Trajectories of Disability and Mortality Among the U.S. Elderly Population: Evidence from the 1984–1999 NLTCS

Eric Stallard, A.S.A., M.A.A.A, F.C.A.
Center for Demographic Studies

Presented at The Living to 100 and Beyond Symposium
Sponsored by the Society of Actuaries

Orlando, Fla.

January 12–14, 2005

Copyright 2005 by the Society of Actuaries.
All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

Abstract

This paper employs a longitudinal form of the Grade of Membership (GoM) model to specify and estimate a multivariate model of the trajectories of disability and mortality among longitudinally followed elderly respondents to the National Long-Term-Care Survey (NLTCS) of 1984, 1989, 1994, and 1999. A distinct trajectory is constructed for each individual respondent to the survey. The trajectories describe the progressive declines over time in physical and cognitive functioning among a nationally representative sample of the U.S. elderly population.

The model is structured to simultaneously represent the essential features of the fixed frailty model (Vaupel et al. 1979) and the model of linearly declining vitality (Strehler and Mildvan 1960). Unlike those models, however, the longitudinal GoM model is designed for easy and direct application to existing longitudinal data sets.

The measurement space in the NLTCS application includes from one to four sets of repeated measures for each survey respondent on 95 independent variables characterizing the nature and intensity of limitations in activities of daily living, instrumental activities of daily living, physical functioning, and cognitive functioning, as well as indicators of behavioral characteristics, medical conditions, subjective health, age, race, sex, and institutional status.

The application shows that the model can be fitted to existing data and that the results are interpretable as generalizations of fixed frailty with linearly declining vitality.

Introduction

Actuaries, demographers, gerontologists, and biomedical researchers have contributed significantly to our understanding of survival at advanced ages. Further advances are possible using the extensive morbidity, disability, and mortality data currently being collected on longitudinally followed populations. Full realization of the analytic potential of such data will require new models and methods.

To date, much of the research on survival at advanced ages has been performed using the life table and related methodologies. Methods for generalizing the life table to represent the effects of observed and unobserved covariates have been proposed. Such methods are required for fully representing the effects of large numbers of time-varying covariates in sample data from longitudinally followed populations.

This paper generalizes the life table to include the effects of observed and unobserved time-varying health-related covariates using a longitudinal form of the

Grade of Membership (GoM) model to specify and estimate a multivariate model of the trajectories of disability and mortality among elderly respondents to the National Long-Term-Care Survey (NLTCS) of 1984, 1989, 1994, and 1999. This model is shown to be a natural extension of currently available methods and one that is consistent with recent research findings on individual differences in mortality and morbidity risks and the impact of those differences on survival at advanced ages.

The paper contains five sections:

- The Background section provides information on existing life-table methods and recent research results that are used to motivate the development of the longitudinal form of the GoM model.

- The Data section provides information on the NLTCS and related administrative files from Medicare that are used as the inputs to the longitudinal GoM model.

- The Methods section describes the longitudinal GoM model, the methods used for estimating its parameters, and the application of the model to longitudinal data such as contained in the NLTCS.

- The Results section presents the parameter estimates and resulting generalizations of the life table to represent trajectories of individual health histories. Validation of the model is based on internal consistency as well as comparisons with published life-table models and other data.

- The Discussion section considers the implications of the results in the context of existing models for analyzing and forecasting future disability and mortality patterns in the United States.

Background

Life-Table Parameters

A life table is a tabular array of age-specific survival probabilities and associated derived annual and cumulative measures (Chiang 1984). Three fundamental parameters are:

1. q_a = the probability of death within one year after attaining age a
2. p_a = the probability of surviving at least one year after attaining age a

3. μ_a = the age-specific death rate, hazard rate, or force of mortality for the corresponding one-year age interval from a to $a + 1$.

The three parameters are functionally related. Parameter (2) is the complement of parameter (1)—that is, the sum of the two probabilities is exactly 1.0. Parameter (2) can be logarithmically transformed, with a sign change, to produce parameter (3) (assuming that μ_a is piecewise constant; Manton and Stallard 1984). Moreover, $\mu_a \geq q_a$ for every age a, with approximate equality for values of μ_a below 0.05, which applies to current U.S. mortality rates up to about age 75–80 years.

Gompterz's Law

The force of mortality exhibits exponential growth above about age 20. This was discovered by Gompertz (1825), who published the famous "law of mortality" that now bears his name. Under Gompterz's law, $\mu_a = \alpha e^{\beta a}$, with proportionality constant α and growth constant β.

Gompterz's law has been confirmed in numerous studies of human (and nonhuman) populations (Wetterstrand 1981; Olshansky and Carnes 1997). Two limitations are that the Gompertz "constants" (α and β) differ from one population to another and over calendar time and cohort within a given population, and moreover, the parameter values for α and β are negatively correlated when calculated for populations with a broad range of mortality conditions (Strehler and Mildvan 1960; Gavrilov and Gavrilova 2001).

Beginning at about 80–90 years of age (for humans; at various postreproductive ages for nonhumans), the increases in observed mortality rates are less than that predicted by Gompertz's law (Olshansky and Carnes 1997; Vaupel et al. 1998), indicating that Gompertz's law should be viewed as a useful approximation rather than an immutable fact.

Two analytic models have been proposed that yield insight into the dynamics of mortality at advanced ages: one based on linearly declining vitality (Strehler and Mildvan 1960) and the other on fixed frailty (Vaupel et al. 1979).

Declining Vitality

Strehler and Mildvan (1960) proposed linearly declining vitality as the underlying physiological mechanism that explains the exponential increase found in Gompertz's law. They defined *vitality* as

> the capacity of an individual organism to stay alive, as measured by an appropriately weighted average of the maximum rate of work output (power output) less the basal power output of all of the functional modalities contributing to survival in the normal environment. (p. 15)

They noted that vitality includes:

> Only the reserve capacity of an organism to do work in overcoming challenges to its existence. It does not include the work it must do to maintain itself in the absence of challenge. The latter, appropriately weighted, is the basal work rate. (p. 18)

To validate these assumptions, they analyzed rates of decline in reserve capacities for eight physiologic functions (nerve conduction velocity, basal metabolic rate, maximal breathing capacity, standard cell water, standard renal plasma flow, vital capacity, standard glomerular filtration rate, and a cardiac index) and found that all of the declines were linear. Assuming that the force of mortality followed Gompertz's law and that the fraction of challenges that are lethal was a negative exponential function of vitality, Strehler and Mildvan (1960, p. 17) found that the implied rate of decline in vitality was "in reasonable agreement" with the observed declines in reserve capacities.

Strehler and Mildvan's (1960) model was developed and tested at the population level. This limited their results in two ways:

> They did not validate their model using individual-level data on the relationships between measures of physiologic reserve capacities and subsequent mortality rates.
>
> They did not account for the impact of individual differences in initial vitality or in the rates of decline in vitality.

Fixed Frailty

Vaupel et al. (1979) developed life-table methods to study the impact of unobserved but persistent individual differences in mortality risks. These differences were represented through the effects of a multiplicative factor termed *frailty* that operated, by assumption, as a fixed multiplier on a standard schedule of age-specific forces of mortality faced by individuals in each birth cohort. Without loss of generality, average frailty was defined to have the value 1 at the initial age of observation (e.g., at birth or any other age at which cohort follow-up began). Thus, using z_i to denote the

frailty value for the *i*-th individual, the model can be specified as $\mu_a(z_i) = z_i \cdot \mu_a(1)$, where $\mu_a(1)$ is the standard force of mortality at age *a*. Because individual frailty values were fixed for life, the distribution of frailty was assumed to be affected only by the selective effects of differential mortality at the different levels of frailty.

The meaning of frailty employed by Vaupel et al. (1979) differed substantially from the meaning of the same term that is being increasingly used in the gerontological literature to characterize a clinical syndrome among elderly persons that includes such characteristics as weight loss, exhaustion, grip strength, walking speed, and physical activity (e.g., see Fried et al. 2001, 2004; Bortz 2002). This use of the term may be closer to the concept of low vitality, as implied by Strehler and Mildvan's (1960) model of declining vitality. The concept of frailty proposed by Vaupel et al. was one in which frailty was a fixed characteristic of an individual; this was synonymous with a concept of relative susceptibility to death.

Part of the motivation for the fixed frailty model was to determine if the slowing of the rate of increase in the force of mortality observed at the oldest ages in a population was an artifact of differential mortality selection on a population in which the forces of mortality faced by individuals continued to increase according to Gompertz's law.

Vaupel et al. (1979) assumed that frailty was gamma distributed. Hougaard (1984) proposed that the inverse Gaussian distribution be considered as an alternative to the gamma. Manton et al. (1986) fitted gamma and inverse Gaussian distributions to U.S. death rates at ages 65–94, using both Gompertz and Weibull models for the standard schedule of age-specific forces of mortality faced by individuals in each birth cohort. The estimates of the standard deviation of frailty were in the range 0.3–0.8 and were larger for females than males, larger for the Gompertz than the Weibull model, and larger for the inverse Gaussian than the gamma distribution. However, because the mathematics of the mortality selection process ensure that the standard deviations were fixed over age for the gamma but declined over age for the inverse Gaussian, the standard deviations for the two distributions converged to equality at age 89 for males and age 93 for females, after which the standard deviations crossed over.

One major limitation of the analysis in Manton et al. (1986) was the need to assume a parametric form such as the Gompertz or Weibull for the standard schedule of age-specific forces of mortality faced by individuals in each birth cohort. Because Strehler and Mildvan (1960, p. 19) argued that their theoretical model applied to individuals, the assumption of a Gompertz function for individual forces of mortality in

Manton et al. was consistent with their model. If that assumption was wrong, however, then the estimate of the standard deviation of frailty could be seriously in error.

To test this assumption, one must apply the model to data for which the functional form of the individual forces of mortality is not restricted to any specific parametric family. Such data were available to Iachine et al. (1998), who applied a correlated gamma-frailty model to survival data on Danish, Swedish, and Finnish twins. They found that the standard schedules of age-specific forces of mortality faced by individuals in each birth cohort increased substantially faster than the Gompertz function and that the standard deviation of frailty was about 1.5, which was about 3 times larger than the comparable estimates in Manton et al. (1986; i.e., 0.46 and 0.54, respectively, for males and females under the gamma-mixed Gompertz formulation).

Iachine et al.'s (1998) results clearly contradicted Strehler and Mildvan's (1960) assumption that their theoretical model applied to individuals. Iachine et al.'s results also affected the interpretation of the analysis of Manton et al. (1986), given that part of the motivation for the fixed frailty model was to determine if the slowing of the rate of increase in the force of mortality at the oldest ages in a population compared to the rate of increase predicted under Gompertz's law was an artifact of mortality selection. As an alternative to the Gompertz function, Manton et al. considered the Weibull function as a model of the forces of mortality faced by individuals. However, when fitted to the same age-specific data, the Weibull function rose slower than the Gompertz function at older ages, so that the Weibull model was likewise contradicted by Iachine et al.'s results. Manton et al. did not consider any models in which the individual forces of mortality rose faster than the Gompertz function, in part because the need for such models was not recognized at that time.

Iachine et al. (1998) found that (1) the variability in individual frailty was substantially larger than expected and (2) the individual forces of mortality rose substantially faster than expected under the assumption that the Gompertz function was an accurate representation of the age trajectories of individual forces of mortality. These findings are relevant to our goal of accurately modeling trajectories of disability and mortality for individual respondents to the NLTCS.

Genetic and Nongenetic Factors

Because their analyses were based on the survival experience of monozygotic and dizygotic twins, Iachine et al. (1998) were able to partition the variability in individual frailty into genetic and nongenetic components. Specifically, they reported country-specific sex-specific estimates of "narrow-sense" heritability in frailty

(reflecting the fraction of the variance of frailty associated with additive genetic effects) in the range 0.36–0.60 and 0.37–0.54 and pooled estimates of 0.57 and 0.51, respectively, for males and females.

Combined with other analyses by the same researchers (e.g., McGue et al. 1993; Yashin and Iachine 1995; Yashin, Iachine, and Harris 1999; De Benedictis et al. 2001), these results indicated that, for both sexes, about 50 percent of the variability in individual frailty was genetic.

Conversely, these results indicated that about 50 percent of the variability in individual frailty was due to nongenetic factors.

Because individual frailty values are assumed to be fixed for life, or at least approximately so, in the fixed frailty model, the finding that nongenetic factors account for about half of the variability in frailty means that the source and stability of such nongenetic factors are relevant.

Stability is relevant because the assumption that frailty is effectively fixed by early adult ages is necessary for the effects of mortality selection to become manifest at older ages. That is, if individual differences in frailty do not persist, then the excess risks faced by any given individual at one set of times would tend to be canceled out by reduced risks at later times. Over time, no cumulative advantage or disadvantage would be experienced by such individuals; hence, the individual and population forces of mortality would be the same, on average, and the impact of nonpersistent individual differences in frailty would be trivial.

The sources of the nongenetic factors are relevant to our understanding of these factors, their stability over time, the persistency of their effects, and the potential for interventions that could mitigate their deleterious effects.

Important insight into the nature of the nongenetic factors was provided by McGinnis and Foege (1993), who introduced the concept of "actual causes of death," which are nongenetic modifiable lifestyle and risk factor behaviors such as smoking, poor nutrition/physical inactivity, alcohol use, microbial agents, toxic agents, motor vehicle accidents, firearms, sexual behaviors, and illicit drug use that were quantitatively associated with the standard concept of "underlying causes of death" through published relative risk estimates. According to these authors, approximately 50 percent of all deaths in the United States in 1990 were attributable to the nine factors studied.

Using similar methods, Mokdad et al. (2004) updated that study and found that 48 percent of all deaths in the United States in 2000 could be attributed to the same nine factors.

Mokdad et al. (2004) identified several limitations to their methodology, including the use of relative risk measures from independent published analyses and the lack of information and controls for the effects of high blood pressure and high serum cholesterol. Moreover, they provided no information on morbidity and disability, which are important to the current analysis. Despite these limitations, the results of their analyses were consistent with the expectation from the twin studies that about 50 percent of the variation in frailty, and hence in mortality, was due to nongenetic factors. Also, the cumulative nature of the damage associated with the various lifestyle and risk factor behaviors could support an assumption that the resulting mortality differentials were stable over time.

Life Tables with Covariates

Hazard-rate (force of mortality) regression models with observed covariates are often specified using Cox's (1972) regression model. This model is similar to the fixed frailty model except that an exponential function of the covariates operates as a fixed multiplier on the age-specific standard force of mortality. For example, using $\mathbf{x}_i$ to denote the vector of covariates for the i-th individual, the model can be specified as $\mu_a(\mathbf{x}_i) = \exp(\mathbf{x}_i'\mathbf{b}) \cdot \mu_a(\mathbf{0})$, where $\mu_a(\mathbf{0})$ is the standard force of mortality at age a, that is, resulting from the condition $\mathbf{x}_i = \mathbf{0}$. Moreover, because $\mu_a(\mathbf{0})$ does not appear in the partial likelihood equations used to estimate the regression parameters, it is not necessary to specify the functional form of the standard force of mortality. Nonparametric estimates of $\mu_a(\mathbf{0})$ can be generated after the regression parameters have been estimated.

Various generalizations of Cox's (1972) regression model have been proposed to allow estimation and testing of the hazard-rate regression coefficients and standard forces of mortality with time-varying covariates (e.g., see Therneau and Grambsch 2000). A common approach is to conduct estimation and testing conditional on the observed sequences of covariate values obtained from all individuals in the sample. This approach neither requires, nor uses, an explicit model of the processes governing the changes in the covariate values.

When the changes in the covariates are a major focus of the analysis, then an explicit model of the processes governing the temporal changes in the covariate values

is required. In this case either the Cox regression model must be embedded within a more general covariate model, or an alternative model must be developed.

Manton et al. (1994a) presented a two-component Gaussian stochastic process model describing (1) the changes in the observable covariates over age and (2) the impact of the age-varying covariates on the individual forces of mortality. Implementation of the model followed the specifications given in Woodbury and Manton (1977) and Tolley and Manton (1991), viz., linear dynamics of the covariates and a quadratic form for the force of mortality.

The changes in the vector of covariate values, $\mathbf{x}_{ia}$, for the i-th individual from age a to $a + 1$ were modeled as a first-order autoregressive process, as follows:

$$\mathbf{x}_{i(a+1)} = \mathbf{u}_a + \mathbf{A}\mathbf{x}_{ia} + \mathbf{e}_{i(a+1)},$$

where $\mathbf{u}_a$ was a vector of age-specific constant changes, $\mathbf{A}$ was a regression matrix representing the effects of $\mathbf{x}_{ia}$ on $\mathbf{x}_{i(a+1)}$, and $\mathbf{e}_{i(a+1)}$ was a vector of normally distributed residual values (i.e., errors or innovations) at age $a + 1$. Estimation of $\mathbf{u}_a$ and $\mathbf{A}$ was generally conducted via ordinary least squares procedures. One exception was the application of the model to GoM scores from the NLTCS, for which estimation of $\mathbf{A}$ (actually $\mathbf{A}_a$) was conducted via the minimum entropy algorithm described in Manton et al. (1992b, pp. 326–329).

The hazard function was modeled as an age-specific quadratic function of the current covariates:

$$\mu_a(\mathbf{x}_{ia}) = \mu_a(\mathbf{0}) + \mathbf{b}'_a\mathbf{x}_{ia} + \tfrac{1}{2}\mathbf{x}'_{ia}\mathbf{B}_a\mathbf{x}_{ia},$$

with parameters $\mu_a(\mathbf{0})$, $\mathbf{b}_a$, and $\mathbf{B}_a$, where $\mu_a(\mathbf{0})$ was the standard force of mortality at age a, that is, resulting from the condition $\mathbf{x}_{ia} = \mathbf{0}$.

To facilitate estimation of the quadratic hazard, it was necessary to impose two types of constraints. One set of constraints was needed to ensure that the quadratic form was non-negative definite. The second set was needed to reduce the number of parameters to a manageable number. This was done by replacing each age-specific term in the quadratic hazard with an exponential function of age, as follows:

$$\begin{aligned} \mu_a(\mathbf{0}) &= \mu(\mathbf{0})\, e^{\beta a}, \\ \mathbf{b}_a &= \mathbf{b}\, e^{\beta a}, \\ \mathbf{B}_a &= \mathbf{B}\, e^{\beta a}, \end{aligned}$$

where the exponential growth constant, β, was the same for all three types of parameters (see Manton et al. 1992a). The expression for $\mu_a(\mathbf{0})$ was identical to Gompertz's formula with the substitution $\alpha = \mu(\mathbf{0})$. Substitution for $\mu_a(\mathbf{0})$, $\mathbf{b}_a$, and $\mathbf{B}_a$ in the expression for $\mu_a(\mathbf{x}_{ia})$ yielded

$$\mu_a(\mathbf{x}_{ia}) = \left(\mu(\mathbf{0}) + \mathbf{b}'\mathbf{x}_{ia} + \tfrac{1}{2}\mathbf{x}'_{ia}\mathbf{B}\mathbf{x}_{ia}\right)e^{\beta a},$$

a multidimensional generalization of Gompertz's law with the substitution $\alpha = \mu(\mathbf{0}) + \mathbf{b}'\mathbf{x}_{ia} + \frac{1}{2}\mathbf{x}'_{ia}\mathbf{B}\mathbf{x}_{ia}$ reflecting multiplicative and interaction effects of observable age-varying covariates. The resulting expression allowed the individual forces of mortality to rise as fast as, faster than, or slower than the Gompertz function, given that the individual forces of mortality were controlled by (1) changes in $\mathbf{x}_{ia}$, which can move the individual up or down the sides of the quadratic function, and (2) the exponential growth constant, β, which increased the height of the quadratic function by a constant factor at each age.

Manton et al. (1994a) found that the exponential growth constant, β, was reduced substantially when covariates were included in the model. Moreover, in comparing the performance of functional status measures from the NLTCS with cardiovascular disease risk factors from the Framingham Heart Study, the authors found that the functional status measures accounted for higher proportions of the variance in mortality risks (79–87 percent vs. 70–71 percent) and produced substantially greater reductions in the exponential growth constant, β (35–48 percent vs. 14–19 percent), than did the cardiovascular disease risk factors.

As the exponential growth constant declined toward 0, one would have greater confidence that the changes in the covariates accounted for changes in mortality over age, that is, because the movement of the covariates to high risk values would be the only mechanism in the model that could account for the increase in mortality risks among affected individuals.

One limitation of the two-component Gaussian stochastic process model was that the changes in the observed covariates were represented as an autoregressive process with individual covariate values tending to regress toward age-specific cohort mean values rather than to individually estimated covariate trajectories. This limitation increases in importance if the analysis is directed toward individual level survival, not population or cohort survival. However, generalizations of the model to represent individually estimated covariate trajectories are yet to be done.

A second limitation was that the two-component stochastic process model was specified as a Gaussian diffusion process (Woodbury and Manton 1977). Application of this model to more general forms of covariate changes required ad hoc procedures. In the case of Manton et al.'s (1994a) analysis of the GoM scores from the NLTCS, ad hoc constraints had to be imposed on (1) the regression matrix $\mathbf{A}$, to maintain constraints on the sums and the signs of the GoM scores, and (2) the covariance matrices of the elements of $\mathbf{e}_{i(a+1)}$, to maintain the restricted ranges of the variances and covariances of the GoM scores.

The remainder of this paper presents a new approach to calculating individually estimated covariate trajectories based on the specification of an explicit temporal structure for the GoM model. The approach is substantially simpler and easier to apply than the prior approach, which adapted the two-component Gaussian stochastic process model to the analysis of the changes in the GoM scores estimated from each wave of the NLTCS. Moreover, with fixed covariate trajectories, the approach logically follows as an integration and generalization of the fixed frailty and the declining vitality models.

Data

The longitudinal GoM model was estimated using data from the second through fifth waves of the National Long-Term-Care Survey, which was conducted in 1982, 1984, 1989, 1994, 1999, and 2004. The first five waves of the NLTCS contained longitudinal and cross-sectional data on a nationally representative sample of 41,947 U.S. elderly persons who were enrolled in Medicare and were aged 65 years or older at some point during 1982–1999. The sixth wave of the survey initiated field operations in late 2004 and is expected to be available for analysis in late 2005.

The NLTCS covered both institutionalized and noninstitutionalized persons. The wave-specific sample sizes during 1982–1999 ranged from 17,286 to 22,139 persons, with from 3,112 to 5,552 persons disabled and living in the community, and 1,036 to 1,946 persons in institutional residence.

The response rates were excellent for the first five waves of the survey (95–97 percent: see Manton et al. 1993, 1997; Manton and Gu 2001; Freedman et al. 2004). All institutionalized persons were designated for a detailed interview except in 1982. A screener interview targeted noninstitutionalized disabled persons for further study using a detailed community interview. At the time of each survey, a replenishment sample of 5,000–5,500 persons who attained their 65th birthday in the period following the prior survey was added to the surviving sample to replace the deaths occurring since the prior survey and to ensure that the new sample was representative of the entire elderly population aged 65 years or older.

The NLTCS provided data on age, sex, race, residence type (community vs. institutional, 1982–1999; and assisted living, 1999–2004), height, weight, alcohol and cigarette use, exercise, 30 major medical conditions, vision, subjective health status, seven activities of daily living (ADLs), nine instrumental activities of daily living (IADLs), eight functional limitation items (Nagi 1976), cognitive status (Short Portable Mental Status Questionnaire [SPMSQ] 1982–1994, 2004; Mini–Mental State Examination

[MMSE] 1999), short-term memory, and aberrant behaviors. A detailed listing of these items is provided in the Appendix.

The NLTCS questionnaire items used to assess ADL limitations identified activities in which the respondent received active physical help from another person during the week prior to the interview. This provided an objective anchor against which the use of standby help or special equipment to cope with lower levels of limitation can be compared; the questionnaire also probed for activities in which help was needed but not received, so that the entire spectrum of ADL limitations was represented in the NLTCS. The only subjective measure occurred when the respondent reported that help was needed but not received. This latter category of limitation was not counted as an ADL limitation in either the traditional NLTCS disability classification algorithm or in the HIPAA classification algorithm (Stallard 2000, 2001; Stallard and Yee 2000). It was included as a mild disability in the five-state Markov chain models developed in the two Stallard papers and Stallard and Yee. Similarly, the assessment of IADL limitations was based on questionnaire items that establish that the respondent cannot perform the activity due to a disability or health problem. This removed socially defined roles as reasons for not performing activities such as cooking, housework, or managing bills (Freedman and Martin 1998).

One frequent misconception about the design of the NLTCS is that the detailed community interview covered only chronically disabled persons. This is not true. Although the screener interview targeted noninstitutionalized disabled persons for further study using the detailed community interview, the NLTCS included over 6,000 detailed community interviews for nondisabled persons.

The size and composition of the nondisabled group changed over time as the survey evolved. For each survey ADL and IADL data were collected in a screener interview given to a nationally representative sample drawn from Medicare lists. In 1994 and 1999 a special "healthy" sample of persons was drawn, using screener interview rejection criteria, who otherwise would not have received the detailed community interview. This yielded a supplementary subsample of 922 nondisabled persons in 1994 and 284 nondisabled persons in 1999. The "healthy" samples can be combined with a second subgroup of 884 persons in 1994 and 1,748 persons in 1999 who were determined to be nondisabled after application of the NLTCS classification protocols in the detailed community interview, yielding a total of 1,806 nondisabled persons in 1994 and 2,032 persons in 1999.

The NLTCS classification protocols in the detailed community interviews identified 536, 980, and 916 additional persons who were nondisabled in the 1982, 1984,

and 1989 surveys, respectively. The large increase in this subgroup between 1982 and 1984 occurred because 1982 was the only year in which all respondents received the complete screener interview. After 1982, persons who previously met the screener criteria were designated for detailed follow-up interviews without further screening.

The NLTCS had large sample sizes at ages 85 and older, ranging from 2,552 persons in 1984 to 3,317 persons in 1999. The 1994 and 1999 surveys introduced supplementary samples of the population aged 95 and older, so that the sample sizes at ages 95 and older increased from 244 in 1989 to 745 in 1994 and 930 in 1999.

All NLTCS records were linked to Medicare vital statistics and beneficiary claims data for calendar years 1982 and later, with ongoing periodic updating to allow virtually complete tracking of mortality (Kestenbaum 1992). This linkage resolved concerns about bias in mortality rates due to the 5 percent per wave nonresponse rate, which can be significant in traditional longitudinal designs in which respondent rosters were not linked to an administrative record system such as Medicare.

A related issue is the impact of nonresponse on the estimated disability trajectories for persons who were initially interviewed at some prior wave. The longitudinal GoM model was designed to be robust in this case. The primary concern is that, even when controlling for prior health status, the missing responses were correlated with poor health at the time of the missing wave. In this case population estimates that do not control for prior health status may be biased (see Manton et al. 1991 for discussion).

Age at last birthday on the date of the NLTCS interview was queried on the interview and verified against the age computed from the Medicare vital statistics data. Discrepancies were resolved using the Medicare data, which were accurate up to about ages 95–99 (Kestenbaum 1992) and were consistent from one survey to the next, because the computed ages were based on recorded dates of birth.

Methods

This section generalizes the basic Grade of Membership technique, a categorical data procedure that allows large numbers of variables to be simultaneously analyzed (Woodbury and Clive 1974; Woodbury et al. 1978; Manton and Stallard 1984; Woodbury et al. 1994; Manton et al. 1994b), to estimate individually defined covariate trajectories from longitudinal data.

The GoM technique is appropriate as the starting point for such a task because GoM generates scores for each individual person included in the analytic data set. The GoM scores are an essential building block for defining covariate trajectories. Moreover, because the GoM scores are derived from data on morbidity, disability, and other health-related characteristics, they represent individual measures of fixed frailty, and the age changes in those scores represent individual measures of declining vitality.

The GoM scores can be included in a system of linked equations to predict the probabilities of future covariate outcomes based on the observed distributions of covariates in the analytic data set. The method permits predictions of morbidity, disability, and other health-related characteristics such as the use, cost, and intensity of long-term-care (LTC) services for the elderly.

The expressions for the age changes in GoM scores can be written in the form of an age-inhomogeneous K-state Markov chain in which the K-element vector of GoM scores for the i-th individual, denoted by $\mathbf{g}_i$, with elements g_{ik}, $k = 1,\ldots, K$, represents the initial state vector for the K states.

In conventional Markov chain models, an individual occupies only one state at a time, and, with this restriction, one can say that the states are "crisp" (Manton et al. 1994b).

When GoM techniques are combined with Markov models, however, two things are different. First, each GoM state is described in relation to imaginary, ideal individuals, or "pure types." Second, the GoM states are not "crisp"; instead they are "fuzzy" (Zadeh 1965; Singpurwalla and Booker 2004). Thus, each individual is characterized by combinations of the GoM states. The extent to which the individual is characterized by any one pure-type state is referred to as the individual's grade of membership (GoM) score for that state.

Each individual has a GoM score for each of K states; the scores fall between 0 and 1, inclusive, and they sum to 1. These constraints imply that the vector of GoM scores is a *stochastic* vector. The GoM scores, however, are not *probabilities* of membership in each of the K states; instead they are *grades* of membership. If they were probabilities, then the GoM model would simplify to a latent class model (Lazarsfeld and Henry 1968), and the initial state vector, $\mathbf{g}_i$, would be a probability distribution vector.

Differences between the GoM and latent class models are manifest when the likelihoods are compared and when the posterior probabilities of classification are

considered (Manton et al. 1992b). Under the latent class model, as more information is obtained on each individual, the posterior probabilities of classification tend toward 1 for the correct class and 0 for all other classes. Under the GoM model, as more information is obtained on each individual, one obtains more precise estimates of the GoM scores without convergence to the boundary values, 0 or 1.

Given the initial stochastic state vector of GoM scores, subsequent values of the stochastic state vector are obtained via matrix multiplication by a sequence of stochastic transition matrices, for which the elements in each row (assuming postmultiplication) are non-negative and sum to one. These operations, which maintain the constraints on the state vectors of GoM scores, are identical to the operations that constitute a conventional age-inhomogeneous K-state Markov chain.

Thus, the set of K GoM scores defines the K-element vector $\mathbf{g}_i$. Disease and disability progression are represented as changes in the GoM score vector from one wave of the NLTCS to the next as the individual ages, possibly moving from the community to an institution, developing impairments, changing behaviors, or dying.

The remainder of this section discusses the details of the GoM model and its application to longitudinal data analysis.

Basic GoM Model

The longitudinal GoM model is most easily understood when described as an extension of the basic GoM model (Woodbury and Clive 1974).

The basic GoM model is used to analyze data for I persons, indexed by i, $i = 1,\ldots, I$, with measurements or observations on J multinomial covariates, indexed by j, $j = 1,\ldots, J$, each of which is measured with L_j distinct outcomes or response levels. The j-th covariate for the i-th individual is typically denoted as x_{ij}. However, given that the outcomes are discrete, it is more convenient to employ binary coding for the data such that $y_{ijl} = 1$ if $x_{ij} = l$ and $y_{ijl} = 0$ if $x_{ij} \neq l$. By convention, this coding rule sets $y_{ijl} = 0$ if the value of x_{ij} is unknown or is missing from the data, or if x_{ij} is undefined because the j-th covariate was properly skipped due to the i-th individual's response on a prior screening question.

From the J covariates, K latent dimensions can be identified, where K is the number of pure-type individuals or ideal states in the specified model. In the longitudinal application, K is the number of components or elements of the fixed frailty vector. Generally, $K < J/2$ is a sufficient condition for identifiability of the GoM model

(Woodbury et al. 1994). Practically, K-values very much smaller than $J/2$ lead to satisfactory models (Wachter 1999).

For example, the longitudinal application in this paper uses $K = 3$, 4, or 5 with $J = 95$, implying that $K \leq J/19$ (where 19 = 95/5). For the four-pure-type model, the NLTCS data set provided 922,303 responses to support the estimation of 46,116 parameters for males and 1,587,710 responses to support 66,480 parameters for females (see Appendix). Although the number of parameters is large, the number of responses used in estimating those parameters is substantially larger by factors of 20.5 and 24.5, respectively.

Two types of parameters are estimated in basic GoM. The first are the λ_{kjl}'s, which are the *pure-type probabilities*; the second are the g_{ik}'s, which are the *GoM scores*.

Like the y_{ijl}'s, the λ_{kjl}'s are matched to the indexes (j and l) of the responses to the J covariates. The λ_{kjl}'s for the j-th covariate are multinomial probabilities, where each parameter λ_{kjl} is the probability that (only) the l-th response is observed for pure type k (or for an individual i who is exactly like pure type k) on variable j, subject to the convexity constraints $0 \leq \lambda_{kjl} \leq 1$ and $\sum_{l=1}^{L_j} \lambda_{kjl} = 1$.

The g_{ik}'s are state variables that quantify how well each individual's observed state is described by each of the K pure-type dimensions or ideal states. An individual with $g_{ik} = 1$ for some index k is exactly like the k-th pure type. The g_{ik}'s are convexly constrained scores for individuals, that is, $0 \leq g_{ik} \leq 1$ and $\sum_{k=1}^{K} g_{ik} = 1$ (Woodbury et al. 1994).

As noted above, the binary coded 0–1 variables for the l-th response to the j-th variable are indicated by y_{ijl}. The fundamental equation of the basic GoM model generates the probability of each discrete outcome (i.e., the probability that $y_{ijl} = 1$ or $x_{ij} = l$) from the "continuous" state variables (g_{ik}) using an inner product form

$$\text{Prob}(y_{ijl} = 1) = \sum_{k=1}^{K} g_{ik} \lambda_{kjl} \ .$$

For a given set of observations, the likelihood, L, is expressed as the product over i, j, and l of the set {Prob($y_{ilj} = 1$)}

$$L = \prod_i \prod_j \prod_l \left(\sum_k g_{ik} \lambda_{kjl} \right)^{y_{ijl}} \ .$$

The product form of the likelihood over the index i represents the standard assumption that individuals are statistically independent. The product form over the index j represents the fundamental assumption of the GoM model that the J covariates

are statistically independent conditional on the responses to any prior screening questions, and conditional on the GoM score vector $\mathbf{g}_i$. The binary coding of the y_{ijl}'s ensures that all but the one term indexed by $l = x_{ij}$ will have unit value for each observed covariate, j, and hence will not affect parameter estimation.

Parameter estimation follows Manton and Stallard (1984) or Woodbury et al. (1994). Parameter estimation may be conducted using fixed-point iteration procedures to update existing estimates of g_{ik} and λ_{kjl} as follows:

$$g_{ik} \leftarrow \frac{g_{ik} \times \sum_j \sum_l y_{ijl} \lambda_{kjl} / p_{ijl}}{\sum_j \sum_l y_{ijl}} \quad \text{and} \quad \lambda_{kjl} \leftarrow \frac{\lambda_{kjl} \times \sum_i y_{ijl} g_{ik} / p_{ijl}}{\sum_{l'} \sum_i y_{ijl'} g_{ik} \lambda_{kjl'} / p_{ijl'}},$$

where

$$p_{ijl} = \sum_k g_{ik} \lambda_{kjl} \quad .$$

If the final values of the g_{ik}'s are known, then the formula for the λ_{kjl}'s may be used to iteratively solve for the final λ-values. Each iteration is structured so that the update computations are run for all combinations of the subscripts k, j, and l using a fixed set of λ-values, after which the λ_{kjl}'s are updated as a set.

If the final values of the λ_{kjl}'s are known, then the formula for the g_{ik}'s may be used to iteratively solve for the final g-values. Each iteration is structured so that the update computations are run for all combinations of the subscripts i and k using a fixed set of g-values, after which the g_{ik}'s are updated as a set.

If both the g_{ik}'s and λ_{kjl}'s are unknown, the formulas may be used by alternating between the g-values and λ-values.

Once the GoM score vectors, $\mathbf{g}_i$, are estimated, they can be used as supplementary input data to conditional probability models for resource utilization, costs of services, and other health-related or survival-related measures.

Longitudinal GoM Model

Under a longitudinal GoM model, each individual observed at multiple times in a longitudinal survey is viewed as following a distinct trajectory as he or she ages. The 1984–1999 NLTCS provides the opportunity to evaluate 15 years of experience along these trajectories, sufficient time to characterize the trajectories for many disabled persons.

To generalize the basic GoM model for longitudinal data, one first needs to add an age (or time) index to the basic GoM score vector, $\mathbf{g}_i$, so that this vector is respecified as $\mathbf{g}_{ia}$ for individual i at age a. For the NLTCS the index a increases by five years with each new wave of the survey; for notational simplicity, however, one-year (or one-unit) age increments are used in the following development.

One way to estimate a sequence of $\mathbf{g}_{ia}$'s for individuals in the NLTCS is to apply the basic GoM model to pooled data from the available waves. In this case the observed data for each individual at each wave are jointly indexed by i and a; the observations corresponding to distinct combinations of i and a are treated as independent observations in the basic GoM likelihood function. In effect, the index i is replaced by the index pair ia in the model's equations.

This approach was used in Manton et al. (1992a) and Manton et al. (1994a). In both cases follow-up assessments of the changes in $\mathbf{g}_{ia}$ over age were required to complete the analyses. This was done using a linear dynamic form for the covariate changes, as assumed in the two-component Gaussian stochastic process model. Thus,

$$\mathbf{g}_{i(a+1)} = \mathbf{A}_a \mathbf{g}_{ia} + \mathbf{e}_{i(a+1)},$$

where $\mathbf{e}_{i(a+1)}$ was a vector of randomly distributed residual values (i.e., errors or innovations) at age $a + 1$. Because the distribution of $\mathbf{e}_{i(a+1)}$ was not Gaussian, ad hoc constraints were imposed on the diffusion process (see Manton et al. 1991).

An alternative approach is to drop the residual term from the updates, obtaining

$$\mathbf{g}_{i(a+1)} = \mathbf{U}'_{a+1}\mathbf{g}_{ia} \quad \text{or, equivalently,} \quad \mathbf{g}'_{ia} = \mathbf{g}'_{i(a-1)}\mathbf{U}_a,$$

where $\mathbf{U}_a$ is the Markov-chain transition matrix governing transitions in $\mathbf{g}_{i(a-1)}$ from age $a - 1$ to a. New notation is needed to distinguish $\mathbf{U}_a$ from $\mathbf{A}_a$, due to the transposed form of the revised update equation and the change in the age index.

The longitudinal GoM model is specified using the sequence of updates to $\mathbf{g}_{ia}$ from age a_0 to age a. The model defines a fixed frailty GoM vector, $\mathbf{g}_i$, as an age-invariant vector of GoM scores or, equivalently, as an initial vector of GoM scores at age a_0. It follows that

$$\begin{aligned}\mathbf{g}'_{ia} &= \mathbf{g}'_i \left\{ \prod_{\alpha=a_0}^{a} \mathbf{U}_\alpha \right\}, \\ &= \mathbf{g}'_i \mathbf{V}_a\end{aligned}$$

where

$$\mathbf{V}_a = \prod_{\alpha=a_0}^{a} \mathbf{U}_\alpha,$$

with $\mathbf{U}_{a_0} = \mathbf{I}$, an identity matrix. The model uses the $\mathbf{V}_a$-matrices to represent declining vitality.

Operationally, the inner product form in the fundamental equation of the basic GoM model is replaced with a bilinear form:

$$\text{Prob}(y_{ijla} = 1) = \mathbf{g}_i' \left\{ \prod_{\alpha=a_0}^{a} \mathbf{U}_\alpha \right\} \boldsymbol{\lambda}_{m_{jl}} = \mathbf{g}_i' \mathbf{V}_a \boldsymbol{\lambda}_{m_{jl}},$$

where the vector subscript m is used to index the combination (j, l), and the subscript a is used to index age. The likelihood is the product over i, j, l, and a of the set {Prob(y_{ijla} = 1)}:

$$L = \prod_i \prod_j \prod_l \prod_a \left(\mathbf{g}_i' \left\{ \prod_{\alpha=a_0}^{a} \mathbf{U}_\alpha \right\} \boldsymbol{\lambda}_{m_{jl}} \right)^{y_{ijla}}.$$

Parameter estimation for the g-values and λ-values follows the fixed-point iteration procedures provided above. To implement those procedures, one needs to rewrite the above expression in an inner product form, isolating either the g-parameters or the λ-parameters as follows:

$$\text{Prob}(y_{ijla} = 1) = \mathbf{g}_i' (\mathbf{V}_a \boldsymbol{\lambda}_{m_{jl}}) = \mathbf{g}_i' \boldsymbol{\lambda}_{m_{jl}a}$$

and

$$\text{Prob}(y_{ijla} = 1) = (\mathbf{g}_i' \mathbf{V}_a) \boldsymbol{\lambda}_{m_{jl}} = \mathbf{g}_{ia}' \boldsymbol{\lambda}_{m_{jl}}.$$

Fixed-point iteration procedures can be implemented for the elements u_{kca} of $\mathbf{U}_a$, subject to the following convexity constraints: $0 \le u_{kca} \le 1$ and $\sum_{c=1}^{K} u_{kca} = 1$. The update equation for u_{kca} is

$$u_{kca} \leftarrow \frac{u_{kca} \times \sum_{t=a} \sum_i \sum_j \sum_l y_{ijlt} g_{ik(a-1)} \lambda_{cjl(a+1)t} / p_{ijlt}}{\sum_{c'} \sum_{t=a} \sum_i \sum_j \sum_l y_{ijlt} g_{ik(a-1)} u_{kc'a} \lambda_{c'jl(a+1)t} / p_{ijlt}},$$

where

$$p_{ijlt} = \sum_k \sum_c g_{ik(a-1)} u_{kca} \lambda_{cjl(a+1)t},$$

and $\lambda_{cjl(a+1)t}$ is the element in row c of the following vector:

$$\boldsymbol{\lambda}_{m_{jl}(a+1)t} = \left(\prod_{s=a+1}^{t} \mathbf{U}_s \right) \boldsymbol{\lambda}_{m_{jl}}.$$

The updating formulas may be used by alternating between the sets of g-values, u-values, and λ-values. All parameter updates are positive and meet the constraints of the model. Certain parameters converge to 0, and these are identifiable as infinitesimal values. Convergence can be tested by perturbing the estimates and verifying that

subsequent solutions converge to the same values. Improvements to the estimation procedures based on modified Newton-Raphson procedures are being investigated.

Conditional GoM Model

Under a longitudinal GoM model, changes in the health statuses of individuals are described by changes over age or time of the GoM score vectors, and the utilization of resources, costs of services, and age-specific probabilities of death are described by a second model in which the GoM score vectors are assumed to be known. This second model is a conditional probability model because it depends on the first model to provide estimates of the GoM score vectors. The parameters of the conditional probability model can be estimated with a likelihood function identical in form to the likelihood functions shown above, the only difference being that the numerical values of the g_{ik}'s or g_{ika}'s are assumed to be known.

Results

The longitudinal GoM model with three, four, and five pure types or ideal states (i.e., $K = 3$, 4, or 5) was independently estimated for 15,102 males and 21,890 females in the NLTCS using data on 95 independent variables characterizing the nature and intensity of limitations in ADLs, IADLs, physical functioning, and cognitive functioning, as well as indicators of behavioral characteristics, medical conditions, subjective health, age, race, sex, and institutional status. This required a total of six sets of parameter estimation.

The selection of three, four, and five pure types was based on prior experience in applying a related form of this model to represent the natural history of Alzheimer's disease in the period beginning immediately after the first symptoms (Kinosian et al. 2000, 2004). In the Alzheimer's application the temporal dimension was time since onset. In the present application the temporal dimension was age. The difference was that a large number of respondents in the NLTCS commenced follow-up after age 65–69, raising the possibility that the age-dependent parameters may be unstable.

Model selection was based on the likelihood ratio test statistics for $K = 1$ versus $K = 3$, 4, or 5 using the standard χ^2 approximation (Wilks 1938). Combining the male and female results, the total χ^2 value for $K = 3$ was 601,826 with 74,786 degrees of freedom (d.f.), indicating that $K = 3$ was significantly better than $K = 1$—that is, the null model for a homogeneous population. The incremental χ^2 value for $K = 4$ versus $K = 3$ was 45,800 with 37,414 d.f., which was statistically significant. The incremental χ^2 value for $K = 5$ versus $K = 4$ was 55,071 with 37,428 d.f., which was also statistically significant.

However, the male χ^2 value of 10,433 with 15,320 d.f. was not statistically significant, indicating that five pure types were too many for males.

All parameter estimates were constrained to lie within the closed interval [0, 1]. Self and Liang (1987) indicated that these constraints imply more complex forms for the distributions of the likelihood ratios than derived by Wilks (1938). As a result, the χ^2 test statistics described above were conservatively biased, and the computed tail probabilities were larger than they should have been (Manton et al. 1994b). Corrections for such bias can be approximated by downward adjustments in the degrees of freedom for the likelihood ratio χ^2 test statistics.

Akaike (1974) indicated that the incremental χ^2 value should be at least twice the degrees of freedom before accepting a more complex model. Using Akaike's criterion, $K = 3$ would be the preferred model. Given the higher K-values indicated by the χ^2 tests and the likely upward bias in the degrees of freedom employed in Akaike's criterion, the model with $K = 4$ was selected as a reasonable compromise.

Logarithms of the likelihood ratios, χ^2 test statistics, and associated degrees of freedom for testing $K = 4$ versus $K = 1$ for the 95 variable are shown in the Appendix for males and females separately. The χ^2 test statistics were statistically highly significant at conventional levels for all 95 variables.

Geometrically, the boundary constraints on a four-pure-type GoM model imply that the entire process occurs within (or on the boundaries of) a three-dimensional object shaped as a regular tetrahedron (i.e., triangular pyramid). The base and side faces of this object are equilateral triangles.

The temporal trajectories of the GoM scores represent paths within this regular tetrahedron along which individuals travel as they age. For example, if the vertex opposite the base corresponds to the healthiest pure type, then a typical trajectory for a healthy 65-year-old person would begin near the top vertex and gradually move toward the base as the person aged. Movement toward one or another base vertex would depend on the nature of the medical conditions and disabilities that the person acquired as well as the correspondence of each base vertex with the other pure types and the likelihood that these pure types would manifest specific medical conditions and disabilities.

In contrast, the boundary constraints on a three-pure-type GoM model imply that the entire process occurs within (or on the sides of) a two-dimensional object shaped as an equilateral triangle; while the constraints on a two-pure-type GoM model

imply that the entire process occurs on a one-dimensional line segment. Two is the minimum number of pure types needed to model fixed frailty with declining vitality. The use of four pure types in the current analysis allows for substantial heterogeneity in initial frailty as well as in the trajectories of declining vitality.

Table 1 displays the average GoM scores for the four-pure-type model with stratifications by age at initial interview and sex. The GoM pure types are indexed by Roman numerals (I–IV). Type I has the highest average GoM scores, ranging from 48.0 to 59.2 percent. The other types share the remaining portion of the membership. Males aged 80–84 and 85–89 show a jump in the Type I average in Table 1 compared to younger and older males; a corresponding jump occurs for females, but only at age 80–84.

The independent estimation of the model by sex means that the GoM pure types are sex-specific. The pure types do not necessarily mean the same thing for males and females.

Comparison across ages indicates that all of the pure types are represented at all ages, although there are some noticeable discontinuities in the age trend (e.g., compare male Type III at ages 70–74 with ages 75–79 in Table 1). Some of the instability in age trend may be due to the commencement of follow-up after age 65–69. Type I has the lowest mortality risks. Hence, some upward trend in Type I, and downward trends in the other types, could be a result of the impact of mortality selection on the pure-type distribution.

Table 2 provides summary statistics on the variances of the GoM scores, whose averages are shown in Table 1. The statistics selected for display are the age-specific Bernoulli relative variances (BRVs), defined as the ratio of the variance of the GoM scores for pure type k within the specified age category a to the maximum variance that can be attained in a population with the mean constrained to the corresponding value, $\bar{g}_{ak}$, for age category a and pure type k reported in Table 1. Hence,

$$BRV_{ak} = \mathrm{var}(g_{ak}) / \left[\bar{g}_{ak}(1-\bar{g}_{ak})\right].$$

The BRVs lie in the range [0, 1], attaining a value of 0 when all persons have the average score and the value 1 when all persons have a score of 1 or 0 with the probabilities $\bar{g}_{ak}$ and $(1-\bar{g}_{ak})$, respectively. The BRVs in Table 2 indicate that the variances of the GoM scores span the middle three quintiles of the range.

Tables 3 and 4 display the age-specific GoM transition matrices ($\mathbf{U}_a$) and the resulting vitality matrices ($\mathbf{V}_a$) for the four-pure-type model for males and females,

respectively. The transition matrices govern the age changes in the GoM scores in a manner identical to a conventional age-inhomogeneous *K*-state Markov chain: that is, the GoM scores for each individual respondent to a given NLTCS interview can be arranged in a row vector that is postmultiplied by the appropriate age-specific transition matrix in Tables 3 or 4 to generate the GoM scores at the next interview.

Each age-specific vitality matrix is a cumulative product of the transition matrices for all ages up to and including the specified age. Examination of the vitality matrices shows that the "membership mass" begins to move away from Types II and III at age 70–74 for females and age 75–79 for males and toward Type IV; by age 85–89 all of the membership for males and females in Types II and III has transitioned to Type IV.

Membership in Type I persists for males to age 95–99, with transitions to Types II and III beginning at age 80–84 (and at age 75–79 for females). For both sexes membership in Type III from ages 80–84 to 100–104 is solely a result of transitions from Type I.

These transitions produce vitality matrices at and above age 80–84 for females and age 85–89 for males in which the only nonzero elements are in the first row and last column, with a pattern resembling the number "7" or, upon transposing the matrices, the letter "L."

Given that the transition and vitality matrices at age 65–69 are diagonal in form, by assumption, the transformation to an L-pattern is a remarkably simple outcome. Understanding the implications of this transformation can yield significant insight into the nature of the aging process.

Table 5 displays the annual probabilities of death by age and sex. These were estimated as λ-values using the conditional form of the longitudinal GoM model with the g-values and the u-values fixed at the values underlying Tables 1, 3, and 4. To exploit the available data fully, the annual mortality probabilities were estimated from all 20,717 deaths (for 8,747 males and 11,970 females) in the 17-year period following the 1984 NLTCS, rather than just for the four one-year periods following each survey (i.e., 1984, 1989, 1994, and 1999). This was done by linearly interpolating the individual GoM scores to single years of age (i.e., age at last birthday) and time, and then regrouping the annual observations into the eight quinquennial exposure-age categories shown in Table 5.

For both sexes Type IV has the highest death probabilities, and Type III ranks second. Types I and II have probabilities close to 0, indicating that these are extremely

healthy types, except for males aged 95 or older, where the results are less credible due to small sample sizes. For females the corresponding probabilities were not estimable because the average GoM scores were equal to 0 within those cells.

Table 6 displays the death probabilities after adjustment for declines in vitality, using the formula $\boldsymbol{\lambda}_{m_{jl}a} = \mathbf{V}_a \boldsymbol{\lambda}_{m_{jl}}$ to make the adjustment. Types II and III converge rapidly to the high death probabilities of Type IV. Type I maintains its advantage throughout the entire age range.

Figures 1 and 2 display the observed and predicted annual probabilities of death for males and females by age at last birthday at the start of the one-year follow-up period. The predicted probabilities were calculated by multiplying the age-specific probabilities in Table 6 by the corresponding age-specific average (interpolated) GoM scores. The observed and predicted probabilities are virtually identical.

Also shown in Figures 1 and 2 are the age-specific probabilities reported in the U.S. decennial life tables (USDLT) for 1989–1991 (NCHS 1997), a period selected because it was close to the 1992–1993 midpoint of the 17-year period following the 1984 NLTCS. The plotted points from the USDLT were annual probabilities of death for persons whose exact ages at last birthday at the start of the one-year follow-up period were at the midpoints of each quinquennial category. The comparisons with the observed/predicted rates indicate that the NLTCS mortality rates closely reproduced the age patterns of increase in the USDLT, which approximates Gompertz's law.

An exact match was not expected because the initial waves of the NLTCS oversampled the disabled population, a group that had mortality rates higher than the general population. On average, the excess mortality was about 1.7 percent for males and 4.4 percent for females. The NLTCS provided sampling weights that can be used to reweight the individual characteristics to exactly match population characteristics. However, such weights were neither needed nor used in the current analysis because the focus was on individual trajectories of disability and mortality, not population averages.

Figures 1 and 2 demonstrate that the averages of the individual trajectories follow known age patterns (i.e., with increases approximated by Gompertz's law) for human mortality above age 65. The averages of the individual trajectories also provide a basis for comparison with the age patterns for individual mortality above age 65, which are not well known.

Figures 3 and 4 display the adjusted individual annual probabilities of death shown in Table 6 for four categories of pure-type males and females by age at last birthday at the start of the one-year follow-up period. Also included in the figures are the average predicted probabilities for the NLTCS sample shown in Figures 1 and 2. The figures show that the pure-type probabilities of death increased slower than (Type IV), faster than (Types II–III, until reaching convergence with Type IV), or about the same rate (Type I) as predicted by Gompertz's law, after adjusting for declines in vitality. The probabilities of death for Type IV were the highest at all ages. For females there were no ages for which the Type IV probabilities clearly accelerated faster than expected under Gompertz's law; for males a rapid acceleration occurred between ages 65–69 and 75–79, but not thereafter. The Type III probabilities converged quickly to those for Type IV, but the initial rate of increase was substantially faster than expected under Gompertz's law, and faster than for Type IV for the younger males.

Figures 5 and 6 display the unadjusted individual annual probabilities of death shown in Table 5 for four categories of pure-type males and females by age at last birthday at the start of the one-year follow-up period. Also included in the figures are the average predicted probabilities for the NLTCS sample shown in Figures 1 and 2. The unadjusted pure-type probabilities of death were either relatively constant (I–II, and III [females]) or increasing (IV and III [males]) over age at a rate slower than that of the average predicted probabilities (i.e., slower than expected under Gompertz's law). The Type I value for males aged 95–99 was 1.00 (which was off the graph), but this declined to 0.27 at age 100–104. These latter two results were less credible due to small sample sizes at these ages (see Table 5).

Table 7 displays the average adjusted GoM scores in the four-pure-type model for males and females by age. The averages are displayed graphically in Figures 7 and 8 using a 100 percent stacked-line format to highlight the depletion or exhaustion (females only) of Types I and II over age, and the corresponding increases in Types III and IV over age. The GoM scores are constrained to sum to 1.0 (100 percent) so that the reductions in the average scores for Types I and II necessitate corresponding increases in the average scores for Types III and IV.

The low average Type I and II GoM scores for males at ages 95 and older in Table 7 combined with the small numbers of person-years at risk for males at ages 95 and older in Table 5 account for the instability noted above for the probabilities of death in Table 5 for Types I and II. The complete absence of females with nonzero Type I and II GoM scores at ages 85–89 or 90–94 and older in Table 7 accounts for the inability to estimate the corresponding unadjusted probabilities of death in Table 5. The use of linear interpolation of the GoM scores allowed one additional probability to be

computed in Table 5 for Types I and II than would be the case if the noninterpolated data in Table 7 were used.

Figures 9 and 10 display the average unadjusted GoM scores shown in Table 1 for the four-pure-type model for males and females by age, using a 100 percent stacked-line format to highlight the age pattern. Ideally, one would have expected Types I and II to increase over age due to differential mortality selection. Under such selection persons who survive to the oldest ages are those who faced the lowest mortality probabilities at younger ages. This expectation is partially fulfilled up to about age 80–84. Beyond that age Type IV exhibits an unanticipated increase.

The increase in Type IV may be an artifact of the structure of the vitality matrices in Tables 3 and 4 in which most of the initial pure-type membership for Types I and II is converted to Types III and IV beyond age 85–89. With such conversion the declines in Types I and II have no impact on the likelihood function for females and little impact for males because the GoM scores for these two pure types convert to Types III and IV after adjustment by the vitality matrices.

Table 8 displays the average adjusted GoM scores in the four-pure-type model for males and females at the initial NLTCS interview, stratified by the number of ADLs expected to meet the HIPAA ADL disability criterion using an adaptation of the NLTCS procedures detailed in Stallard and Yee (2000). Dramatic changes in the membership levels for Type IV can be seen as one moves from 0 to 1 to 2 HIPAA ADLs. Beyond that point the GoM membership continues to move toward Type IV, but at a more gradual pace.

Table 9 displays the average adjusted GoM scores in the four-pure-type model for males and females at the initial NLTCS interview, stratified by the number of IADLs that meet the NLTCS triggering criteria. The overall pattern is similar to that seen in Table 8, except that the changes at the lowest IADL counts are more gradual than for the lowest ADL counts.

The shift in membership from Types I–III to Type IV with increasing ADL and IADL disability counts was consistent with the finding (not shown) that Type IV was the only pure type that had nontrivial λ-values for the ADL and IADL disability responses.

Alternatively, Types I–III had 0 or trivial λ-values for all of the ADL and IADL disability responses. This was consistent with the low mortality probabilities for Types I and II, the disappearance of these pure types for females above age 85–89, and the near-

disappearance of these pure types for males at these ages. The average GoM scores for Type III for males and females aged 85–89 and older with no HIPAA ADLs were 0.545 and 0.731, respectively. The corresponding averages for Type IV for males and females with no HIPAA ADLs were 0.109 and 0.173, respectively. These results indicated that membership in Type III at the oldest ages was associated with greatly reduced risk of ADL and IADL disabilities.

Discussion

This paper had two goals: to introduce the reader to a broad range of research on survival at advanced ages, and to integrate the findings of that research into a coherent model of the trajectories of change in health and survival characteristics of individuals. The model was structured to simultaneously represent the essential features of the fixed frailty model (Vaupel et al. 1979) and the model of linearly declining vitality (Strehler and Mildvan 1960). Unlike those models, however, the new model was designed for easy and direct application to existing longitudinal data sets.

The application was successful in characterizing the health and survival experience of respondents to the NLTCS. The model was structured to represent the effects of variables that directly measured individual health, disability, and health-related behaviors. The results of the analysis indicated that the four-pure-type longitudinal GoM model is a parsimonious and powerful model of the changes in health and survival characteristics of individuals.

Several findings were noteworthy:

The likelihood ratio statistical tests comparing $K = 1$ with $K = 3$, 4, or 5 demonstrated that the elderly population is not homogeneous. This seems obvious, but it is a critical step in establishing the dimensionality of the longitudinal GoM model.

The GoM scores exhibited substantial variability relative to the Bernoulli bounds. The GoM scores varied over individuals, but they were not exclusively clustered at the boundaries with values of 0 or 1.

Approximately 20–35 percent of the initial age-invariant GoM membership was in Types III and IV, the pure types with the highest mortality rates. Membership in these pure types indicated impaired health at age 65.

The distributions of the initial age-invariant GoM scores provided an effective and quantifiable implementation of the concept of fixed frailty.

The cumulative transition or vitality matrices at and above age 80–84 for females and age 85–89 for males had patterns in which only the first row and last column contained nonzero values, forming a pattern resembling a transposed "L."

A transposed L-shape also occurred for males and females with the three-pure-type model and for females with the five-pure-type model. For males the transposed L-shape also occurred with the five-pure-type model except for terms on the second diagonal (for Type II). The transposed L-shapes of the vitality matrices were robust with respect to the dimensionality of the model.

The cumulative transition or vitality matrices provided an effective and quantifiable implementation of the concept of declining vitality.

The combination of the initial distributions of GoM scores with the transposed L-shapes of the vitality matrices was consistent with a multistage process with random initial defects. The transposed L-shape arose because membership in Types II and III at older ages represented "replacement" membership drawn from Type I, not persisting initial membership in Types II and III.

The differences in mortality probabilities between the four pure types were substantial (see Figures 3 and 4). At age 65 these differences imply life expectancies that range from 5.4 to 22.9 years for males and from 6.2 to 25.9 years for females, based on the survival probabilities for Types IV and I, respectively. The substantially greater life expectancy for Type I is consistent with published data on centenarians that report substantially better health and less disability, on average, in the decades prior to reaching 100 than that of their contemporaries (Perls et al. 1999).

The adjusted mortality probabilities for the four pure types (see Figures 3 and 4) exhibited patterns of increase that were slower than (Type IV), faster than (Types II–III, prior to convergence with Type IV), or about the same rate (Type I) as predicted by Gompertz's law. Thus, Iachine et al.'s (1998) finding that the standard schedules of age-specific forces of mortality faced by individuals in each birth cohort increased substantially faster than the Gompertz function may have been an artifact of the assumption of correlated gamma-distributed frailty among twins. Moreover, the relative ratios of the probabilities for Type IV versus Types I, II, and III declined substantially over age, indicating that the assumption that frailty operated multiplicatively on mortality may not be correct. The longitudinal GoM model made no assumptions about the parametric form of the frailty distribution or about the parametric form of the age-specific probabilities of death.

The unadjusted mortality probabilities were either relatively constant (I–II and III [females]) or increasing (IV and III [males]) over age at a rate slower than that of the average predicted probabilities (i.e., slower than expected under Gompertz's law). This was consistent with the results in Manton (1994a), who found that the exponential growth constant in the two-component Gaussian stochastic process model was reduced substantially when covariates were included in the model. The difference in the current analysis was that most of the remaining increase was restricted to Type IV, especially for females.

Further refinement of the GoM model should focus on the statistical stability and smoothness of the progression of the age-specific parameter estimates. This could be done by imposing additional constraints on the GoM transition matrices, by introducing additional parameters to represent period and cohort effects in the study population, and by expanding the longitudinal follow-up of the individual respondents when the 2004 NLTCS data become available.

The NLTCS also provided contextual data that could be used to extend the model to better understand the causes and consequences of individual differences in the initial GoM scores and in the subsequent trajectories of health, disability, and mortality. These data include information on education, marital status, family structure, income and assets, military service, housing and neighborhood characteristics, health insurance, medical providers and prescription medicines, the number and relationship (to respondent) of caregivers, and caregiver hours/days and type of activity for which help is provided.

Brown and McDaid (2003) conducted a comprehensive literature search, identifying 12 factors affecting retiree mortality that could be considered in developing risk classes for impaired life annuities in the individual annuity market. The 12 factors were age, gender, education, income, occupation, marital status, religion, smoking, alcohol, other health-related behaviors, obesity, and race/ethnicity. Seven of the 12 factors (age, gender, smoking, alcohol, other health-related behaviors, obesity, and race) were included in the current analysis; five factors (education, income, occupation, marital status, and religion) were excluded because they did not directly measure individual health, disability, and health-related behaviors. Four of the five factors (education, income, marital status, and religion) were provided as contextual measures in the NLTCS and could be included in further analyses of these data.

The Appendix provided statistical tests of the 95 variables in the analysis along with sex-specific rank orderings of the significance of each variable using ratios of χ^2 statistics to the associated degrees of freedom. The analysis was stratified by gender,

with age having a special role in structuring the transition matrices. The five factors considered by Brown and McDaid (2003) that were included in the Appendix were rank ordered for males and females, respectively, as follows: smoking (88 and 92), alcohol (69 and 67), other health-related behaviors (51 and 53; for moderate exercise), obesity (49 and 31), and race (16 and 41). While these five factors were significant, the most significant factors were the measures of physical functioning based on the list of IADLs, ADLs, and functional limitations. Measures of cognitive functioning based on the Short Portable Mental Status Questionnaire (SPMSQ) were also highly significant.

The GoM model provided a solution to a problem identified by Brown and McDaid (2003, p. 41): how should one account for the joint effects of large numbers of risk factors that are potentially correlated or statistically dependent? The structure of the GoM model is such that the GoM scores are introduced as explanatory variables that resolve the correlations and statistical dependencies. Thus, conditional on the GoM scores, all risk factors are statistically independent. This condition was enforced in the current analysis via the product form of the GoM likelihood over all measured variables and tested by considering a sequence of models with increasing numbers of pure types (i.e., *K*-values). The statistical tests indicated that these conditions were achieved using a four-pure-type specification.

The NLTCS provides a broad range of data on the use, cost, and intensity of long-term-care and Medicare-funded acute care services at or following the time of each of the four NLTCS interviews. Additional applications of the current model are being developed using the conditional form of the GoM model to analyze these measures for use in conjunction with the individual trajectories to estimate and project lifetime cost measures at the individual level. These individual level results will be aggregated into population estimates and projections using longitudinal sampling weights based on the NLTCS sampling design.

Applications of the GoM model are not restricted to just the NLTCS. The model can be applied to any longitudinal data with sufficient numbers of variables measured on each subject to support parameter estimation. For example, prior work with the Framingham Heart Study (e.g., Manton et al. 1994a) suggests that it would be feasible to develop a longitudinal GoM model for these data. The Framingham Heart Study consisted of 25 biennial examinations with longitudinal follow-up over 48 years on a sample of 5,209 persons initially aged 28–62 years in 1948–1950. Each examination included information on a broad range of cardiovascular disease risk factors and disease events. The later waves of the study (beginning at Exam 14) introduced health, disability, and cognitive functioning measures comparable to those in the NLTCS, suggesting that, with appropriate coding of the common variables, one could conduct

pooled analyses of the two sets of data. Such analyses would double the size of the age range of the current model, with individual trajectories being tracked from age 30 onwards. This would improve the characterization of the early stages of the disablement process using traditional cardiovascular disease risk factors as precursors to the morbidity and disability measures jointly provided by the NLTCS and the later waves of the Framingham Heart Study.

Acknowledgments

Support for the research presented in this paper was provided by the National Institute on Aging, through grants R01AG01159, P01AG17937, P01AG08761, and U01AG07198, by the Department of Veteran Affairs, and by the National Council on the Aging. David L. Straley provided programming support.

References

Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AC-19(6): 716–723.

Bortz, W. M. 2002. "A Conceptual Framework of Frailty: A Review." *Journal of Gerontology: MEDICAL SCIENCES* 57A(5): M283–M288.

Brown, R. L, and J. McDaid. 2003. "Factors affecting Retirement Mortality." *North American Actuarial Journal* 7(2): 24–43.

Chiang, C. L. 1984. *The Life Table and Its Applications*. Malabar, FL: Robert E. Krieger.
Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society: Series B (Methodological)* 34(2): 187–220.

De Benedictis, G., Q. Tan, B. Jeune, K. Christensen, S. V. Ukraintseva, M. Bonafidè, C. Franceschi, J. W. Vaupel, and A. I. Yashin. 2001. "Recent Advances in Human Gene–Longevity Association Studies." *Mechanisms of Ageing and Development* 122: 909–920.
Freedman, V. A., E. Crimmins, R. F. Schoeni, B. C. Spillman, H. Aykan, E. Kramarow, K. L. Land, J. Lubitz, K. Manton, L. G. Martin, D. Shinberg, and T. Waidmann. 2004. "Resolving Inconsistencies in Trends in Old-Age Disability: Report from a Technical Working Group." *Demography* 41(3): 417–441.

Freedman, V. A., and L. G. Martin. 1998. "Understanding Trends in Functional Limitations among Older Americans." *American Journal of Public Health* 88(10): 1457–1462.

Fried, L. P., L. Ferrucci, J. Darer, J. D. Williamson, and G. Anderson. 2004. "Untangling the Concepts of Disability, Frailty, and Comorbidity: Implications for Improved Targeting and Care." *Journal of Gerontology: MEDICAL SCIENCES* 59(3): 255–263.

Fried, L. P., C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, and M. A. McBurnie. 2001. "Frailty in Older Adults: Evidence for a Phenotype." *Journal of Gerontology: MEDICAL SCIENCES* 56A(3): M146–M156.

Gavrilov, L. A., and N. S. Gavrilova. 2001. "The Reliability Theory of Aging and Longevity." *Journal of Theoretical Biology* 213: 527–545.

Gompertz, B. 1825. "On the Nature of the Function Expressive of the Law of Human

Mortality, and on a New Mode of Determining the Value of Life Contingencies." *Philosophical Transactions of the Royal Society of London* 115: 513–583.

Hougaard, P. 1984. "Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity." *Biometrika* 71(1): 75–83.

Iachine, I. A., N. V. Holm, J. R. Harris, A. Z. Begun, M. K. Iachina, M. Laitinen, J. Kaprio, and A. I. Yashin. 1998. "How Heritable Is Individual Susceptibility to Death? The Results of an Analysis of Survival Data on Danish, Swedish and Finnish Twins." *Twin Research* 1: 196–205.

Kestenbaum, B. 1992. A Description of the Extreme Aged Population Based on Improved Medicare Enrollment Data. *Demography* 29(4): 565–580.

Kinosian, B., E. Stallard, J. Lee, M. A. Woodbury, A. Zbrozek, and H. A. Glick. 2000. Predicting 10-Year Outcomes of Alzheimer's Disease. *Journal of the American Geriatrics Society* 48(6): 631–638.

Kinosian, B., E. Stallard, K. G. Manton, D. L. Straley, A. Zbrozek, and H. A. Glick. 2004. "The Expected Outcomes and Costs of U.S. Patients with Incident Suspected Alzheimer's Disease (AD) over 15 Years." Abstract of Poster Session at the 9th International Conference on Alzheimer's Disease and Related Disorders, Philadelphia, PA, July 17–22.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin.

Manton, K. G., L. S. Corder, and E. Stallard. 1993. "Estimates of Change in Chronic Disability and Institutional Incidence and Prevalence Rates in the U.S. Elderly Population from the 1982, 1984, and 1989 National Long Term Care Survey." *Journal of Gerontology: Social Sciences* 48(4): S153–S166.

———. 1997. "Chronic Disability Trends in Elderly United States Populations: 1982–1994." *Proceedings of the National Academy of Sciences* 94: 2593–2598.

Manton, K. G., and X. Gu. 2001. "Changes in the Prevalence of Chronic Disability in the United States Black and Nonblack Population above Age 65 from 1982 to 1999." *Proceedings of the National Academy of Sciences* 98(11): 6354–6359.

Manton, K. G., and E. Stallard. 1984. *Recent Trends in Mortality Analysis*. Orlando, FL:

Academic Press.

Manton, K. G., E. Stallard, and B. Singer. 1992a. "Projecting the Future Size and Health Status of the U.S. Elderly Population." *International Journal of Forecasting* 8: 433–458.

Manton, K. G., E. Stallard, and J. W. Vaupel. 1986. "Alternative Models for the Heterogeneity of Mortality Risks among the Aged." *Journal of the American Statistical Association* 81(395): 635–64.

Manton, K. G., E. Stallard, and M. A. Woodbury. 1991. "A Multivariate Event History Model Based upon Fuzzy States: Estimation from Longitudinal Surveys with Informative Nonresponse." *Journal of Official Statistics* 7(3): 261–293.

Manton, K. G., E. Stallard, M. A. Woodbury, and J. E. Dowd. 1994a. "Time-Varying Covariates in Models of Human Mortality and Aging: Multidimensional Generalizations of the Gompertz." *Journal of Gerontology: BIOLOGICAL SCIENCES* 49(4): B169–B190.

Manton, K. G., M. A. Woodbury, E. Stallard, and L. S. Corder. 1992b. "The Use of Grade of Membership to Estimate Regression Relationships." In *Sociological Methodology, 1992,* ed. P. Marsden, pp. 321–381. London: Basil Blackwell.

Manton. K. G., M. A. Woodbury, and H. D. Tolley. 1994b. *Statistical Applications Using Fuzzy Sets*. New York: John Wiley & Sons.

McGinnis, J. M., and W. H. Foege. 1993. "Actual Causes of Death in the United States." *Journal of the American Medical Association* 270(18): 2207–2212.

McGue, M., J. W. Vaupel, N. Holm, and B. Harvald. 1993. "Longevity Is Moderately Heritable in a Sample of Danish Twins Born 1870–1880." *Journal of Gerontology: BIOLOGICAL SCIENCES* 48(6): B237–B244.

Mokdad, A. H., J. S. Marks, D. F. Stroup, and J. L. Gerberding. 2004. "Actual Causes of Death in the United States, 2000." *Journal of the American Medical Association* 291(10): 1238–1245.

Nagi, S. Z. 1976. "An Epidemiology of Disability among Adults in the United States." *Milbank Memorial Fund Quarterly—Health and Society* (fall): 439–467.

National Center for Health Statistics (NCHS). 1997. *U.S. Decennial Life Tables for 1989–91*.

Vol. 1, no. 1. Hyattsville, MD: NCHS.

Olshansky, S. J., and B. A. Carnes. 1997. "Ever since Gompertz." *Demography* 34(1): 1–15.

Perls, T. T., M. H. Silver, and J. F. Lauerman. 1999. *Living to 100*. New York: Basic Books.

Self, S. G., and K. Y. Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions." *Journal of the American Statistical Association* 82(398): 605–610.

Singpurwalla, N. D., and J. M. Booker. 2004. "Membership Functions and Probability Measures of Fuzzy Sets." *Journal of the American Statistical Association* 99(467): 867–889.

Stallard, E. 2000. "Retirement and Health: Estimates and Projections of Acute and Long-Term Care Needs and Expenditures of the U.S. Elderly Population." In *Retirement Needs Framework*, SOA Monograph M-RS00-1, chapter 15. Schaumburg, IL: Society of Actuaries.

———. 2001. "Estimates of the Incidence, Prevalence, Duration, Intensity, and Cost of Chronic Disability among the U.S. Elderly." Paper presented at the annual meeting of the Population Association of America, March 29–31, Washington, DC. Silver Spring, MD: PAA.

Stallard, E., and R. K. Yee. 2000. *Non-insured Home and Community-Based Long-Term Care Incidence and Continuance Tables.* Actuarial Report Issued by the Long-Term-Care Experience Committee. Schaumburg, IL: Society of Actuaries.

Strehler, B. L., and A. S. Mildvan. 1960. "General Theory of Mortality and Aging." *Science* 132(3418): 14–2.

Therneau, T. M., and P. M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Tolley, H. D., and K. G. Manton. 1991. "Intervention Effects among a Collection of Risks." *Transactions of the Society of Actuaries* 43: 443–468.

Vaupel, J. W., J. R. Carey, K. Christensen, T. E. Johnson, A. I. Yashin, N. V. Holm, I. A. Iachine, V. Kannisto, A. A. Khazaeli, P. Liedo, V. D. Longo, Y. Zeng, K. G. Manton, and J. W. Curtsinger. 1998. "Biodemographic Trajectories of Longevity." *Science* 280(5365): 855–860.

Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16(3): 439–454.

Wachter, K. W. 1999. "Grade of Membership Models in Low Dimensions." *Statistical Papers* 40: 439–457.

Wetterstrand, W. H. 1981. "Parametric Models for Life Insurance Mortality Data: Gompertz's Law over Time." *Transactions of the Society of Actuaries* 33: 443–468.

Wilks, S. S. 1938. "The Large Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *Annals of Mathematical Statistics* 9(1): 60–62.

Woodbury, M. A., and J. Clive. 1974. "Clinical Pure Types as a Fuzzy Partition." *Journal of Cybernetics* 4(3): 111–121.

Woodbury, M. A., J. Clive, and A. Garson. 1978. "Mathematical Typology: A Grade of Membership Technique for Obtaining Disease Definition." *Computers and Biomedical Research* 11(3): 277–298.

Woodbury, M. A., and K. G. Manton. 1977. "A Random-Walk Model of Human Mortality and Aging." *Theoretical Population Biology* 11(1): 37–48.

Woodbury, M. A., K. G. Manton, and H. D. Tolley. 1994. "A General Model for Statistical Analysis Using Fuzzy Sets: Sufficient Conditions for Identifiability and Statistical Properties." *Information Sciences* 1: 149–180.

Yashin, A. I., and I. A. Iachine. 1995. "Genetic Analysis of Duration: Correlated Frailty Model Applied to Survival of Danish Twins." *Genetic Epidemiology* 12: 529–538.

Yashin, A. I., I. A. Iachine, and J. R. Harris. 1999. "Half of the Variation in Susceptibility to Mortality Is Genetic: Findings from Swedish Twin Survival Data." *Behavior Genetics* 29(1): 11–19.

Zadeh, L. A. 1965. "Fuzzy Sets." *Information and Control* 8(3): 335–353.

Appendix

Table 1: Average GoM Scores in Four Pure-Type Model, by Age at Initial Interview

Males

		Average Gom Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	9,194	0.470	0.132	0.360	0.037
70-74	2,863	0.428	0.252	0.241	0.079
75-79	1,472	0.497	0.362	0.031	0.109
80-84	864	0.656	0.168	0.032	0.143
85-89	416	0.627	0.153	0.055	0.165
90-94	151	0.476	0.092	0.077	0.355
95-99	122	0.532	0.074	0.080	0.315
100-104	20	0.369	0.054	0.121	0.456
Total	15,102	0.480	0.179	0.273	0.068

Females

		Average Gom Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	11,355	0.639	0.203	0.118	0.040
70-74	3,854	0.553	0.330	0.076	0.041
75-79	2,469	0.540	0.288	0.109	0.063
80-84	1,792	0.626	0.206	0.104	0.064
85-89	1,223	0.509	0.044	0.348	0.099
90-94	525	0.402	0.031	0.219	0.349
95-99	575	0.416	0.037	0.123	0.424
100-104	97	0.354	0.028	0.118	0.500
Total	21,890	0.592	0.217	0.124	0.068

Table 2: Bernoulli Relative Variances of GoM Scores in Four Pure-Type Model, by Age at Initial Interview

Males

		Relative Variance of GoM Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	9,194	0.770	0.639	0.802	0.584
70-74	2,863	0.806	0.684	0.800	0.634
75-79	1,472	0.728	0.641	0.143	0.663
80-84	864	0.585	0.428	0.184	0.748
85-89	416	0.586	0.493	0.267	0.760
90-94	151	0.619	0.173	0.077	0.660
95-99	122	0.688	0.161	0.125	0.677
100-104	20	0.831	0.170	0.221	0.739
Total	15,102	0.759	0.645	0.793	0.660

Females

		Relative Variance of GoM Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	11,355	0.822	0.729	0.682	0.601
70-74	3,854	0.700	0.675	0.450	0.432
75-79	2,469	0.661	0.573	0.470	0.445
80-84	1,792	0.555	0.631	0.406	0.332
85-89	1,223	0.584	0.389	0.577	0.319
90-94	525	0.633	0.237	0.580	0.395
95-99	575	0.727	0.297	0.322	0.527
100-104	97	0.867	0.303	0.271	0.634
Total	21,890	0.742	0.688	0.597	0.540

Source: Author's calculations based on data from the NLTCS.

Table 3: GoM Transition and Vitality Matrices, Four Pure-Type Model, Males

	5-Year Transition Matrix U				Attained Age	Vitality Matrix V			
	To ...								
From ...	I	II	III	IV		I	II	III	IV
					65-69				
I	1.0000	0.0000	0.0000	0.0000		1.0000	0.0000	0.0000	0.0000
II	0.0000	1.0000	0.0000	0.0000		0.0000	1.0000	0.0000	0.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	1.0000	0.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					70-74				
I	1.0000	0.0000	0.0000	0.0000		1.0000	0.0000	0.0000	0.0000
II	0.0000	1.0000	0.0000	0.0000		0.0000	1.0000	0.0000	0.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	1.0000	0.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					75-79				
I	1.0000	0.0000	0.0000	0.0000		1.0000	0.0000	0.0000	0.0000
II	0.0000	0.5052	0.4948	0.0000		0.0000	0.5052	0.4948	0.0000
III	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					80-84				
I	0.7119	0.0000	0.2881	0.0000		0.7119	0.0000	0.2881	0.0000
II	0.0000	0.7891	0.0000	0.2109		0.0000	0.3986	0.0000	0.6014
III	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					85-89				
I	0.5299	0.0951	0.3750	0.0000		0.3772	0.0677	0.5551	0.0000
II	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					90-94				
I	0.4552	0.0658	0.4742	0.0047		0.1717	0.0378	0.6993	0.0912
II	0.0000	0.1916	0.1181	0.6903		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.9231	0.0769		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					95-99				
I	0.4503	0.1808	0.3610	0.0078		0.0773	0.0491	0.7524	0.1212
II	0.0000	0.4762	0.5102	0.0136		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.9597	0.0403		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					100-104				
I	0.0000	0.0000	0.3234	0.6766		0.0000	0.0000	0.5911	0.4089
II	0.0000	0.0000	0.3269	0.6731		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.7311	0.2689		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000

Source: Authors' calculations based on data from the NLTCS.

Table 4: GoM Transition and Vitality Matrices, Four Pure-Type Model, Females

	5-Year Transition Matrix U				Attained Age	Vitality Matrix V			
	To ...								
From ...	I	II	III	IV		I	II	III	IV
					65-69				
I	1.0000	0.0000	0.0000	0.0000		1.0000	0.0000	0.0000	0.0000
II	0.0000	1.0000	0.0000	0.0000		0.0000	1.0000	0.0000	0.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	1.0000	0.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					70-74				
I	1.0000	0.0000	0.0000	0.0000		1.0000	0.0000	0.0000	0.0000
II	0.0000	0.7844	0.2156	0.0000		0.0000	0.7844	0.2156	0.0000
III	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					75-79				
I	0.6020	0.3980	0.0000	0.0000		0.6020	0.3980	0.0000	0.0000
II	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	1.0000	0.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					80-84				
I	0.4199	0.1348	0.4453	0.0000		0.2528	0.4325	0.3147	0.0000
II	0.0000	0.8829	0.1171	0.0000		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					85-89				
I	0.6003	0.0000	0.3997	0.0000		0.1518	0.0000	0.8482	0.0000
II	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	1.0000	0.0000		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					90-94				
I	0.0000	0.0000	0.9571	0.0429		0.0000	0.0000	0.7355	0.2645
II	0.0000	0.0000	0.0506	0.9494		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.6958	0.3042		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					95-99				
I	0.0000	0.0000	0.0693	0.9307		0.0000	0.0000	0.6729	0.3271
II	0.0000	0.0000	0.0685	0.9315		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.9149	0.0851		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000
					100-104				
I	0.0000	0.0000	0.0494	0.9506		0.0000	0.0000	0.4553	0.5447
II	0.0000	0.0000	0.0493	0.9507		0.0000	0.0000	0.0000	1.0000
III	0.0000	0.0000	0.6765	0.3235		0.0000	0.0000	0.0000	1.0000
IV	0.0000	0.0000	0.0000	1.0000		0.0000	0.0000	0.0000	1.0000

Source: Authors' calculations based on data from the NLTCS.

Table 5: Probabilities of Death within One Year in Four Pure-Type GoM Model, by Attained Age at Time of Exposure

Males

		Annual Probability by Type					
Exposure Age	Number of Person-Years At Risk	I	II	III	IV	Observed Probability	Predicted Probability
65-69	25,250	0.001	0.004	0.077	0.152	0.030	0.030
70-74	39,563	0.004	0.002	0.159	0.258	0.043	0.043
75-79	31,291	0.008	0.003	0.189	0.354	0.067	0.067
80-84	19,170	0.000	0.104	0.189	0.323	0.105	0.106
85-89	8,117	0.000	0.000	0.218	0.296	0.154	0.154
90-94	2,728	0.000	0.000	0.240	0.351	0.228	0.229
95-99	793	0.298	1.000	0.144	0.452	0.301	0.301
100-104	155	0.269	0.269	0.269	0.522	0.400	0.401
Total	127,067	0.005	0.002	0.145	0.344	0.069	0.069

Females

		Annual Probability by Type					
Exposure Age	Number of Person-Years At Risk	I	II	III	IV	Observed Probability	Predicted Probability
65-69	31,095	0.003	0.001	0.085	0.141	0.017	0.018
70-74	53,631	0.003	0.003	0.091	0.161	0.027	0.027
75-79	48,498	0.004	0.000	0.083	0.191	0.043	0.043
80-84	35,563	0.000	0.000	0.091	0.185	0.070	0.071
85-89	20,404	0.000	0.032	0.087	0.218	0.115	0.115
90-94	9,577	0.000	—	0.100	0.282	0.183	0.183
95-99	3,804	—	—	0.101	0.378	0.264	0.264
100-104	992	—	—	0.043	0.472	0.325	0.325
Total	203,564	0.003	0.001	0.091	0.214	0.059	0.060

Note: "—" denotes cells with average GoM scores equal to zero for which probabilities can not be estimated.

Source: Author's calculations based on data from the NLTCS.

Table 6: Probabilities of Death within One Year in Four Pure-Type GoM Model, Adjusted for Declines in Vitality, by Attained Age at Time of Exposure

Males

		Annual Probability by Type					
Exposure Age	Number of Person-Years At Risk	I	II	III	IV	Observed Probability	Predicted Probability
65-69	25,250	0.001	0.004	0.077	0.152	0.030	0.030
70-74	39,563	0.004	0.002	0.159	0.258	0.043	0.043
75-79	31,291	0.008	0.095	0.354	0.354	0.067	0.067
80-84	19,170	0.054	0.236	0.323	0.323	0.105	0.106
85-89	8,117	0.121	0.296	0.296	0.296	0.154	0.154
90-94	2,728	0.200	0.351	0.351	0.351	0.228	0.229
95-99	793	0.235	0.452	0.452	0.452	0.301	0.301
100-104	155	0.372	0.522	0.522	0.522	0.400	0.401
Total	127,067	0.026	0.090	0.231	0.276	0.069	0.069

Females

		Annual Probability by Type					
Exposure Age	Number of Person-Years At Risk	I	II	III	IV	Observed Probability	Predicted Probability
65-69	31,095	0.003	0.001	0.085	0.141	0.017	0.018
70-74	53,631	0.003	0.022	0.161	0.161	0.027	0.027
75-79	48,498	0.003	0.083	0.191	0.191	0.043	0.043
80-84	35,563	0.028	0.185	0.185	0.185	0.070	0.071
85-89	20,404	0.074	0.218	0.218	0.218	0.115	0.115
90-94	9,577	0.148	0.282	0.282	0.282	0.183	0.183
95-99	3,804	0.192	0.378	0.378	0.378	0.264	0.264
100-104	992	0.277	0.472	0.472	0.472	0.325	0.325
Total	203,564	0.026	0.102	0.178	0.186	0.059	0.060

Source: Author's calculations based on data from the NLTCS.

Table 7: Average Adjusted GoM Scores at Initial Interview in Four Pure-Type Model, by Age at Initial Interview

Males

		Average GoM Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	9,194	0.470	0.132	0.360	0.037
70-74	2,863	0.428	0.252	0.241	0.079
75-79	1,472	0.497	0.183	0.179	0.140
80-84	864	0.467	0.067	0.189	0.277
85-89	416	0.237	0.042	0.348	0.373
90-94	151	0.082	0.018	0.333	0.567
95-99	122	0.041	0.026	0.400	0.533
100-104	20	0.000	0.000	0.218	0.782
Total	15,102	0.450	0.152	0.310	0.088

Females

		Average GoM Score by Type			
Initial Age	Number of Respondents	I	II	III	IV
65-69	11,355	0.639	0.203	0.118	0.040
70-74	3,854	0.553	0.259	0.071	0.117
75-79	2,469	0.325	0.215	0.288	0.172
80-84	1,792	0.158	0.271	0.197	0.374
85-89	1,223	0.077	0.000	0.432	0.491
90-94	525	0.000	0.000	0.295	0.705
95-99	575	0.000	0.000	0.280	0.720
100-104	97	0.000	0.000	0.161	0.839
Total	21,890	0.483	0.197	0.162	0.158

Source: Author's calculations based on data from the NLTCS.

Table 8: Average Adjusted GoM Scores at Initial Interview in Four Pure-Type Model, by HIPAA ADL Count at Initial Interview

Males

		Average GoM Score by Type			
Number of HIPAA ADLs[1]	Number of Respondents	I	II	III	IV
0	13,587	0.489	0.156	0.331	0.024
1	381	0.212	0.207	0.222	0.358
2	225	0.130	0.160	0.196	0.514
3	145	0.122	0.148	0.136	0.594
4	193	0.082	0.087	0.078	0.753
5	276	0.044	0.057	0.040	0.859
6	290	0.012	0.012	0.008	0.968
Total	15,097	0.450	0.152	0.310	0.088

Females

		Average GoM Score by Type			
Number of HIPAA ADLs	Number of Respondents	I	II	III	IV
0	18,332	0.560	0.225	0.170	0.046
1	933	0.180	0.120	0.245	0.454
2	452	0.134	0.073	0.201	0.591
3	370	0.089	0.055	0.131	0.725
4	411	0.057	0.038	0.076	0.829
5	596	0.032	0.021	0.033	0.914
6	786	0.004	0.005	0.007	0.984
Total	21,880	0.483	0.197	0.162	0.158

Note 1: Number of ADLs for which the respondent requires either standby or active personal assistance. See Stallard and Yee (2000) for details.

Source: Author's calculations based on data from the NLTCS.

Table 9: Average Adjusted GoM Scores at Initial Interview in Four Pure-Type Model, by IADL Count at Initial Interview

Males

		Average GoM Score by Type			
Number of IADLs[1]	Number of Respondents	I	II	III	IV
0	12,638	0.505	0.152	0.336	0.008
1	509	0.335	0.242	0.271	0.152
2	235	0.256	0.259	0.250	0.235
3	210	0.203	0.221	0.264	0.312
4	177	0.192	0.194	0.219	0.395
5	183	0.166	0.164	0.215	0.454
6	168	0.132	0.156	0.151	0.562
7	150	0.094	0.090	0.098	0.718
8	131	0.056	0.057	0.068	0.819
9	176	0.021	0.016	0.031	0.933
Total	14,577	0.464	0.155	0.317	0.064

Females

		Average GoM Score by Type			
Number of IADLs	Number of Respondents	I	II	III	IV
0	16,274	0.600	0.232	0.154	0.015
1	830	0.329	0.221	0.274	0.176
2	626	0.291	0.153	0.269	0.287
3	634	0.226	0.140	0.257	0.376
4	478	0.155	0.108	0.268	0.470
5	312	0.135	0.081	0.225	0.560
6	320	0.094	0.066	0.176	0.663
7	254	0.050	0.047	0.121	0.782
8	257	0.030	0.033	0.088	0.849
9	268	0.007	0.008	0.046	0.940
Total	20,253	0.520	0.210	0.167	0.103

Note 1: Number of IADLs for which the respondent requires assistance because of a disability or health problem. See Stallard and Yee (2000) for details.

Source: Author's calculations based on data from the NLTCS.

Figure 1 – Probability of Death Within One Year
Males

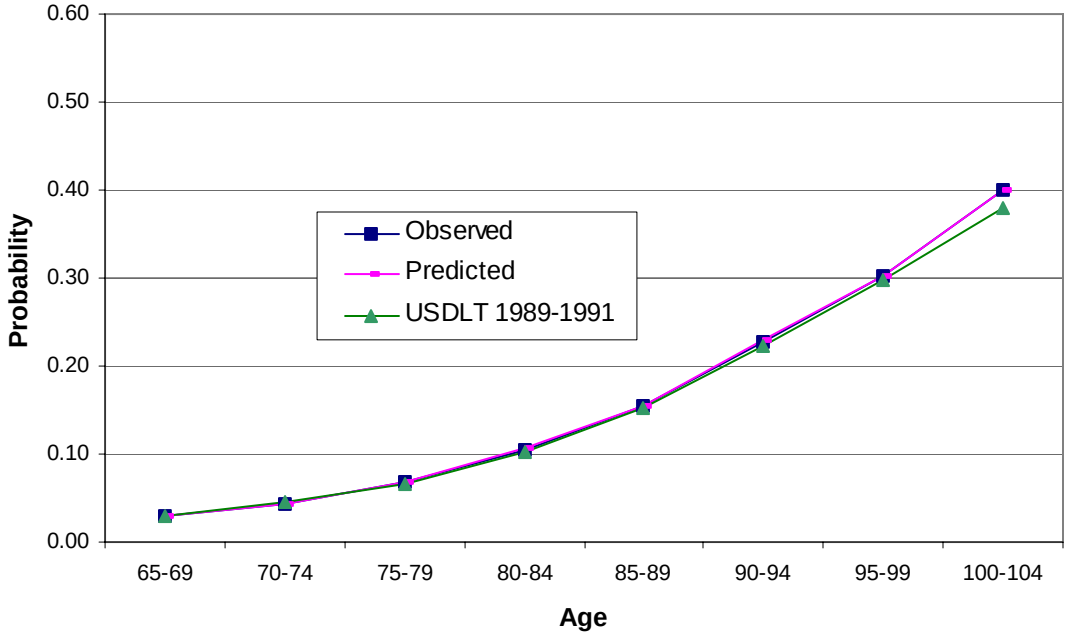


Figure 2 – Probability of Death Within One Year
Females

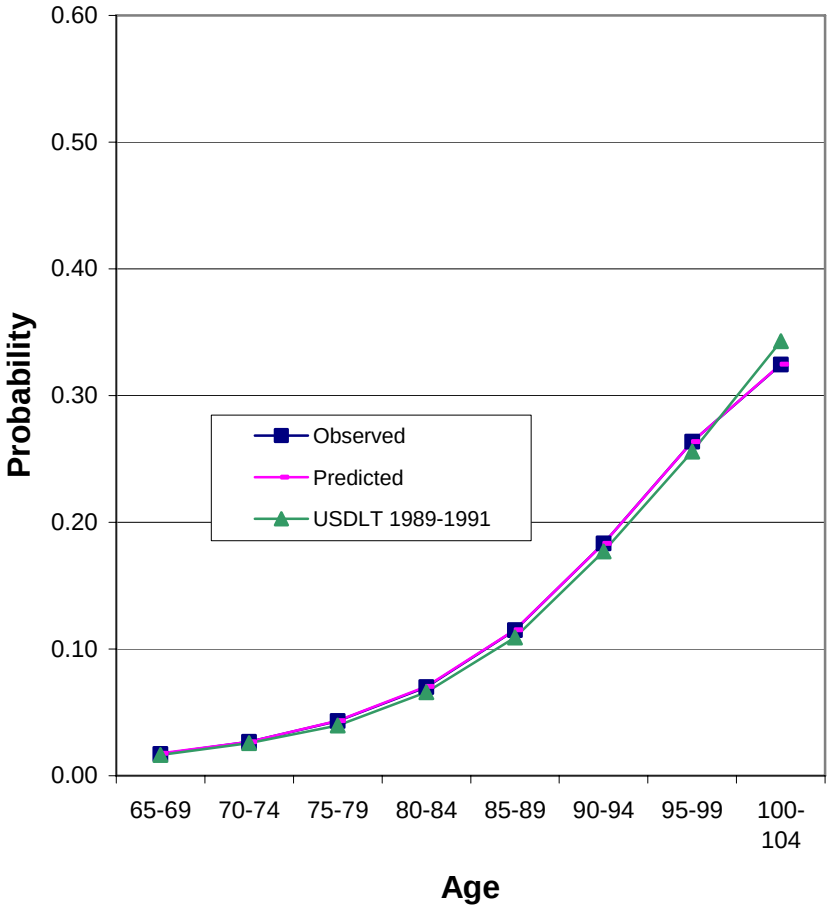


Figure 3 – Adjusted Annual Probabilities of Death in Four Pure-Type Model, Males

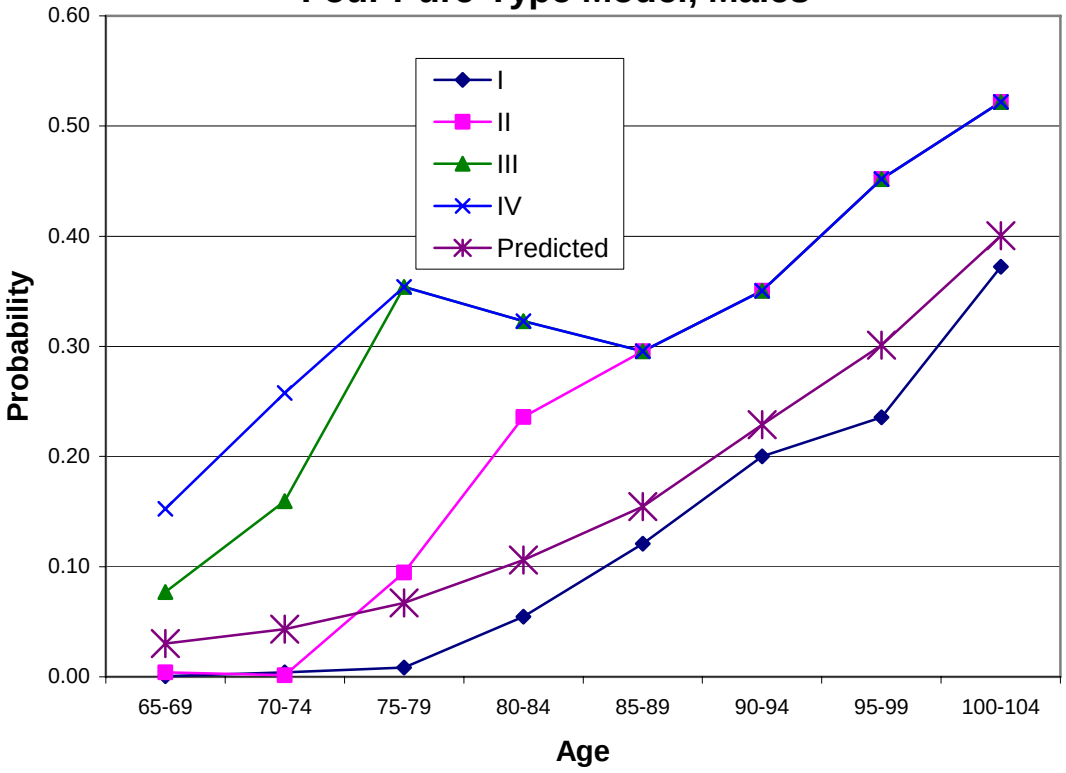


Figure 4 – Adjusted Annual Probabilities of Death in Four Pure-Type Model, Females

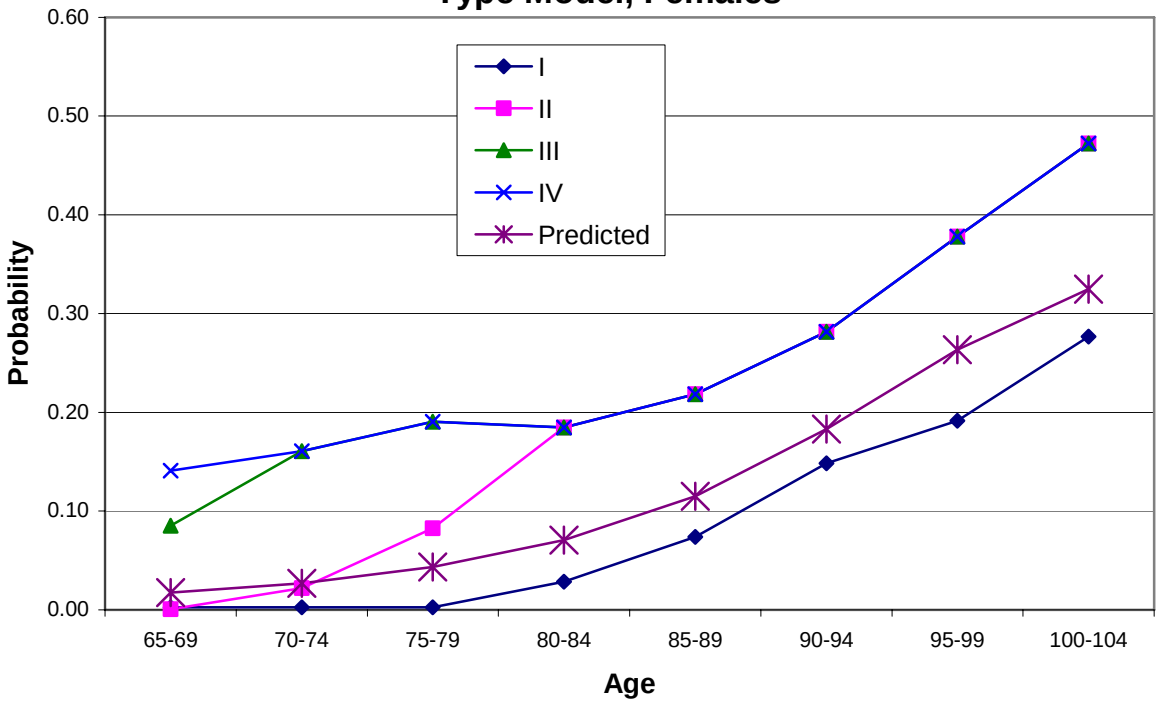


Figure 5 – Unadjusted Annual Probabilities of Death in Four Pure-Type Model, Males

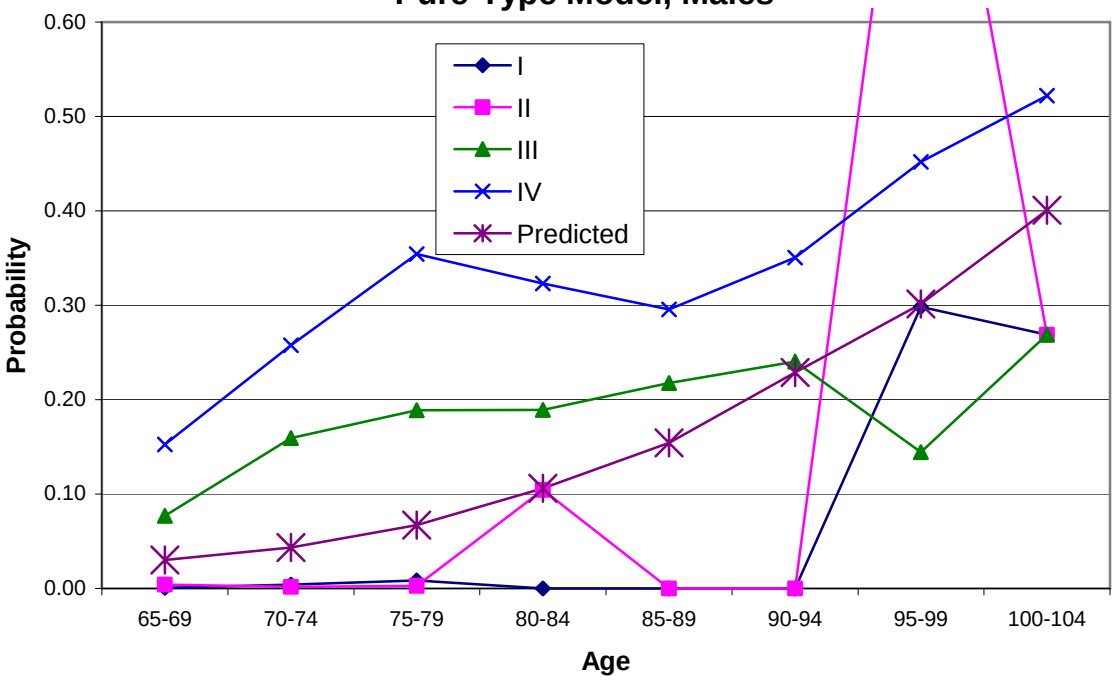


Figure 6 – Unadjusted Annual Probabilities of Death in Four Pure-Type Model, Females

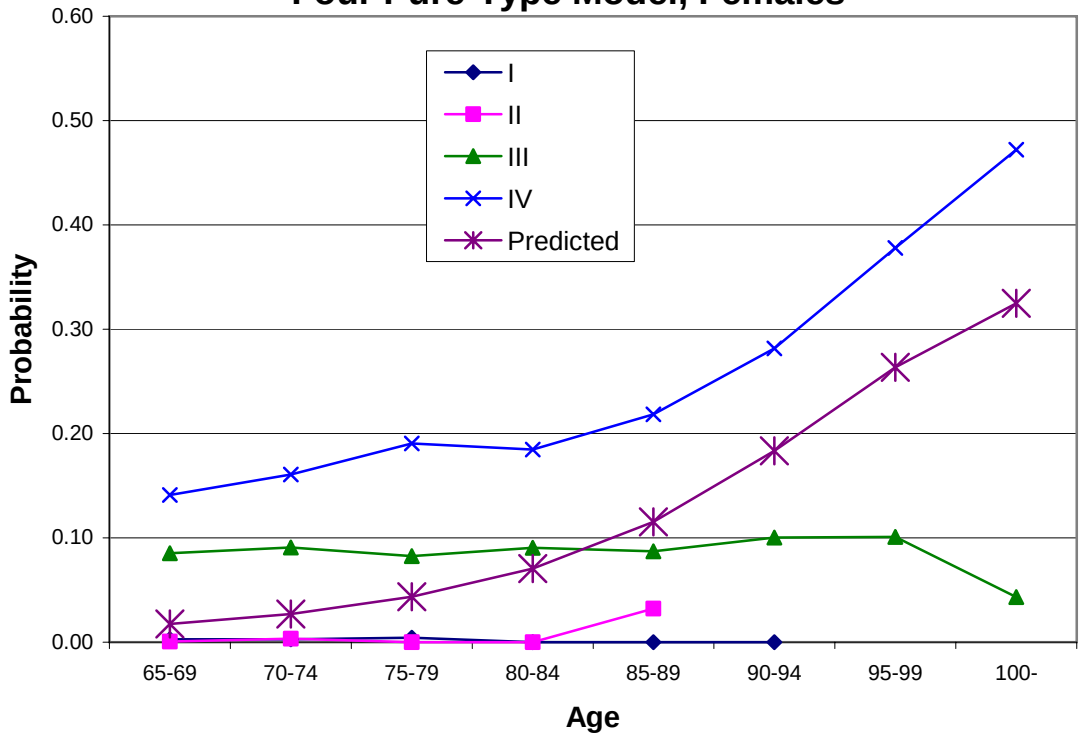


Figure 7 – Adjusted Age-Specific GoM-Score Distribution, Males

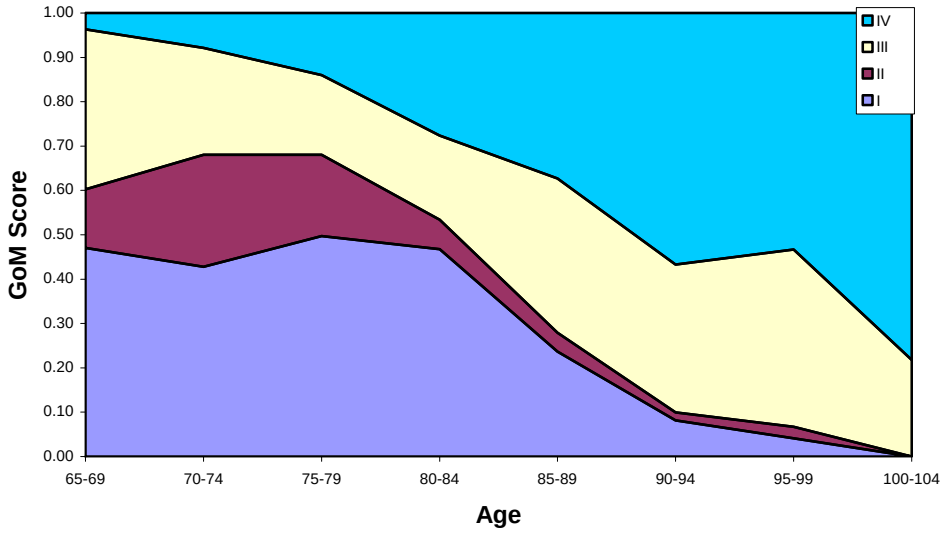


Figure 8 – Adjusted Age-Specific GoM-Score Distribution, Females

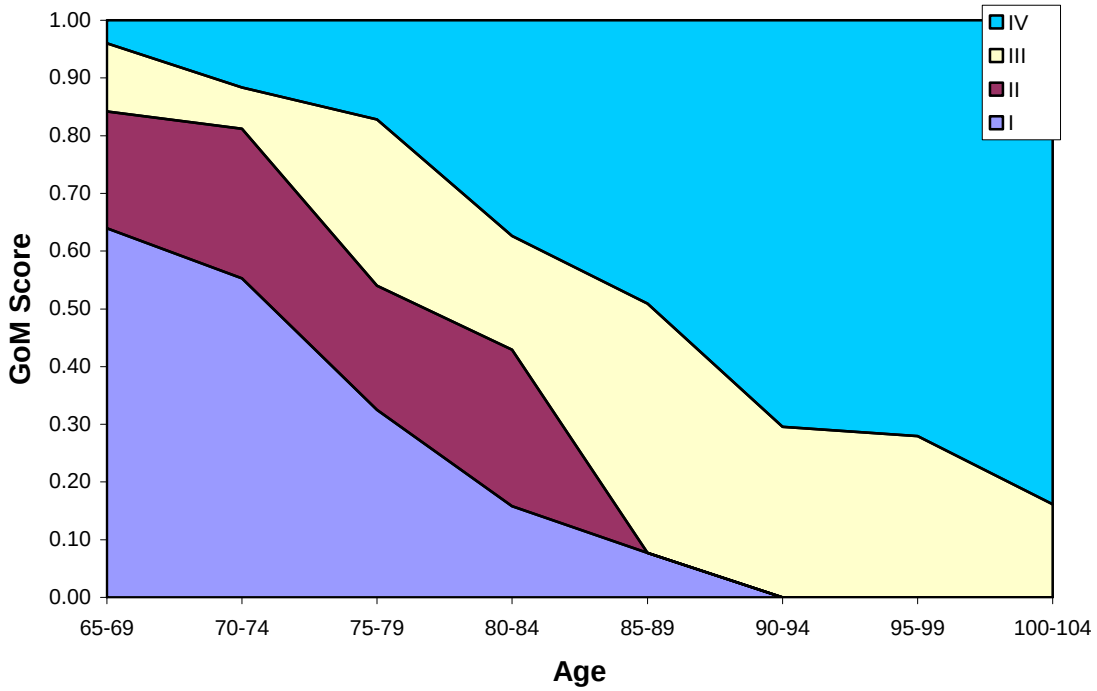
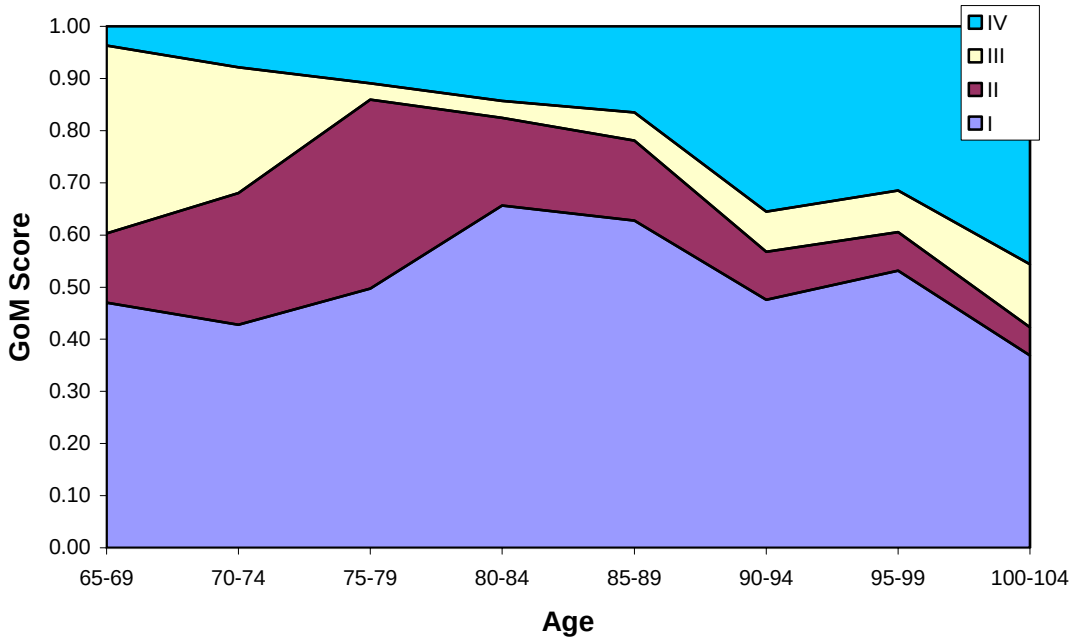


Figure 9 – Unadjusted Age-Specific GoM-Score Distribution, Males



Appendix: NLTCS Variables, Log-Likelihood Values, and Chi-Squared Statistics by Variable, by Sex

			Males					Females					
#	Variable Name/Description	Number of Response Levels	Number of Respond-ents	Log of Likelihood-Ratio for K=4 vs. K=1	Approximate Chi-Squared Statistic for K=4 vs. K=1	d.f for Testing K=4 vs. K=1[2]	Chi-Squared per d.f.	Rank Order of Chi-Squared per d.f.	Number of Respond-ents	Log of Likelihood-Ratio for K=4 vs. K=1	Approximate Chi-Squared Statistic for K=4 vs. K=1	d.f for Testing K=4 vs. K=1[2]	Chi-Squared per d.f.	Rank Order of Chi-Squared per d.f.
1	5-Year Survival Status[1]	2	24,207	6,394.70	12,789.41	3	4,263.14	1	38,280	7,764.52	15,529.04	3	5,176.35	5
2	Respondent is Proxy	2	8,329	756.25	1,512.49	3	504.16	20	17,633	2,077.06	4,154.13	3	1,384.71	18
3	Race	3	39,047	2,210.77	4,421.54	6	736.92	16	59,687	953.50	1,906.99	6	317.83	41
4	Residence Type: Institutional vs.non-institutional	2	27,819	2,519.16	5,038.32	3	1,679.44	10	43,764	6,303.42	12,606.84	3	4,202.28	6
5	Height	5	3,400	47.63	95.26	12	7.94	91	6,392	58.77	117.53	12	9.79	93
6	Body Mass Index – Current BMI class	4	3,349	346.26	692.52	9	76.95	65	6,218	679.52	1,359.04	9	151.00	59
7	Body Mass Index – BMI class at age 50 years	4	3,113	188.82	377.63	9	41.96	73	5,635	246.35	492.71	9	54.75	74
8	Body Mass Index – BMI class 12 months prior to interview	4	3,250	333.92	667.85	9	74.21	66	5,983	647.74	1,295.48	9	143.94	61
9	Alcohol use	3	3,483	186.70	373.40	6	62.23	69	6,578	222.67	445.33	6	74.22	67
10	Cigarette use	3	3,491	39.01	78.01	6	13.00	88	6,588	30.19	60.38	6	10.06	92
11	Exercise – Hours/minutes of vigorous activities	5	3,424	437.57	875.13	12	72.93	67	6,450	411.00	821.99	12	68.50	68
12	Exercise – Hours/minutes of moderate activities	5	3,397	784.42	1,568.85	12	130.74	51	6,410	1,193.40	2,386.81	12	198.90	53
13	Exercise – Hours/minutes of light activities	5	3,364	558.69	1,117.38	12	93.12	61	6,291	824.12	1,648.25	12	137.35	63
14	Medical – Rheumatism or arthritis	2	7,026	274.71	549.41	3	183.14	45	13,514	329.91	659.83	3	219.94	51
15	Medical – Other permanent numbness or stiffness	2	7,021	399.75	799.49	3	266.50	32	13,487	456.86	913.72	3	304.57	42
16	Medical – Paralysis	2	7,030	359.39	718.78	3	239.59	37	13,518	429.43	858.86	3	286.29	46
17	Medical – Multiple sclerosis	2	7,028	11.40	22.79	3	7.60	92	13,520	30.28	60.55	3	20.18	87
18	Medical – Cerebral palsy	2	7,029	8.07	16.14	3	5.38	94	13,517	12.59	25.18	3	8.39	94
19	Medical – Epilepsy	2	7,023	16.59	33.18	3	11.06	90	13,514	28.17	56.35	3	18.78	88
20	Medical – Parkinson's disease	2	7,028	79.51	159.02	3	53.01	70	13,510	86.21	172.43	3	57.48	72
21	Medical – Glaucoma	2	7,023	28.96	57.92	3	19.31	80	13,500	64.62	129.23	3	43.08	79
22	Medical – Diabetes	2	7,026	101.99	203.98	3	67.99	68	13,509	296.11	592.22	3	197.41	54
23	Medical – Cancer	2	7,023	19.86	39.72	3	13.24	87	13,499	23.93	47.86	3	15.95	89
24	Medical – Frequent constipation	2	7,002	312.58	625.16	3	208.39	40	13,485	455.52	911.04	3	303.68	43
25	Medical – Frequent trouble sleeping	2	7,017	500.56	1,001.12	3	333.71	26	13,491	542.75	1,085.49	3	361.83	39
26	Medical – Frequent severe headaches	2	7,010	265.02	530.05	3	176.68	46	13,496	410.22	820.44	3	273.48	47
27	Medical – Obesity or medically overweight	2	7,021	221.50	443.00	3	147.67	49	13,490	834.00	1,668.01	3	556.00	31
28	Medical – Arteriosclerosis or hardening of the arteries	2	6,956	284.93	569.85	3	189.95	42	13,406	449.35	898.71	3	299.57	45
29	Medical – A heart attack in 12 months prior to interview	2	7,010	62.95	125.90	3	41.97	72	13,477	84.37	168.74	3	56.25	73
30	Medical – Any other heart problem in 12 months prior to interview	2	7,012	200.93	401.87	3	133.96	50	13,487	322.71	645.42	3	215.14	52
31	Medical – Hypertension or high blood pressure in 12 months prior to interview	2	7,012	187.94	375.89	3	125.30	52	13,472	224.95	449.89	3	149.96	60
32	Medical – A stroke in 12 months prior to interview	2	7,010	171.47	342.94	3	114.31	55	13,471	285.10	570.21	3	190.07	55
33	Medical – Circulation trouble in arms or leg in 12 months prior to interview	2	7,003	812.96	1,625.92	3	541.97	19	13,464	955.82	1,911.65	3	637.22	25
34	Medical – Pneumonia in 12 months prior to interview	2	7,010	118.15	236.30	3	78.77	63	13,457	92.66	185.33	3	61.78	71
35	Medical – Flu or influenza in 12 months prior to interview	2	7,012	117.40	234.80	3	78.27	64	13,471	208.18	416.37	3	138.79	62
36	Medical – Bronchitis in 12 months prior to interview	2	7,012	178.36	356.73	3	118.91	53	13,474	259.99	519.98	3	173.33	56
37	Medical – Emphysema in 12 months prior to interview	2	7,015	262.15	524.29	3	174.76	47	13,471	94.07	188.13	3	62.71	70
38	Medical – Asthma in 12 months prior to interview	2	7,015	168.26	336.52	3	112.17	56	13,483	187.80	375.60	3	125.20	64
39	Medical – A broken hip in 12 months prior to interview	2	7,012	28.21	56.42	3	18.81	81	13,482	68.42	136.84	3	45.61	76
40	Medical – Other broken bones in 12 months prior to interview	2	7,007	17.83	35.66	3	11.89	89	13,469	23.47	46.95	3	15.65	90
41	Medical – Senility	2	7,181	420.53	841.06	3	280.35	31	14,176	847.25	1,694.50	3	564.83	30
42	Medical – Alzheimer's disease	2	5,175	148.37	296.74	3	98.91	58	10,342	342.81	685.61	3	228.54	50
43	Medical – Mental Retardation	2	7,182	55.46	110.92	3	36.97	75	14,180	73.10	146.20	3	48.73	75
44	See well enough to read newspaper	2	6,965	435.73	871.46	3	290.49	29	13,413	737.90	1,475.81	3	491.94	35
45	Subjective Health Status	4	6,543	1,392.69	2,785.38	9	309.49	27	12,782	1,649.06	3,298.12	9	366.46	38
46	ADL Personal Assistance Level – Bathing	6	27,814	7,258.32	14,516.65	15	967.78	12	43,753	14,783.77	29,567.54	15	1,971.17	11
47	ADL Personal Assistance Level – Dressing	6	27,814	5,322.59	10,645.19	15	709.68	17	43,753	10,444.85	20,889.71	15	1,392.65	16
48	ADL Personal Assistance Level – Toileting	6	27,814	4,947.94	9,895.88	15	659.73	18	43,753	10,416.89	20,833.77	15	1,388.92	17
49	ADL Personal Assistance Level – Transferring in/out bed	6	27,814	6,046.80	12,093.61	15	806.24	14	43,753	12,178.80	24,357.60	15	1,623.84	14
50	ADL Personal Assistance Level – Eating	6	27,814	2,789.04	5,578.09	15	371.87	25	43,753	5,451.76	10,903.53	15	726.90	22
51	ADL Personal Assistance Level – Continence	4	27,814	3,369.48	6,738.97	9	748.77	15	43,753	6,825.72	13,651.45	9	1,516.83	15
52	ADL Personal Assistance Level – Indoor mobility	6	27,814	6,802.78	13,605.56	15	907.04	13	43,753	13,720.91	27,441.82	15	1,829.45	12

#	Variable Name/Description	Number of Response Levels	Males: Number of Respondents	Males: Log of Likelihood-Ratio for K=4 vs. K=1	Males: Approximate Chi-Squared Statistic for K=4 vs. K=1	Males: d.f for Testing K=4 vs. K=1[2]	Males: Chi-Squared per d.f.	Males: Rank Order of Chi-Squared per d.f.	Females: Number of Respondents	Females: Log of Likelihood-Ratio for K=4 vs. K=1	Females: Approximate Chi-Squared Statistic for K=4 vs. K=1	Females: d.f for Testing K=4 vs. K=1[2]	Females: Chi-Squared per d.f.	Females: Rank Order of Chi-Squared per d.f.
53	IADL Limitations – Light housework	2	26,528	3,832.14	7,664.27	3	2,554.76	7	39,646	4,919.21	9,838.43	3	3,279.48	9
54	IADL Limitations – Laundry	2	26,528	4,996.93	9,993.87	3	3,331.29	4	39,646	7,865.91	15,731.83	3	5,243.94	4
55	IADL Limitations – Cooking	2	26,528	4,201.96	8,403.93	3	2,801.31	6	39,646	5,940.95	11,881.89	3	3,960.63	7
56	IADL Limitations – Grocery shopping	2	26,528	5,444.45	10,888.91	3	3,629.64	3	39,646	10,303.69	20,607.37	3	6,869.12	2
57	IADL Limitations – Outside mobility	2	26,758	5,876.35	11,752.70	3	3,917.57	2	40,447	10,827.49	21,654.99	3	7,218.33	1
58	IADL Limitations – Travel	2	26,528	4,893.02	9,786.04	3	3,262.01	5	39,646	10,024.35	20,048.70	3	6,682.90	3
59	IADL Limitations – Managing money	2	26,528	3,103.56	6,207.13	3	2,069.04	9	39,646	5,035.84	10,071.67	3	3,357.22	8
60	IADL Limitations – Taking medicines	2	26,528	3,375.14	6,750.29	3	2,250.10	8	39,646	4,544.16	9,088.32	3	3,029.44	10
61	IADL Limitations – Phoning	2	26,528	2,239.50	4,479.01	3	1,493.00	11	39,646	2,711.92	5,423.84	3	1,807.95	13
62	Functional Limitations – Climbing 1 flight of stairs	4	6,590	2,063.74	4,127.47	9	458.61	22	12,625	2,786.30	5,572.59	9	619.18	26
63	Functional Limitations – Bending to put on socks or stockings	4	6,933	1,804.96	3,609.92	9	401.10	23	13,287	2,776.28	5,552.57	9	616.95	27
64	Functional Limitations – Lifting and holding a 10 lb. package	4	6,914	2,174.98	4,349.96	9	483.33	21	13,240	3,287.98	6,575.96	9	730.66	21
65	Functional Limitations – Reaching above head	4	6,983	1,179.12	2,358.25	9	262.03	33	13,416	1,996.57	3,993.13	9	443.68	36
66	Functional Limitations – Combing or brushing hair	4	6,998	966.05	1,932.10	9	214.68	39	13,450	2,227.36	4,454.71	9	494.97	33
67	Functional Limitations – Washing hair	4	6,990	1,701.00	3,402.00	9	378.00	24	13,420	3,432.01	6,864.01	9	762.67	19
68	Functional Limitations – Using fingers to grasp and handle small objects	4	6,987	933.87	1,867.74	9	207.53	41	13,439	1,355.24	2,710.49	9	301.17	44
69	SPMSQ – What is the date today?	2	5,361	421.15	842.31	3	280.77	30	11,874	1,131.49	2,262.97	3	754.32	20
70	SPMSQ – What day of week is this?	2	5,354	363.09	726.17	3	242.06	35	11,860	877.09	1,754.19	3	584.73	29
71	SPMSQ – What is your street address?	2	5,355	458.50	917.01	3	305.67	28	11,847	1,042.45	2,084.90	3	694.97	23
72	SPMSQ – In what State is this?	2	4,401	171.99	343.99	3	114.66	54	9,959	523.20	1,046.39	3	348.80	40
73	SPMSQ – How old are you?	2	4,407	348.36	696.71	3	232.24	38	9,961	752.45	1,504.90	3	501.63	32
74	SPMSQ – When were you born? (month, day, year)	2	5,132	252.52	505.03	3	168.34	48	11,295	739.94	1,479.89	3	493.30	34
75	SPMSQ – Who is the President of the United States now?	2	4,404	362.37	724.74	3	241.58	36	9,958	999.07	1,998.14	3	666.05	24
76	SPMSQ – Who was the President just before him?	2	4,403	380.92	761.84	3	253.95	34	9,960	905.36	1,810.71	3	603.57	28
77	SPMSQ – What was your mother's maiden name?	2	4,385	142.71	285.43	3	95.14	59	9,923	374.38	748.75	3	249.58	48
78	SPMSQ – Subtract 3 from 20 & keep subtracting ...	2	4,409	280.67	561.34	3	187.11	44	9,951	645.13	1,290.26	3	430.09	37
79	Behavior – Lose temper & throw, kick, slam, destroy things	3	6,951	93.21	186.43	6	31.07	76	13,374	64.07	128.14	6	21.36	86
80	Behavior – Lose your way and not find your way back	2	6,964	118.33	236.66	3	78.89	62	13,385	102.56	205.13	3	68.38	69
81	Behavior – Take anything not yours without realizing	2	6,955	31.17	62.34	3	20.78	77	13,375	66.23	132.46	3	44.15	77
82	Behavior – Forget to do important things like eating	2	6,956	283.86	567.72	3	189.24	43	13,369	372.83	745.67	3	248.56	49
83	Memory – List as many animals as possible in one minute	4	723	28.66	57.31	9	6.37	93	1,341	54.77	109.53	9	12.17	91
84	Memory – Delayed 12-Word Recall	12	714	51.32	102.63	33	3.11	95	1,314	86.09	172.18	33	5.22	95
85	MMSE – Orientation: Day, date, month, year, season	6	1,022	104.66	209.31	15	13.95	85	2,115	259.25	518.50	15	34.57	82
86	MMSE – Orientation: Country, city, street, floor #, address	6	1,022	145.88	291.77	15	19.45	79	2,115	323.25	646.49	15	43.10	78
87	MMSE – Registration: 3-word memory	4	1,022	63.46	126.92	9	14.10	84	2,115	121.99	243.98	9	27.11	84
88	MMSE – Attention: Subtract 7 from 100 & keep subtracting ...	6	1,022	134.62	269.25	15	17.95	82	2,115	224.85	449.70	15	29.98	83
89	MMSE – Recall: 3-word memory	4	1,022	65.48	130.97	9	14.55	83	2,115	156.91	313.82	9	34.87	81
90	MMSE – Language: Point and name	3	1,022	40.47	80.94	6	13.49	86	2,115	74.86	149.72	6	24.95	85
91	MMSE – Language: Repeat phrase	2	1,022	59.35	118.70	3	39.57	74	2,115	121.89	243.79	3	81.26	66
92	MMSE – Language: 3-stage command	4	1,022	89.98	179.95	9	19.99	78	2,115	170.32	340.64	9	37.85	80
93	MMSE – Language: Read and obey	2	1,022	69.27	138.54	3	46.18	71	2,115	143.46	286.93	3	95.64	65
94	MMSE – Language: Sentence writing	2	1,022	140.30	280.59	3	93.53	60	2,115	232.93	465.86	3	155.29	58
95	MMSE – Language: Figure drawing	2	1,022	161.65	323.29	3	107.76	57	2,115	253.33	506.65	3	168.88	57
	Total	285	922,303	112,656	225,312	570	395.28		1,587,710	196,998	393,996	570	691.22	

Note 1: 5-Year Survival Status was included in the analysis to code for missing data due to death at follow-up. For this one variable the predicted probabilities were computed using GoM scores at both the start and end of each follow-up interval to approximate the GoM score changes during the interval. A separate survival analysis with age-specific annual probabilities of death was conducted using the outputs of the initial analysis. See text for details.

Note 2: Incremental degrees of freedom refer to the difference in the number of λ-values between the 4 pure-type model and the 1 pure-type model for the indicated variable. The total of the incremental degrees of freedom does not include 45,306 degrees of freedom for the GoM scores and 42 degrees of freedom for the transition matrix parameters. The chi-squared statistics indicate the incremental effect of each variable assuming that it was the last variable added to the model. Equivalently, the chi-squared statistics indicate the effect of constraining the λ-values for each variable to be equal across the 4 pure types of the 4 pure-type model.

Source: Author's calculations based on data from the NLTCS.