the probability of vital status and any nursing home use, conditional on not reporting any nursing home use between the previous wave and the present wave

	Vital status and nursing home use between present wave (t) and following wave (t+1)								
	alive & nursing home use			**died & no use**			**died & nursing home use**		
	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value
Age									
50-54	-1.933	0.296	0.000	-0.897	0.144	0.000	-1.547	0.365	0.000
55-59	-1.166	0.113	0.000	-0.703	0.075	0.000	-1.300	0.176	0.000
60-64	-0.683	0.094	0.000	-0.377	0.067	0.000	-0.589	0.135	0.000
65-69	(ref)			(ref)			(ref)		
70-74	0.677	0.087	0.000	0.167	0.078	0.031	0.657	0.129	0.000
75-79	1.188	0.081	0.000	0.742	0.070	0.000	1.297	0.117	0.000
80-84	1.750	0.076	0.000	1.148	0.067	0.000	1.888	0.109	0.000
85-89	2.276	0.081	0.000	1.645	0.075	0.000	2.695	0.110	0.000
90-94	2.549	0.104	0.000	2.061	0.101	0.000	3.493	0.123	0.000
95 or older	2.774	0.201	0.000	2.984	0.169	0.000	4.008	0.195	0.000
lagged NH use	1.305	0.105	0.000	0.826	0.128	0.000	1.329	0.128	0.000
male	-0.408	0.047	0.000	0.365	0.040	0.000	0.232	0.059	0.000
Education									
< high school	(ref)			(ref)			(ref)		
high school	-0.139	0.053	0.008	-0.389	0.047	0.000	-0.130	0.068	0.055
college	-0.257	0.055	0.000	-0.522	0.048	0.000	-0.404	0.072	0.000
nonwhite	-0.020	0.059	0.740	0.183	0.049	0.000	0.058	0.078	0.457
married at 50	-0.317	0.053	0.000	-0.168	0.050	0.001	-0.407	0.069	0.000
Number of children									
none	0.107	0.098	0.276	0.011	0.095	0.906	0.048	0.122	0.691
1-3	(ref)			(ref)			(ref)		
4 or more	-0.051	0.047	0.285	0.101	0.042	0.015	-0.182	0.064	0.004
any daughter	-0.025	0.062	0.686	-0.088	0.057	0.121	-0.087	0.079	0.272
ever smoked	0.259	0.045	0.000	0.588	0.043	0.000	0.447	0.061	0.000
constant	-3.585	0.104	0.000	-3.616	0.092	0.000	-4.571	0.142	0.000

N= 90,338

Notes:

Outcome is vital status at wave $t+1$ and whether any nursing home use between waves t and $t+1$.

Reference category is "alive at wave $t+1$ and no nursing home use between waves t and $t+1$."

Estimation over sample of respondents who had no nursing home use between waves $t-1$ and t.

"lagged NH use" is an indicator variable for any nursing home use between waves $t-2$ and $t-1$.

Table S5. Multinomial logit estimates of the probability of vital status and any nursing home use, conditional on reporting some nursing home use between the previous wave and the present wave

	Vital status and nursing home use between present wave and following wave								
	alive & no use			**died & no use**			**died & nursing home use**		
	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value
Age									
50-54	1.104	0.820	0.178	1.437	1.043	0.168	0.408	1.014	0.688
55-59	0.305	0.278	0.272	-0.824	0.661	0.212	-0.697	0.374	0.063
60-64	0.293	0.228	0.197	-0.383	0.459	0.404	-0.612	0.289	0.035
65-69	(ref)			(ref)			(ref)		
70-74	-0.030	0.203	0.883	-0.158	0.375	0.673	-0.160	0.226	0.480
75-79	-0.336	0.193	0.082	0.154	0.328	0.639	-0.152	0.209	0.466
80-84	-0.382	0.180	0.034	0.034	0.315	0.914	0.451	0.186	0.015
85-89	-1.001	0.186	0.000	-0.121	0.315	0.701	0.472	0.183	0.010
90-94	-1.304	0.226	0.000	-0.057	0.354	0.873	0.840	0.193	0.000
95 or older	-1.869	0.379	0.000	0.014	0.472	0.976	0.831	0.231	0.000
lagged NH use	-1.343	0.121	0.000	-1.078	0.203	0.000	0.112	0.084	0.186
male	0.036	0.107	0.739	0.578	0.171	0.001	0.429	0.097	0.000
Education									
< high school	(ref)			(ref)			(ref)		
high school	0.340	0.115	0.003	-0.137	0.194	0.481	-0.229	0.098	0.019
college	0.589	0.121	0.000	0.197	0.196	0.315	-0.206	0.106	0.052
nonwhite	-0.286	0.135	0.035	0.221	0.211	0.295	0.024	0.114	0.832
married at 50	0.371	0.117	0.002	0.240	0.201	0.233	0.088	0.099	0.374
Number of children									
none	-0.356	0.216	0.099	-0.445	0.363	0.221	-0.116	0.163	0.474
1-3	(ref)			(ref)			(ref)		
4 or more	0.071	0.106	0.502	-0.070	0.182	0.699	0.039	0.099	0.691
any daughter	0.076	0.140	0.585	0.042	0.228	0.853	-0.012	0.117	0.919
ever smoked	-0.201	0.101	0.046	0.208	0.173	0.230	0.069	0.089	0.438
constant	-0.083	0.236	0.724	-2.074	0.404	0.000	-0.558	0.222	0.012

N=3,560
Notes:
Outcome is vital status at wave $t+1$ and whether any nursing home use between waves t and $t+1$.
Reference category is “alive at wave $t+1$ and had some nursing home use between waves t and $t+1$”
Estimation over sample of respondents who had some nursing home use between waves $t-1$ and t
“lagged NH use” is an indicator variable for any nursing home use between waves $t-2$ and $t-1$.

Table S6. Regression of log nights spent in a nursing home, conditional on some nursing home use between the present wave and following wave but no use between the previous wave and the present wave

	Vital status at following wave					
	alive			dead		
	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value
Age						
50-54	-0.135	0.302	0.655	-0.090	0.382	0.814
55-59	-0.142	0.178	0.424	-0.173	0.262	0.508
60-64	-0.233	0.154	0.130	-0.188	0.209	0.368
65-69	(ref)			(ref)		
70-74	0.026	0.135	0.845	-0.037	0.185	0.840
75-79	0.203	0.123	0.100	0.052	0.168	0.755
80-84	0.482	0.119	0.000	0.416	0.163	0.011
85-89	0.680	0.124	0.000	0.471	0.162	0.004
90-94	1.030	0.153	0.000	0.448	0.174	0.010
95 or older	1.001	0.276	0.000	0.465	0.246	0.059
Male	-0.107	0.069	0.118	-0.213	0.080	0.008
Education						
< high school	(ref)			(ref)		
high school	-0.048	0.076	0.532	-0.124	0.090	0.172
college	-0.207	0.078	0.008	-0.179	0.096	0.064
nonwhite	0.435	0.088	0.000	-0.069	0.109	0.528
married at 50	-0.087	0.077	0.259	0.130	0.097	0.179
Number of children						
none	-0.162	0.139	0.244	0.232	0.158	0.142
1-3	(ref)			(ref)		
4 or more	-0.220	0.070	0.002	0.081	0.089	0.361
any daughter	-0.223	0.091	0.014	-0.206	0.106	0.051
ever smoked	0.092	0.065	0.159	-0.040	0.084	0.635
constant	3.850	0.158	0.000	3.488	0.197	0.000
	N= 2,766			N=1,597		

Notes:
Outcome is log nights in a nursing home between waves t and $t+1$ conditioned on vital status at wave $t+1$.
Estimation is on sample of respondents who had no nursing home use between waves $t-1$ and t but some nursing home use between waves t and $t+1$.

Table S7. Regression of log nights spent in a nursing home, conditional on some nursing home use between the current wave and following wave and some nursing home use between the previous wave and the present wave

	Vital status at following wave					
	alive			dead		
	Coefficient	Std. Error	P-value	Coefficient	Std. Error	P-value
Age						
50-54	1.507	1.078	0.162	-0.734	0.944	0.437
55-59	0.419	0.309	0.175	-0.383	0.412	0.353
60-64	0.118	0.251	0.638	0.067	0.326	0.837
65-69	(ref)			(ref)		
70-74	0.499	0.212	0.019	0.014	0.230	0.953
75-79	0.431	0.192	0.025	0.182	0.199	0.361
80-84	0.617	0.184	0.001	0.039	0.183	0.833
85-89	0.831	0.181	0.000	0.269	0.181	0.136
90-94	0.849	0.196	0.000	0.336	0.184	0.068
95 or older	1.175	0.237	0.000	0.474	0.209	0.023
male	-0.086	0.098	0.381	-0.117	0.082	0.154
Education						
< high school	(ref)			(ref)		
high school	-0.191	0.096	0.048	-0.092	0.085	0.284
college	-0.340	0.103	0.001	-0.146	0.092	0.111
nonwhite	0.384	0.111	0.001	0.071	0.103	0.493
married at 50	0.038	0.095	0.687	-0.194	0.087	0.025
Number of children						
none	0.164	0.156	0.292	-0.090	0.137	0.511
1-3	(ref)			(ref)		
4 or more	0.065	0.096	0.498	-0.123	0.087	0.158
any daughter	-0.268	0.115	0.020	-0.109	0.099	0.270
ever smoked	-0.059	0.087	0.500	-0.051	0.077	0.509
constant	5.285	0.215	0.000	5.230	0.207	0.000
	N=1426			N=1369		

Notes:
Outcome is log nights in a nursing home between waves t and $t+1$ conditioned on vital status at wave $t+1$.
Estimation is on sample of respondents who had some nursing home use between waves $t-1$ and t and some nursing home use between waves t and $t+1$.

Table S8. Logit estimation of probability that out-of-pocket spending on nursing homes is positive, conditional on reporting nursing home use

	Probability that OOP is positive		
	Coefficient	Std. Error	P-value
Age			
50-54	0.342	0.361	0.344
55-59	-0.224	0.216	0.300
60-64	0.131	0.167	0.434
65-69	(ref)		
70-74	0.243	0.140	0.081
75-79	0.575	0.125	0.000
80-84	0.642	0.120	0.000
85-89	0.628	0.119	0.000
90-94	0.770	0.127	0.000
95 or older	0.869	0.162	0.000
log nights in NH	0.320	0.016	0.000
male	0.050	0.056	0.374
Education			
< high school	(ref)		
high school	0.528	0.063	0.000
college	0.622	0.066	0.000
nonwhite	-0.655	0.079	0.000
married at 50	0.173	0.065	0.008
constant	-2.902	0.146	0.000

N=7,221

Table S9. Regression of log out-of-pocket spending, conditional on positive spending

	OOP Spending		
	Coefficient	Std. Error	P-value
Age			
50-54	-0.357	0.417	0.392
55-59	-0.526	0.269	0.051
60-64	0.092	0.196	0.639
65-69	(ref)		
70-74	0.030	0.160	0.851
75-79	-0.027	0.142	0.849
80-84	0.132	0.136	0.334
85-89	0.254	0.136	0.062
90-94	0.338	0.142	0.017
95 or older	0.452	0.169	0.008
log nights	1.039	0.419	0.013
log nights^2	-0.756	0.216	0.000
log nights^3	0.185	0.043	0.000
log nights^4	-0.013	0.003	0.000
male	-0.028	0.058	0.631
Education			
less than high school	(ref)		
high school	0.069	0.065	0.289
college	0.261	0.067	0.000
nonwhite	-0.345	0.090	0.000
married at age 50	0.238	0.068	0.001
constant	7.701	0.304	0.000

N=2,707

References

1. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/tables.pdf
2. http://www.census.gov/population/age/data/files/2002/table1.pdf
3. The MetLife Mature Market Institute, 2002, "MetLife Market Survey of Nursing Home and Home Care Costs"