Accounting for Survival in Longitudinal Analyses of Health Care Expenditures

Sally C. Stearns, Ph.D.
Health Economics Research Unit, University of Aberdeen

Edward C. Norton, Ph.D.
University of North Carolina at Chapel Hill

November 27, 2000

Abstract

Longitudinal analyses of medical expenditures are perceived as increasingly important in the policy arena, given the aging and increasing longevity of the population. The Royal Commission on Long Term Care (the Sutherland Report) recommended that longitudinal research should track the outcomes of preventive interventions and indicated that projections of the cost of long-term care are required. Analytically, analyses using data for the elderly population are complicated by the censoring caused by death and the underlying endogeneity of death that results in a high positive correlation between such censoring and the level of health care expenditures.

This paper combines two different ways to incorporate survival in longitudinal analyses of health care expenditures. One estimation approach uses methods proposed by biostatisticians and econometricians to estimate average total cost in a given period as product of the probability of being alive at the start of the interval times the average cost conditional upon surviving to the end of the interval. A second approach provides for separate estimation of costs according to proximity to death. Two different approaches are estimated using the National Survey of Self-care and Aging. The estimations show that the likelihood and level of Medicare expenditures varies with proximity to death, and that models that do not account for such differences may overstate the impact of increased longevity on expenditure predictions.

Corresponding author: Sally C. Stearns, Ph.D.
Health Economic Research Unit
University Medical Buildings
Foresterhill, University of Aberdeen
Aberdeen AB25 2ZD

Office: +44 (0)1224 552494
Fax: +44 (0)1224 662994
e-mail: s.c.stearns@abdn.ac.uk

DRAFT: Please do not cite without permission.

Overview

Longitudinal analyses of medical expenditures are perceived as increasingly important in the policy arena, given the aging and increasing longevity of the population. The Royal Commission on Long Term Care (the Sutherland Report) recommended that longitudinal research should track the outcomes of preventive interventions and indicated that projections of long term care costs are required. Analytically, investigations using data for the elderly population are complicated by the underlying endogeneity of death, the documented increase in health care expenditures at the time of death, and empirical censoring prior to death in many longitudinal surveys.

This paper combines two different ways to incorporate survival in longitudinal analyses of health expenditures. Of particular interest is the relationship between age, proximity to death, and expenditures. Both approaches assume that the longitudinal data are available in multiple discrete time intervals. The two methods differ in the way in which they incorporate proximity to death. One method has been used by biostatisticians and econometricians to estimate total expenditures over a given time period as sum of the product of the probability of being alive at the start of an interval times the average cost conditional upon surviving to the start of the interval. The original goal of this approach, which incorporates the idea of proximity to death by predicting survival, was to obtain expenditure estimates without explicit modeling based on covariates. The approach does allow, however, for separate estimates of independent groups of people, as would occur, for example, in a randomized trial with treatment and control groups. The other approach controls directly for proximity to death by undertaking separate estimations of expenditures for people who die during the observation period. This method allows the estimated model, upon which the predictions depend, to depend on proximity to death and other covariates. Ignoring proximity to death in either the model estimation or in simulations using the model results will result in biased estimates if health care expenditures depend on proximity to death.

Estimation combining aspects of both approaches are conducted using data from a longitudinal United States survey called the National Survey of Self-care and Aging. Medicare expenditure data are available for up to four years following the baseline survey. Three years of follow-up data are used for survivors to avoid measurement error in the time to death variable. The estimation results are used to simulate health care expenditures by age.

The estimation shows, not surprisingly, that time to death is an important predictor of health care expenses and that the likelihood and level of health expenditures varies by proximity to death. The paper concludes with a discussion of some of the problems inherent in controlling for proximity to death and consideration of alternative approaches to address the underlying endogeneity of this measure.

Background

Several recent economic studies have explored the economic implications of an aging population from both a theoretical and empirical viewpoint, and have reached surprising conclusions (Lubitz, Beebe, and Baker, 1995; Zweifel, Felder, and Meier 1996; Lee and Skinner, 1997; Lakdawalla and Philipson, 1998; Cutler and Sheiner, 1998). Although conventional wisdom says that the demographic trends of increased longevity and decreased fertility rates will greatly increase the burden of paying for health care in the future, these studies argue that the burden will not be as great as feared. Health care expenditures will definitely increase because of the increased number of aged persons. These articles argue, however, that increased longevity by itself may not contribute substantially to increased health expenditures. In large part, future expenditures will depend on whether increased longevity means primarily a delay in the more intensive use of health services prior to death. Other factors, such as decreases in age-adjusted disability rates, may work together to limit absolute increases in health care expenditures.

The theoretical argument depends on the relationship between expenditures, age, and time until death. If health care expenditures are largely independent of age, and relatively modest until a person is close to death, then increased longevity will not increase per capita expenditures. Instead, the rise in expenditures found prior to death will be delayed to later ages, if fundamentally expenditures are linked largely to death rather than age. This is perhaps most easily thought of with acute care expenditures, which may rise just prior to death. The situation may differ for long-term care expenses (McGrail et al., 2000).

Figures 1a and 1b provide a descriptive assessment of expenditures using data from the National Survey of Self-care and Aging (described later in this paper). Average monthly expenditures are graphed by proximity to death, where survivor status (represented as time to death equals zero, or TTD=0) means not being in the last year of life. Being in the last year of life ranges from the first month of the year prior to death (TTD=1) to the month of death (TTD=12). Figure 1a, which includes a 95% confidence interval around the estimates, shows that the increase in expenditures is particularly steep during the last six months of life among decedents and that the last two quarters of life may therefore matter the most. Figure 1b shows that there is some differentiation between the pattern of expenditures by age group, with persons age 85 or over having generally lower average monthly expenditures during their last year of life than younger persons (age 65-74 or 75-84). This pattern is consistent with findings by Lubitz and colleagues (1995) that the increase in expenditures during the last year of life is greater for relatively younger persons than for relatively older persons.

Several empirical issues arise in addressing this problem in models for predicting health expenditures and, in particular, estimating the impact of increased longevity on predicted health expenditures. First, many available longitudinal surveys or follow-up data for randomized trials contain data for both survivors and decedents during a given time period. Analyzing data simply on decedents would bias analysis of factors related to health expenditure trends and lead to incorrect estimates of total health expenditures. Lin and colleagues (1997) noted that studies that use either average total costs based on all study subjects or data on just decedents are going to be biased downward (due to censoring) or upward, respectively. Yet simple inclusion of an indicator for decedents versus survivors ignores the potential endogeneity of death, the documented increase in expenditures during the year prior to death, and the censoring of expenditures that occurs for survivors beyond the time period under study.

Lin and colleagues proposed an estimator for medical costs with incomplete follow-up data by tracking costs over small time intervals. Their method does not use covariates, and therefore it does not address modelling issues, but it can be used to estimate aggregate total costs over a specified time period for a group of patients (or subgroups of patients according to some independent variable of interest). Aggregate total costs E over T follow-up periods are estimated by summing the product of the Kaplan-Meier probability of being alive at the start of the interval S_t by the average cost conditional upon surviving to the start of the interval E_t.

$$E = \sum_{t=1}^{T} S_t E_t \qquad \text{(eq. 1)}$$

The method is essentially the same as that used by Manning and colleagues (1989) to estimate the aggregate lifetime costs LC of smokers versus non-smokers, except that Manning and colleagues control for one important covariate, smoking:

$$LC = \sum_{t=20}^{95} d^{\,t-20} \Pr(A \mid SM)_t \times C(SM)_t - \sum_{t=20}^{95} d^{\,t-20} \Pr(A \mid NSM)_t \times C(NSM)_t \quad \text{(eq. 2)}$$

where $\Pr\left(A \mid SM\right)_t$ is the probability of surviving from 20 to t years of age conditional upon smoking, $C\left(SM\right)_t$ is the annual costs minus taxes and premium for smokers at age t, and the second half of the equation contains the corresponding components for nonsmokers (*NSM*).

The value of dividing follow-up periods into relatively small intervals for longitudinal studies of health care expenditures seems quite high. Lin and colleagues conducted Monte Carlo simulations of their method of aggregating costs and found that bias in the estimator is reduced when the intervals are small. They also showed that in the absence of relatively small intervals it might be preferable to make prorated adjustments for any person whose data was censored for reasons other than death during the interval. The separate components of survival and expenditures can both be modeled as functions of various explanatory variables (e.g., demographics, health status, preventive interventions).

While expenditure models using dummy explanatory variables to indicate decedent versus survivor status for long follow-up periods are excessively simplistic, models using multiple small follow-up periods (e.g., monthly or quarterly) might be appropriate for incorporating a time-varying indicator of number of months prior to death:

$$E_{it} = b_0 + b_1 X_{it} + b_2 TTD_{it} + e_{it} \quad \text{(eq. 3)}$$

where TTD measures time to death and X reflects other relevant covariates. Such an approach to estimating per period expenditures E_{it} would enable a more precise control for individuals once they enter a terminal phase, as the year prior to death is often associated with significantly higher expenditures (refs). Implementing this approach would require knowledge for each individual of the point at which he/she begins the last year of life during a longitudinal follow-up and can only be estimated retrospectively.

One might argue that the latter approach is flawed because of potential endogeneity of death. For example, death may be affected by certain practices undertaken by the individual or by the level of health services received in a prior period. The approach followed by Lin and colleagues and by Manning and colleagues allows for explicit estimation of the probability of survival, since expenditures are estimated conditional upon surviving to a specific period. Yet the Lin and Manning approaches impose an assumption of independence between survival and health care expenditures that seems unrealistic. This assumption is not a problem if the purpose is simply to use data on two groups from a randomized trial to estimate differences in total expenditures between those groups (i.e., the original purpose of the estimations by these authors). The assumption, however, may be a major limitation if the purpose of the model is to estimate the impact of increased longevity on health care expenditures. In particular, the factors leading to increased survival will affect the probability of survival as well as (at least potentially) the expected expenditures conditional upon surviving to each age.

Methods

Given the work described above, we consider an approach that combines the two approaches. We estimate expenditures E_{it} for individual i in period t using a three component model as follows:

$$E_{it} = \Pr(S \mid X)_{it} \times \Pr(AnyExp \mid S, X)_{it} \times (LevExp \mid S, X, AnyExp)_{it} \qquad \text{(eq. 4)}$$

where each component on the right-hand side represents an equation that is estimated separately. We call this model the "Naï ve Model" because the concept of proximity to death is not reflected in the expenditure estimations and only enters the estimation as the probability of survival. The first component for the probability of survival (S) uses a discrete time hazard model to estimate the probability of surviving to the start of a particular period as a function of a set of exogenous explanatory variables. More specifically:

$$P(S)_{it} = 1 - \sum_{r=1}^{t-1} P(DIE)_{ir}$$

where $P(DIE)_{ir}$ is the probability of dying in each period r. The second and third components represent a two-part model for health expenditures due to the distribution of monthly expenditures. The second component, the likelihood of having any expenditures during a period, is estimated using a panel data probit model of the likelihood of having any Medicare expenditures during the period conditional upon surviving to the start of the period. The third component, the level of expenditures in a month conditional both upon surviving to the period and having had positive expenditures during that period, is estimated using a logged value of expenditures to account for the highly skewed nature of the data. As is standard in such estimations, a smearing transformation is used in transforming expected log expenditures to expected expenditures. All three components are modeled as functions of a vector X of independent explanatory variables, such as age and gender.

The above model assumes independence of all three parts of the equation. Ultimately, it would be desirable to relax this assumption and estimate the equations jointly using a flexible, non-parametric approach for the joint distribution(s) of the error terms. Yet further attention to the joint distribution of the error terms may not fully reflect the underlying implications of proximity to death, and more comprehensive modification of the equation in other ways may also be appropriate. One possible modification would involve incorporating time to death indicators in the expenditure components (the second and third components), as suggested in equation 3 above. Alternatively, one can decompose the probability of survival to reflect different states of surviving and being close to death versus surviving but not yet being close to death. Given the patterns in Figures 1a and 1b and the increase in expenditures in the last two quarters of life, such a decomposition could control whether someone is in the penultimate quarter before death, or whether they die in a particular quarter.

Conceptually, we can think of the surviving population at any point in time as being composed of three different types of people: (S1) those who are not entering a terminal phase (i.e., those who have a good prospect for continued survival); (S2) those who are not going to die this period but are indeed entering a terminal phase (i.e., those who are in the penultimate period before death), and (S3) those who are going to die during the period. Under this modification, the specification becomes:

$$\begin{aligned} E_{it} = P(S)*[&P(S1)_{it} * P(AnyExp \mid S1, X)_{it} * (LevExp \mid S1, X, AnyExp=1)_{it} \\ &+ P(S2)_{it} * P(AnyExp \mid S2, X)_{it} * (LevExp \mid S2, X, AnyExp=1)_{it} \\ &+ P(S3)_{it} * P(AnyExp \mid S3, X)_{it} * (LevExp \mid S3, X, AnyExp=1)_{it}\,] \end{aligned} \qquad \text{(eq. 5)}$$

Equation 5 is subject to the following constraint:

$$P(S1)_{it} + P(S2)_{it} + P(S3)_{it} = 1 \qquad \text{(constraint for eq. 5)}$$

We call this model (equation 5 and the associated constraint) the "Full Model." In a sense, the resulting estimation is analogous to a fully interacted estimation in that the proximity to death variables are not included as explanatory variables in any of the estimations, yet the parameter estimates for the explanatory variables in the likelihood and level of expenses equations differ according to the proximity to death. The advantage of the Full Model, therefore, may be to reduce the potential bias from inclusion of endogenous proximity to death indicators while still controlling explicitly for different proximities to death at the cost of degrees of freedom used up by separate parameter estimation for each component of proximity to death. As discussed at the end of the paper, separate estimation by survival prognosis does not necessarily escape possible bias from the underlying endogeneity of death.

Given these possible models, our estimation strategy proceeds as follows. First, we estimate each component of the Naï ve and Full Models using a quarterly data set described below. (Aggregation of the data to a quarterly level simplifies estimation of the models without sacrificing the main comparisons.) We control for a range of socio-demographic time-invariant covariates. The only time-varying covariate (aside from proximity to death) is age. We then use the data set to predict the components of the models and compare those predictions to consider the potential implication of proximity to death for the likelihood and level of expenditures by age. Finally, we combine these components to compare simulations of quarterly expenditures by age between the Naï ve and Full Models.

Data

Data for this analysis come from the National Survey of Self-Care and Aging (NSSCA), a longitudinal, nationally representative survey of 3,485 community-dwelling Medicare beneficiaries aged 65 and over in the contiguous United States (Norburn et al., 1995). The sampling universe consisted of all Medicare beneficiaries in the contiguous United States who were 65 years of age or older in 1989 and who were living in the community (i.e., did not reside in nursing homes or domiciliary care facilities) at the time of selection. Baseline in-person interviews were completed during the fall and winter of 1990-1991 with 3,485 adults. Subjects were selected from the Medicare Beneficiary Files according to a stratified random sampling design, with approximately equal numbers of adults by gender in each of three age categories, 65 to 74, 75 to 84, and 85 and older. The sample was clustered within 50 primary sampling units in 38 urban and 12 rural areas across the United States. The baseline in-person interviews, conducted in 1990-1991, contain detailed self-reported information on demographic measures, health status, and self-care activities. Medicare Part A and Part B claims data from 1989-1994 and death records were linked to the survey data. Approximately 91 percent of the respondents were matched successfully to the Medicare claims data based on available identifiers. Medicare expenditures include most acute and ambulatory expenditures by the elderly but exclude deductibles, copayments, and most payments for drugs and long-term care. The expenditures in this paper, therefore, represent only roughly half of total health care expenditures by the elderly.

The original analysis file uses the person-month as the unit of analysis and includes only persons who were continuously enrolled in the fee-for-service program for their entire follow-up period (as expenditure data are not available for HMO enrollees). Months are excluded from the analysis file subsequent to the death of a study subject.

While the claims and date of death data were available for up to four years following the baseline survey, we dropped the last year of follow-up data for people who did not die during the follow-up period to ensure that proximity to death was measured as accurately as possible. The final monthly analysis file consists of 95,787 monthly observations for 2,862 people. Figures 1a and 1b were generated using this monthly file, but we simplify the estimation by aggregating the data to the quarterly observations. The final quarterly analysis file consists of 32,174 observations for the same 2,862 people. A total of 667 sample members (23 percent) died during the follow-up. Table 1 provides descriptive statistics for the covariates included in the model (age, sex, race, education, income, supplemental insurance, marital status, smoking and alcohol consumption status, geographic area, physician and hospital bed supply variables, and measures of proximity to death). As noted earlier, age and proximity to death are the only time-varying variables. Future models will consider inclusion of additional time-varying covariates such as functional status and disease indicators.

Results

Components of the Naïve Model

Table 2 provides the results the estimations for the three components of the Naï ve Model (probability of survival, likelihood of expenditures in a quarter conditional upon survival, and level of expenditures conditional upon survival and some expenditures in a quarter). Many of the covariates are statistically significant in an expected direction. Age, gender and smoking/alcohol status are significant in all three components. (The positive coefficients for no alcohol consumption and former smoker relative to non-smoker in the likelihood and level of expenditure equations may reflect cases where the presence of disease precludes consumption of alcohol or tobacco). More educated people live longer, insurance and income affect the likelihood of use of services, and there is substantial geographic variation in the likelihood and level of expenditures. Figure 2 shows the simulations for the sample of these three components by age. The probability of survival (Figure 2a) conditional upon survival to age 66 shows the expected substantial decline by age 100. (The decline may not be as rapid as would be expected in general for the elderly population conditional upon living to age 66, but the less rapid decline may be attributable to the fact that the community-based sample excluded the institutionalized population. Future estimations may use a more representative sample of the full elderly population or survival estimates from vital statistics sources.) Figure 2b shows that *rate* of increase in the likelihood of expenditures in a quarter decreases somewhat with age, though the overall likelihood increases. The level of quarterly expenditures (conditional upon some expenditures during the quarter) increases linearly with age (Figure 2c). It should be noted that the smearing factor used to transform the predicted log expenditures was rather high, possibly indicating uncontrolled heterogeneity in expenditures between survivors and persons approaching death. Finally, Figure 2d puts the three components together to provide simulated quarterly expenditures by age. The u-shaped pattern reflects initial increases with age due to the increasing likelihood and level of use that are ultimately offset by the decreasing probability of survival.

Components of the Full Model

Tables 3a, 3b and 3c provide the results the estimations for the components of the Full Model for survivors (i.e., persons not in the quarter of death or penultimate quarter), penultimates, and decedents, respectively. (The estimation of the likelihood of being in the quarter of death is exactly the same estimation as was used to construct the probability of survival to the start of a period P(S) in the Naï ve Model.) The likelihood of being in the penultimate quarter or quarter of death increases with age, former or current smoking, and no alcohol consumption at baseline. Males and persons without a high school education are also more likely to be in a penultimate quarter or die relative to persons without these characteristics. Estimated coefficients in the likelihood and level estimations for survivors are similar in sign and significance to the coefficients in the Naï ve Model. Many of the coefficients in the likelihood and level estimations for penultimates and decedents are not significant, possibly due to the substantially smaller number of observations available for these estimations. (Note that the same observations are used in the likelihood of expenditures equations for penultimates and decedents except for two people who died within the first quarter after the baseline interview.)

Figure 3 shows the simulations for the sample of the model components by age. The figures for the probability of expenditures in a quarter show that the likelihood of use first increases and then decreases with age for survivors and penultimates (Figures 3.1b and 3.2b), but it decreases and then increases with age for decedents (Figure 3.3b). The level of expenditures (conditional upon some expenditures) increases and then decreases with age for survivors (Figure 3.1.c), but it declines continuously with age for penultimates and decedents (Figures 3.2c and 3.3c). (Age was not significant in the level of use estimations for penultimates or decedents due to the small number of observations, but graphically the effect of age appears to be strong.) In total, these figures are consistent with some use of acute services for terminally ill persons but with some constraint on the level of use at increasing ages.

Comparison of Expenditures from the Naïve and Full Models

Estimations of the components of the two models reveal reflect important underlying differences in Medicare expenditures in relation to age and proximity to death. From a policy perspective, these differences should be incorporated into expenditure predictions, especially under the expected scenario of increased longevity. Table 4 provides a simulated comparison of predicted expenditures by age for the Naï ve and Full Models. These simulations are constructed by using separate predictions for each of the components in each model. The results show that while the Naï ve Model predictions are slightly lower than the Full Model predictions up to age 80, the Naï ve Model predictions are greater than the Full Model predictions from age 81 to 100 (by as much as 37% at age 100). Figure 4 graphs the simulated expenditures by age for each model. Standard errors or confidence intervals have not yet been constructed for these estimates (as can be done using bootstrapping).   Yet it appears that the Naï ve Model may overpredict expenditures in later periods as would be expected if proximity to death is not sufficiently controlled for in a model. Table 4 shows that the cumulative difference in expenditures (overprediction by the Naï ve Model) can reach as high as $1,511 per quarter for someone living until age 100. Construction of standard errors may ultimately provide evidence supportive of such overprediction.

Discussion

As developed in this paper, proximity to death may be incorporated into expenditure estimations in several ways. At a minimum, estimations conditional upon probability of survival (i.e., the Naï ve Model) are appropriate. Yet further incorporation of proximity to death in period-specific estimations of the likelihood and level of expenditures appears to be appropriate (as in the Full Model). While the Full Model incorporates the impact of proximity to death on the likelihood and level of expenditures without direct inclusion of proximity to death as a covariate, the specification does embody division of the estimation sample by proximity to death. Therefore, the Full Model does not necessarily avoid possible complications due to endogeneity of proximity to death as a covariate, as in a sense it is imposing choice-based sampling on the estimation method. Further attention to this issue is appropriate.

The empirical estimations are not definitive proof of the superiority of one estimation over another, and future work using Monte Carlo simulations may be insightful. Yet the estimations indicate important differences in patterns of the likelihood and level of use (expenditures) with age and proximity to death, and the net differences in the predictions between the Naï ve and Full

Models are indicative of potentially important implications for appropriate prediction of future expenditures.

Other estimation approaches may also merit future consideration. For example, joint estimation of the components of the model allowing for correlated error terms may be important. Given concerns about possible endogeneity bias from division of the sample by proximity to death in the Full Model, consideration of approaches to use instruments to deal with the bias, such as those used by Lindrooth et al. (2000) may be fruitful. Furthermore, the distinction between persons in the penultimate quarter before death and the quarter of death may be limited, and a more simple version of the Full Model combining the penultimate and decedent estimation may provide more efficient estimates. The discussion and estimations in this paper therefore provide further support to the expanding literature on the importance of proximity to death but do not yet provide a full answer to the question of the best way to incorporate proximity to death in expenditure models.

References

Cutler, D.M. and L. Sheiner. (1998) “Demographics and Medical Care Spending: Standard and Non-Standard Effects”, NBER Working Paper #6866.

Lakdawalla, D. and T. Philipson. (1998) “The Rise in Old Age Longevity and the Market for Long-Term Care”, University of Chicago working paper.

Lee, R. and J. Skinner. (1997) “Will Aging Baby Boomers Bust the Federal Budget?” University of California at Berkeley working paper.

Lin, D.Y., E.J. Feuer, R. Etzioni, and Y. Wax. (1997) "Estimating Medical Costs From Incomplete Follow-up Data." *Biometrics* 53(2):419-434.

Lindrooth, R.C., T.J. Hoerger, and E.C. Norton (2000) "Expectations Among the Elderly About Nursing Home Entry," *Health Services Research*, 35(5)Part II:121-142.

Lubitz, J., J. Beebe, and C. Baker. (1995) “Longevity and Medicare Expenditures,” *New England Journal of Medicine* 332(15):999-1003.

Manning, W. et al. (1989) "The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?" *JAMA* March 17, 1989. 261(11):1604-1609.

McGrail, K., B. Green, M.L. Barer, R.G. Evans, C. Hertzman, and C. Normand (2000) "Age, Costs of Acute and Long-term Care and Proximity to Death: Evidence for 1987-1988 and 1994-1995 in British Columbia." *Age and Ageing*, 29(3):249-253.

Norburn, J.E.K., S.L. Bernard, T.R. Konrad, A.Wommert, G. DeFriese, W.D. Kalsbeek WD, G.G. Koch, and M.G. Ory. (1995) "Self-care and Assistance from Others in Coping with Functional Status Limitations Among a National Sample of Older Adults." *Journal of Gerontology*, 50B(2):S101-S109.

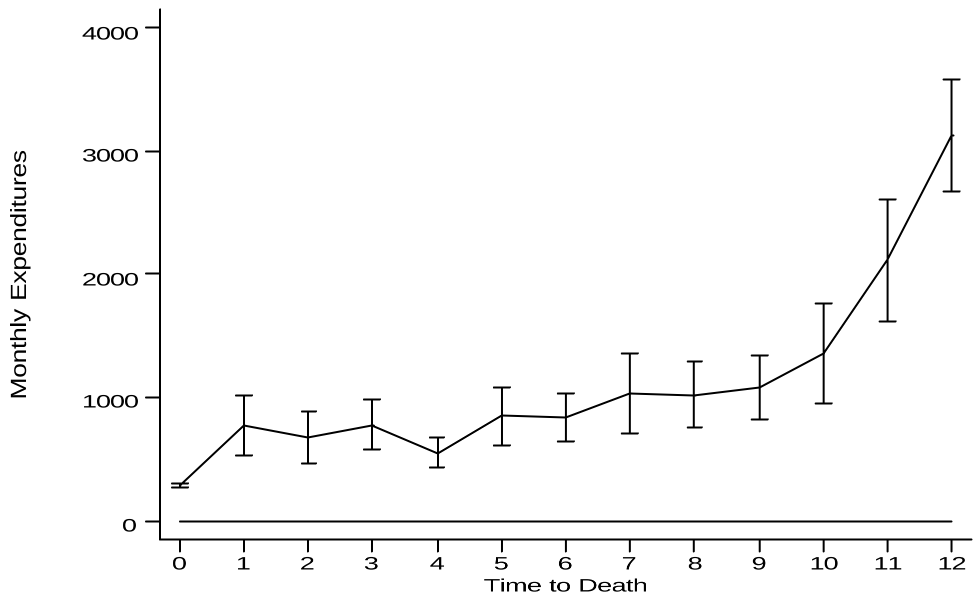

Figure 1a: Monthly Medicare Expenditures by Time to Death (with 95% CI)

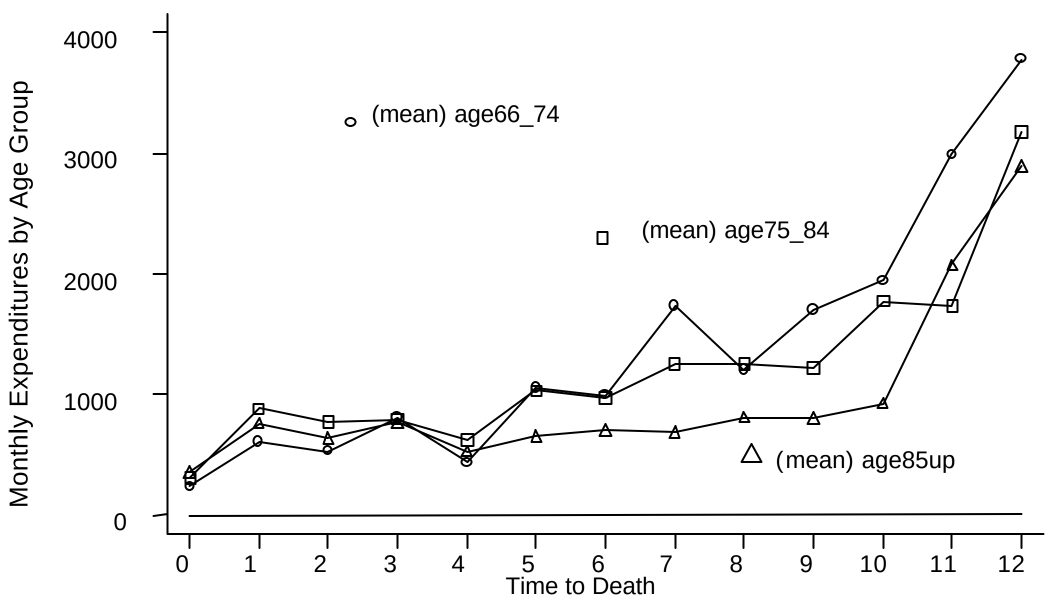

Figure 1b: Monthly Medicare Expenditures by Time to Death and by Age Group

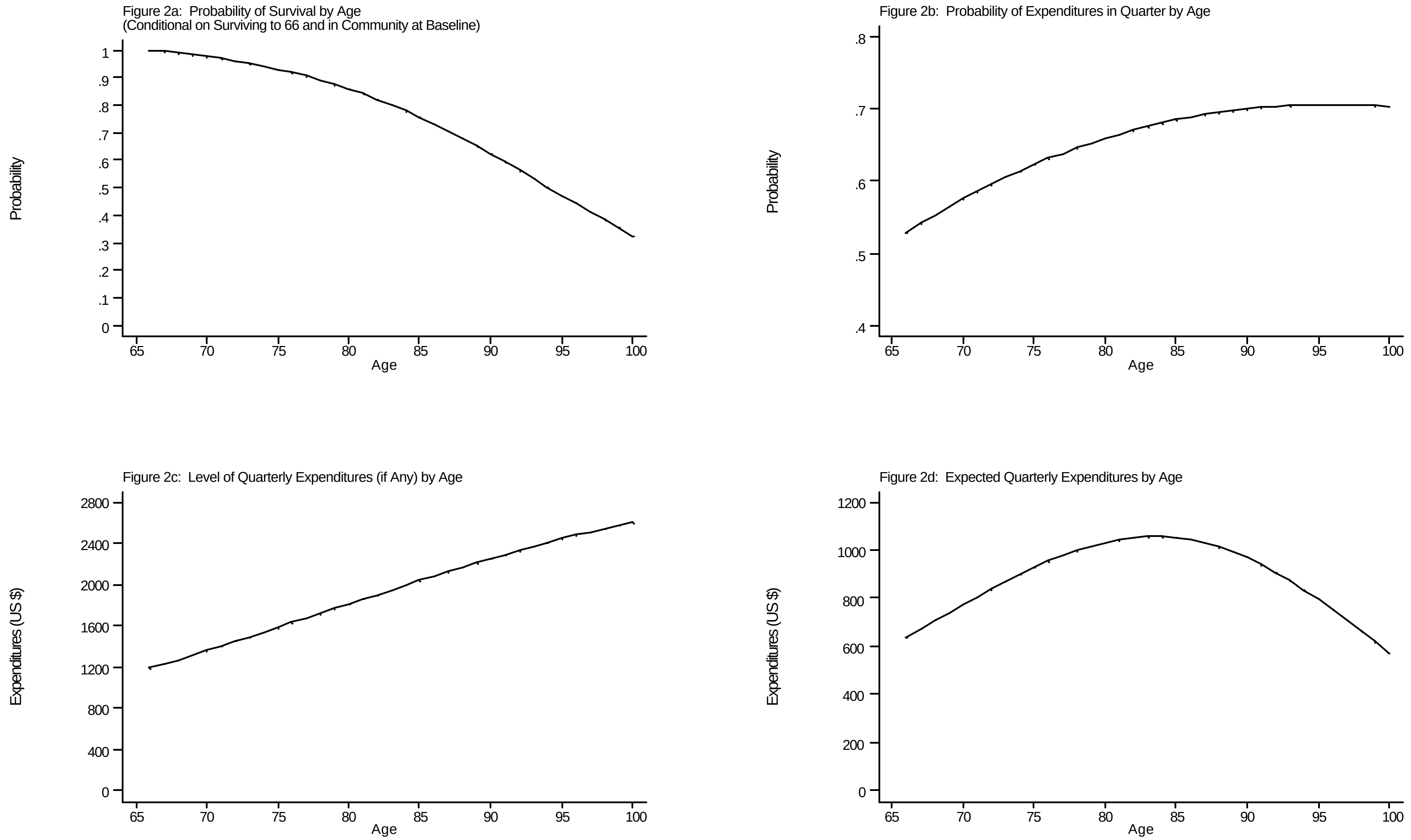

Figure 2: Naive Model Simulated Components

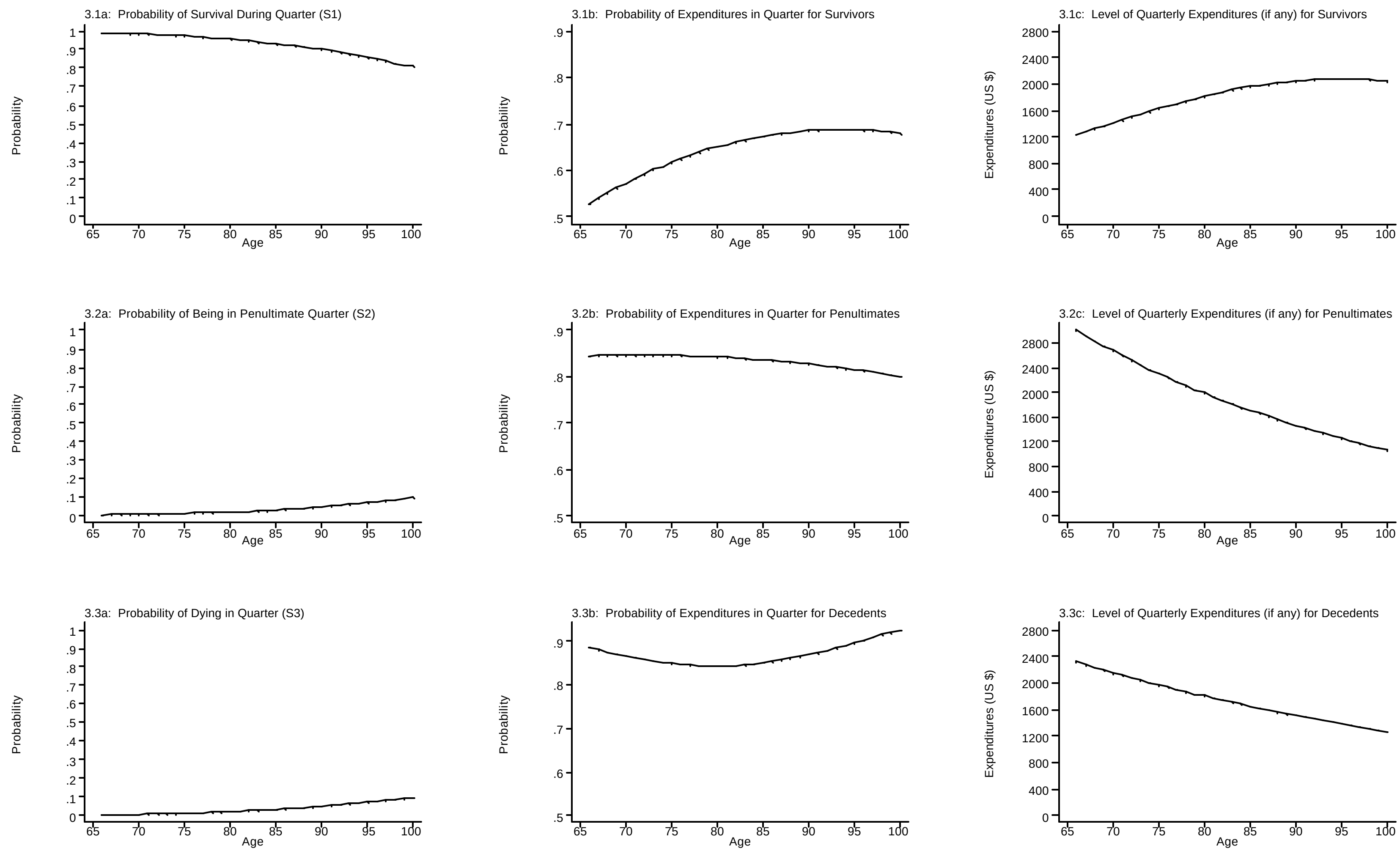

Figure 3: Full Model Simulated Components

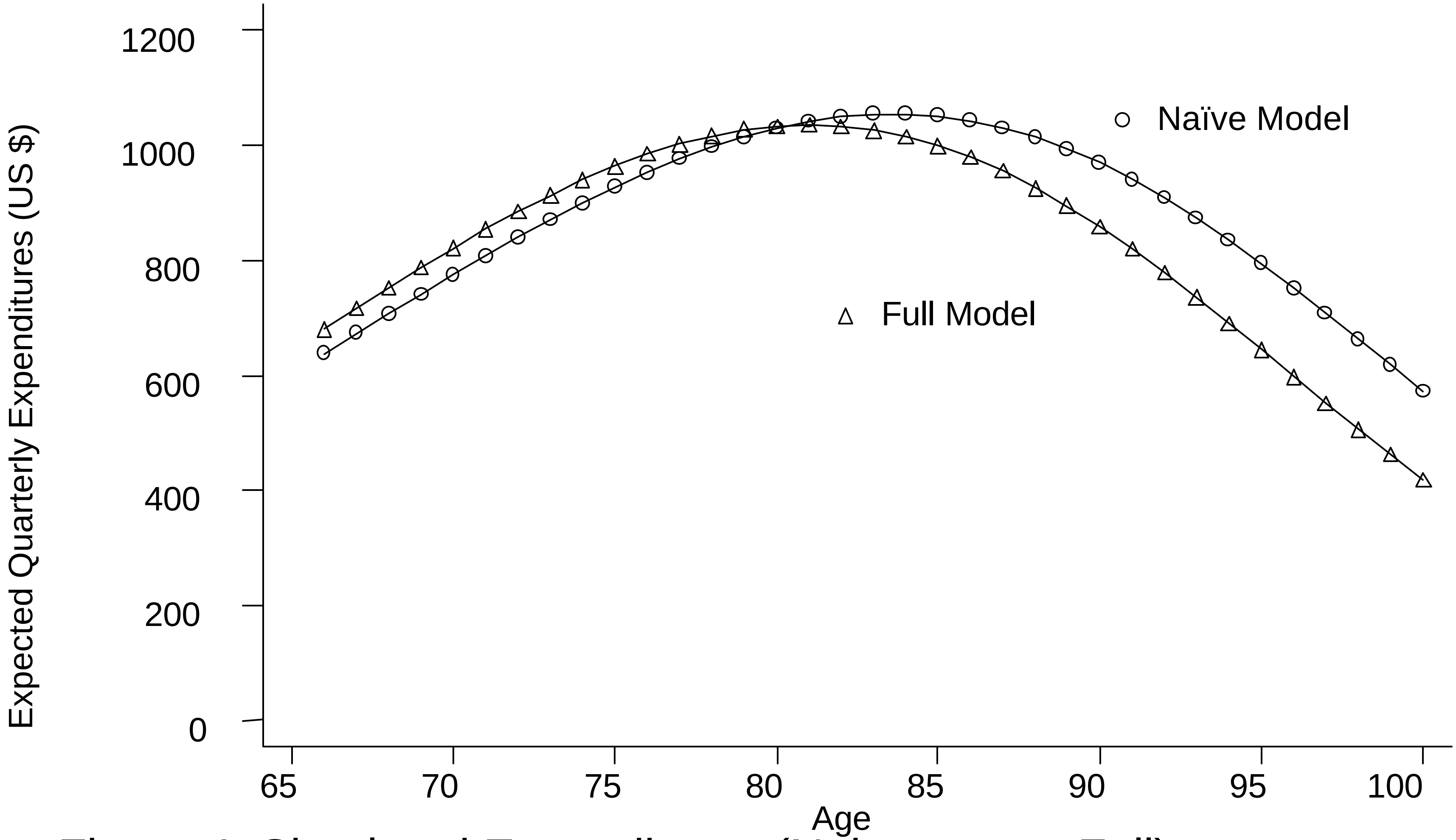

Figure 4: Simulated Expenditures (Naive versus Full)

Table 1: Descriptive Statistics (N=32,174)

Variable	Mean	Std. Dev.	Minimum	Maximum
Age	77.667	7.445	66	102
Modified Age	0.000	7.445	-11.667	24.333
Modified Age Squared	55.424	60.238	0.111	592.118
Male	0.510	0.500	0	1
Black	0.057	0.233	0	1
High School Education or Higher	0.584	0.493	0	1
Married	0.549	0.498	0	1
Supplemental Insurance	0.715	0.451	0	1
150% of Poverty Line or Higher	0.180	0.384	0	1
100-149% of Poverty Line	0.657	0.475	0	1
No Alcohol Consumption	0.587	0.492	0	1
Former Smoker	0.373	0.484	0	1
Current Smoker	0.110	0.313	0	1
Rural Area	0.283	0.450	0	1
Midwest	0.280	0.449	0	1
Northeast	0.262	0.440	0	1
South	0.126	0.332	0	1
South West	0.058	0.234	0	1
West	0.132	0.338	0	1
Physicians per 1K Population	2.028	1.227	0	9.450
Hospital Beds per 10K Population	5.036	2.743	0	21.240
In Penultimate Quarter	0.021	0.142	0	1
In Death Quarter	0.021	0.142	0	1

Note: Modified age has the mean of age subtracted from age.

Table 2: Estimation Results for Naive Model

Dependent Variable	Die During Quarter				Any Expenditure				Quarterly Expenditures if Any		
	Coefficient		Std. Error		Coefficient		Std. Error		Coefficient		Std. Error
Modified Age	0.1002	***	0.0081		0.0195	***	0.0023		0.0268	***	0.0037
Modified Age Squared	-0.0009		0.0007		-0.0005	**	0.0003		-0.0004		0.0004
Male	0.4493	***	0.0973		-0.1223	***	0.0338		0.1958	***	0.0534
Black	0.0048		0.1771		-0.0892		0.0652		-0.0526		0.1056
High School Education or Higher	-0.1602	*	0.0887		0.0197		0.0333		-0.0320		0.0524
Married	-0.0895		0.0936		0.0043	*	0.0341		0.0063		0.0538
Supplemental Insurance	-0.0056		0.0909		0.0906	***	0.0342		-0.0373		0.0545
150% of Poverty Line or Higher	0.0781		0.1264		-0.1475	***	0.0507		-0.0770		0.0795
100-149% of Poverty Line	-0.0015		0.1183		-0.0510		0.0463		-0.2045	***	0.0718
No Alcohol Consumption	0.2345	***	0.0898		0.1112	***	0.0323		0.1957	***	0.0512
Former Smoker	0.3010	***	0.0912		0.1719	***	0.0335		0.1195	**	0.0524
Current Smoker	0.5093	***	0.1419		-0.0015		0.0494		0.1149		0.0800
Rural	-0.1238		0.1101		-0.1216	***	0.0405		-0.2809	***	0.0633
Midwest	-0.1038		0.1423		-0.3217	***	0.0528		-0.4812	***	0.0802
Northeast	-0.0479		0.1429		-0.2256	***	0.0530		-0.2410	***	0.0798
South	-0.1446		0.1661		-0.3127	***	0.0602		-0.3587	***	0.0928
South West	-0.1014		0.2060		-0.4423	***	0.0749		-0.5302	***	0.1192
West	-0.0334		0.1538		-0.3098	***	0.0579		-0.2354	***	0.0884
Physicians per 1K Population	0.0299		0.0456		0.0127		0.0172		0.0238		0.0270
Hospital Beds per 10K Population	-0.0075		0.0168		-0.0150	**	0.0063		-0.0244	**	0.0101
Constant	-4.4358	***	0.2250		0.6323	***	0.0847		6.0238	***	0.1304
Stata Estimation Technique	Logit				Xtprobit				Xtreg		
Observations	32,174				32,174				32,174		
Persons Represented	2,862				2,862				2,862		
Wald (Chi2(20))					215.52				207.58		
Log Likelihood	-3043.52										

*** p<0.01 ** p<0.05 *p<0.01

Table 3a: Estimation Results for Full Model - Survivors (All Results Conditional Upon Survival to Quarter)

Dependent Variable	Survivors (S1)	Any Expenditure (Survivors)			Survivors Quarterly Expenditures if Any		
		Coefficient		Std. Error	Coefficient		Std. Error
Modified Age		0.0178	***	0.0023	0.0214	***	0.0036
Modified Age Squared		-0.0006	**	0.0003	-0.0006		0.0004
Male		-0.1321	***	0.0341	0.1540	***	0.0522
Black		-0.1053		0.0658	-0.1000		0.1034
High School Education or Higher		0.0142		0.0337	-0.0188		0.0512
Married		0.0050		0.0344	0.0124		0.0526
Supplemental Insurance		0.0875	**	0.0346	-0.0269		0.0533
150% of Poverty Line or Higher		-0.1430	***	0.0512	-0.0995		0.0779
100-149% of Poverty Line		-0.0438		0.0468	-0.2350	***	0.0702
No Alcohol Consumption		0.1120	***	0.0326	0.1693	***	0.0499
Former Smoker		0.1611	***	0.0339	0.0946	**	0.0512
Current Smoker		-0.0168		0.0497	0.0845		0.0780
Rural		-0.1198	***	0.0410	-0.2563	***	0.0618
Midwest		-0.3227		0.0534	-0.4961	***	0.0784
Northeast		-0.2131		0.0535	-0.2364	***	0.0779
South		-0.3032		0.0609	-0.3540	***	0.0907
South West		-0.4494		0.0757	-0.5151	***	0.1169
West		-0.3058		0.0585	-0.2252	***	0.0865
Physicians per 1K Population		0.0116		0.0174	0.0270		0.0263
Hospital Beds per 10K Population		-0.0147	**	0.0063	-0.0244	**	0.0098
Constant		0.6184	***	0.0857	6.0021	***	0.1277
Stata Estimation Technique	None	Xtprobit			Xtreg		
Observations	(Calculated	30,842			19,126		
Persons Represented	from constraint)	2,837			2,679		
Wald (Chi2(20))		198.82			171.48		
Log Likelihood							
R-squared							
*** p<0.01 ** p<0.05 *p<0.01							

Table 3b: Estimation Results for Full Model - Penultimates				(All Results Conditional Upon Survival to Quarter)							
Dependent Variable	In Penultimate Quarter (S2)				Any Expenditure (Penultimates)				Penultimate s		
									Quarterly Expenditures if Any		
	Coefficient		Std. Error		Coefficient		Std. Error		Coefficient		Std. Error
Modified Age	0.1019	***	0.0082		-0.0028		0.0128		-0.0298		0.0190
Modified Age Squared	-0.0009		0.0007		-0.0002		0.0009		0.0000		0.0015
Male	0.4605	***	0.0975		-0.2157		0.1428		0.2371		0.2200
Black	0.0027		0.1772		0.3063		0.2842		0.3715		0.3957
High School Education or Higher	-0.1650	*	0.0889		0.1763		0.1308		-0.0342		0.2032
Married	-0.1005		0.0939		0.1877		0.1367		0.2454		0.2106
Supplemental Insurance	-0.0118		0.0910		0.2798	**	0.1338		-0.0496		0.2122
150% of Poverty Line or Higher	0.0772		0.1265		-0.3166	*	0.1927		-0.1714		0.2838
100-149% of Poverty Line	-0.0065		0.1185		-0.4081	**	0.1823		0.1341		0.2626
No Alcohol Consumption	0.2345	***	0.0899		-0.0296		0.1355		0.4121	**	0.2076
Former Smoker	0.3102	***	0.0915		0.1230		0.1354		-0.2825		0.2048
Current Smoker	0.5179	***	0.1420		-0.1557		0.2129		-0.4452		0.3372
Rural	-0.1221		0.1104		-0.1037		0.1589		-0.5739	**	0.2508
Midwest	-0.0917		0.1429		-0.1362		0.2205		-0.4070		0.3270
Northeast	-0.0347		0.1435		-0.1928		0.2267		0.2387		0.3342
South	-0.1346		0.1666		-0.2729		0.2574		0.1681		0.3803
South West	-0.0938		0.2065		-0.3401		0.2995		-0.5946		0.4676
West	-0.0134		0.1543		-0.0648		0.2393		0.0159		0.3517
Physicians per 1K Population	0.0268		0.0458		-0.0128		0.0650		-0.0419		0.1056
Hospital Beds per 10K Population	-0.0067		0.0167		-0.0236		0.0250		-0.0447		0.0378
Constant	-4.4207	***	0.2257		1.3715	***	0.3409		7.2221	***	0.5147
Stata Estimation Technique	Logit				Probit				Reg (OLS)		
Observations	31,507				665				549		
Persons Represented	2,860				665				549		
Wald (Chi2(20))											
Log Likelihood	-3014.13				-297.77						
R-squared									0.056		
*** p<0.01 ** p<0.05 *p<0.01											

Table 3c: Estimation Results for Full Model - Decedents											
Dependent Variable	Die During Quarter (S3)				Any Expenditure (Decedents)				Decedents		
									Quarterly Expenditures if Any		
	Coefficient		Std. Error		Coefficient		Std. Error		Coefficient		Std. Error
Modified Age	0.1002	***	0.0081		-0.0046		0.0134		-0.0178		0.0179
Modified Age Squared	-0.0009		0.0007		0.0011		0.0011		0.0000		0.0014
Male	0.4493	***	0.0973		-0.1951		0.1480		0.1487		0.2080
Black	0.0048		0.1771		0.1552		0.2713		0.3884		0.3911
High School Education or Higher	-0.1602	*	0.0887		0.2600	*	0.1369		0.0975		0.1918
Married	-0.0895		0.0936		0.1621		0.1395		0.1201		0.2002
Supplemental Insurance	-0.0056		0.0909		0.0880		0.1404		-0.3184		0.1997
150% of Poverty Line or Higher	0.0781		0.1264		-0.1692		0.1971		0.0501		0.2705
100-149% of Poverty Line	-0.0015		0.1183		-0.2534		0.1842		0.0504		0.2524
No Alcohol Consumption	0.2345	***	0.0898		0.0034		0.1384		0.3509	*	0.1982
Former Smoker	0.3010	***	0.0912		-0.0174		0.1402		0.2169		0.1950
Current Smoker	0.5093	***	0.1419		-0.2409		0.2185		-0.1141		0.3213
Rural	-0.1238		0.1101		0.2185		0.1712		-0.3673		0.2320
Midwest	-0.1038		0.1423		0.0619		0.2300		-0.0748		0.3080
Northeast	-0.0479		0.1429		-0.2056		0.2348		0.1447		0.3154
South	-0.1446		0.1661		-0.3024		0.2578		-0.0829		0.3694
South West	-0.1014		0.2060		-0.1943		0.3044		-0.2816		0.4467
West	-0.0334		0.1538		0.0019		0.2514		-0.2460		0.3316
Physicians per 1K Population	0.0299		0.0456		0.1143		0.0727		-0.0650		0.0980
Hospital Beds per 10K Population	-0.0075		0.0168		-0.0418	*	0.0241		0.0082		0.0398
Constant	-4.4358	***	0.2250		1.0700	***	0.3496		7.3917	***	0.4865
Stata Estimation Technique	Logit				Probit				Reg (OLS)		
Observations	32,174				667				566		
Persons Represented	2,862				667				566		
Wald (Chi2(20))											
Log Likelihood	-3043.52				-273.88						
R-squared									0.0307		
*** p<0.01 ** p<0.05 *p<0.01											

Table 4: Simulated Quarterly Expenditures (US $) by Age: Naïve versus Full Model
(Conditional upon survival to age 66)

Age	Naïve	Full	Percentage Difference (Naïve versus Full)	Cumulative Difference
66	$638.56	$680.19	-6.12%	-$41.63
67	$672.86	$716.53	-6.09%	-$85.29
68	$707.12	$752.37	-6.01%	-$130.54
69	$741.14	$787.42	-5.88%	-$176.82
70	$774.70	$821.40	-5.69%	-$223.52
71	$807.60	$854.01	-5.43%	-$269.93
72	$839.58	$884.93	-5.12%	-$315.28
73	$870.41	$913.85	-4.75%	-$358.73
74	$899.82	$940.45	-4.32%	-$399.36
75	$927.53	$964.41	-3.82%	-$436.24
76	$953.27	$985.40	-3.26%	-$468.36
77	$976.76	$1,003.13	-2.63%	-$494.74
78	$997.71	$1,017.32	-1.93%	-$514.34
79	$1,015.84	$1,027.69	-1.15%	-$526.20
80	$1,030.87	$1,034.02	-0.30%	-$529.34
81	$1,042.56	$1,036.11	0.62%	-$522.90
82	$1,050.66	$1,033.83	1.63%	-$506.07
83	$1,054.96	$1,027.05	2.72%	-$478.16
84	$1,055.28	$1,015.75	3.89%	-$438.63
85	$1,051.49	$999.95	5.15%	-$387.09
86	$1,043.50	$979.73	6.51%	-$323.33
87	$1,031.25	$955.25	7.96%	-$247.33
88	$1,014.77	$926.72	9.50%	-$159.28
89	$994.13	$894.44	11.15%	-$59.58
90	$969.48	$858.75	12.89%	$51.15
91	$941.01	$820.06	14.75%	$172.10
92	$908.98	$778.82	16.71%	$302.26
93	$873.72	$735.54	18.79%	$440.44
94	$835.60	$690.71	20.98%	$585.33
95	$795.04	$644.89	23.28%	$735.49
96	$752.50	$598.60	25.71%	$889.39
97	$708.46	$552.35	28.26%	$1,045.50
98	$663.41	$506.66	30.94%	$1,202.24
99	$617.85	$461.97	33.74%	$1,358.11
100	$572.26	$418.71	36.67%	$1,511.67