Estimation of Multi-State Life Table Functions and Their Variability Using the SPACE Program*

Liming Cai, James Lubitz, Mark D. Hayward,
Yasuhiko Saito, Aaron Hagedorn and Eileen Crimmins

* Liming Cai and James Lubitz, Office of Analysis and Epidemiology, National Center for Health Statistics. Mark Hayward, Population Research Center and Department of Sociology, University of Texas at Austin. Yasuhiko Saito, Advanced Research Institute for the Sciences and Humanities, Nihon University, Japan. Aaron Hagedorn and Eileen Crimmins, Davis School of Gerontology, University of Southern California. Address correspondence to: Liming Cai, 3311 Toledo Road, Room 6330, Hyattsville, MD 20782; Email: lcai@cdc.gov.

The multi-state life table (MSLT) model is a "time-inhomogeneous, finite-space, continuous-time" first-order Markov model (Schoen 1988:64). Demographers frequently use it to analyze expected duration in various states when a stochastic process involves multiple and recurrent events, such as functional limitations (Crimmins, Hayward and Saito 1994; 1996; Land, Guralnik and Blazer 1994), HIV/AIDS (Palloni 1996), labor force participation (Hayward and Grady 1990; Hayward, Grady and McLaughlin 1988), cohabitation and marriage (Bumpass and Lu 2000; Espenshade and Braun 1982; Hofferth 1985; Schoen and Land 1979), poverty (Duncan and Rodgers 1988), living in poor neighborhoods (Quillian 2003). It has also been used in studies linking longevity to medical spending for the aging population (Goldman et al. 2005; Lubitz et al. 2003). When the MSLT model was originally developed, life tables were calculated using population-level rates – hence there was limited attention given to estimation techniques and variability in the life table functions. Increasingly, however, the inputs to the life table (i.e., transition rates or probabilities) are being derived from panel data obtained via large-scale survey sampling. Sampling variability is thus introduced into estimates of MSLT functions and it becomes important to measure its magnitude.

Estimating sampling variability is important for hypothesis testing of group differences in MSLT functions such as life expectancy and survivorship. MSLT functions arise from a complex set of transitions and group differences in these functions may occur in ways that are not immediately obvious from analyzing parameter estimates in event-specific models in traditional event history modeling. For example, suppose one wants to test the hypothesis that males and females differ in health over the life cycle, where health is measured by health expectancy (i.e. expected years in various health states) and trajectories of expected experiences, such as incidences and recoveries from functional limitations. This hypothesis is global in the sense that it takes into account sex differences in *all* of the transitions defining the life table model of life cycle health. How sex is associated with an overall process made up of multiple events, however, may be unclear for a number of reasons: 1) sex may significantly affect some transitions and not others; 2) sex effects, even though statistically significant, may be offsetting; and 3) sex effects may be non-significant for the whole set of transitions, yet the consistency of effects for a lengthy period of time may combine in a way in which the sex effect is reinforced and magnified with age. Since a covariate such as sex may affect life cycle health in many ways, we need a way to statistically evaluate sex differences in the *overall* process.

From a methodological standpoint, an appropriate variance estimation procedure must also take into account the complex data structure of panel surveys (e.g., the Medicare Current Beneficiary Survey, the Established Populations for Epidemiologic Studies of the Elderly survey, the National Long-Term Care Survey and the Longitudinal Studies of Aging). These longitudinal surveys are the primary data source for many MSLT applications (e.g., Cai and Lubitz 2007; Crimmins et al. 1994, 1996; Goldman et al. 2005; Guralnik et al. 1993; Lubitz et al. 2003; Manton, Corder and Stallard 1993). They all have design elements such as stratification and multi-stage clustering with opposing effect on variance estimates. If these design factors are not adequately controlled, variance estimates for MSLT estimates will be incorrect and statistical inferences will be invalid (Lohr 1999).

In order to address this problem of variance estimation, we provide researchers with a new statistical program, the SPACE (Stochastic Population Analysis for Complex Events) program, which derives consistent MSLT variance estimates from complex survey samples. It joins two other publically available programs – the IMaCh (**I**nterpolated **Ma**rkov **Ch**ain) program and the GSMLT (Gibbs Sampling Multistate Life Table) program –developed for the estimation of MSLT functions and their sampling variability. The IMaCh program was developed on the basis of Lièvre et al. (2003), and has been used in a number of recent studies (e.g., Al Snih et al. 2007; Jagger et al. 2007; Reynolds, Saito and Crimmins 2005). The variance of health expectancy estimates is derived from the variance of the transition probabilities. The GSMLT program was developed on the basis of Lynch and Brown (2005), which adopts a Bayesian approach to estimating MSLT. The variance of life table estimates is derived from samples of their posterior distributions. Since the GSMLT program does not address the issue of sampling design, the variance estimates are incorrect when complex survey data are used and thus the GSMLT program is not compared with the SPACE program in this study.[1]

There are a number of important differences between the SPACE program and the IMaCh program. First, they differ in their treatment of the design factors of the

---

[1] There is another publically available program to estimate MSLT health expectancy (Weden, 2005, downloadable from http://www.ssc.wisc.edu/~mweden/). Since this program does not produce variance estimates, we will not discuss it here.

survey data. The IMaCh program assumes that the survey design affects the sample weight only. Sample weight is the inverse of the probability of the selection of each sampled *individual*. In surveys with single- or multi-stage clustering, one must also account for the selection probabilities of the *clusters*. The problem of variance overestimation in clustered survey data cannot be corrected by using sample weight alone (Lohr 1999). The SPACE program, on the other hand, uses a version of the rescaling bootstrap method developed specifically for complex surveys (Rao and Wu 1988; Sitter 1992a). It produces consistent variance estimates by resampling clusters within each stratum, and has been implemented in earlier studies (Cai and Lubitz 2007; Cai, Schenker and Lubitz 2006).

Second, the two programs differ in the choice of event history models to fit the panel data. Nearly all large panel surveys interview their subjects infrequently at intervals of one to two years, or even five years. This data collection schedule is likely to miss many events of short duration between scheduled follow-up interviews (Hardy and Gill 2004). It thus seems desirable to apply the MSLT model to shorter transition periods (e.g., monthly, quarterly, etc.) that are “embedded” within the longer observation interval. Laditka and Wolf (1998) first applied the embedded Markov chain (eMC) approach to MSLT estimations. Their research inspired Lièvre et al. (2003) and was incorporated in the development of the IMaCh program. The SPACE program, on the other hand, uses the traditional event history approach that assumes one single spell between two successive interviews. This assumption is used in many health expectancy studies (e.g., Cai and Lubitz 2007; Crimmins et al. 1994, 1996; Hayward and Grady 1990; Land et al. 1994; Rogers, Rogers and Branch 1989; Rogers, Rogers and Belanger 1989).

To date there is little conclusive evidence favoring one approach over the other, despite more realistic assumption employed by the eMC method. If health expectancy is the main goal of study, research suggests that both approaches may yield comparable results (Laditka and Wolf 1998). Gill et al. (2005) also noted that health expectancy estimates are insensitive to the time between observations for up to two years. For more direct analysis of events, such as the incidence of or recovery from disability, both approaches may be equally biased, however (Wolf and Gill 2008).

Third, the two programs differ in the degree of modeling flexibility. The IMaCh program assumes that the underlying functional form for all events (i.e., the risk of disability onset, the risk of recovery, the risk of death from disability and from other states) in the state space follows a Gompertz failure time distribution. The Gompertz assumption may hold relatively well for changes in functional disability. But, if other processes were being modeled – e.g., retirement, marriage/divorce, etc., the Gompertz assumption need not hold and in fact could be inappropriate. In contrast, the SPACE program allows a more general set of failure time distributions, including the Gompertz function, so that it may be more applicable to a broader array of demographic analyses.

Finally, the programs differ in computation techniques. The IMaCh program uses a deterministic approach (i.e., the radix population) to estimate health expectancy, while the SPACE program uses stochastic approach (i.e., microsimulation). Microsimulation is a computation technique that has become popular in recent years (e.g., Cai and Lubitz 2007; Cai et al. 2006; Laditka and Wolf 1998; Lubitz et al. 2003; Wolf 1986). It uses MSLT transition probability estimates to produce person-level health trajectories; from this collection of individual data a broad range of population statistics can be computed directly. Many of these statistics may be difficult, or even impossible, to obtain otherwise (e.g., the probability of becoming disabled again at age 85 after two prior episodes, each lasting one year or longer).

Another advantage of microsimulation is the enhanced capability to analyze the variability of health expectancy about its expected value (Wolf and Laditka 1997). Traditionally, researchers focus on the average value of health expectancy due primarily to the limitation of computation techniques. By tabulating individual life histories as a reflection of the underlying stochastic process, microsimulation allows analysis of the full frequency distribution of statistics of interest, rather than just their average values. Using such technique, Wolf and Laditka (1997) found substantial variation in the distribution of health expectancy. By combining microsimulation with the bootstrap method in the SPACE program, researchers will therefore have a powerful tool to assess the sampling variability as well as the uncertainty associated with the distribution of health expectancy as well as a broad array of population health measures.

In the following sections we will describe the data set and the methods used in the SPACE program. We will then present two sets of results: one to compare the variance estimates from the SPACE and the IMaCh programs to assess the design effect (DEFF), and another to highlight the usefulness of microsimulation by presenting measures of the distribution of health expectancy at 65 years of age. These results may help readers better understand the differences between the SPACE and the IMaCh programs.

DATA

The study population is drawn from the 1998-2002 panels of the Medicare Current Beneficiary Survey (MCBS). The MCBS is a nationally representative, multistage, longitudinal panel survey of the Medicare population, sponsored by the Centers for Medicare and Medicaid Services, and conducted continuously since 1991. The survey gathers data on a wide range of topics such as health status, socio-demographic information, and use and costs of medical services. Survey records are linked to administrative data on use and expenditures of Medicare-covered services (hospital, physician, etc.) and on vital status. Interested readers can go to their website (http://www.cms.hhs.gov/mcbs) for more information.

The MCBS has all the elements of a complex survey. Strata are created based on the characteristics of PSUs, which are basically large geographical areas (i.e., a Metropolitan Statistical Area (MSA) or group of contiguous counties). The largest MSAs in the country are selected with probability one, each is essentially a "stratum." Within these certainty strata, the first stage of selection is zip clusters, which are paired up to form pseudo strata; the individual zip clusters are considered PSUs for variance estimation. For each of the noncertainty strata, two PSUs are selected. The analysis sample used in this study has a total of 112 strata and 1,168 PSUs. The individual Medicare beneficiaries are then selected in the third stage (i.e., within each zip cluster) stratified by seven age groups (under 45, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84 and 85 and over). The oldest (85 and over) and the disabled (64 and under) are oversampled to allow detailed analysis of their health status and health care needs.

The MCBS follows a rotating panel design with three in-person interviews per year. Each person is scheduled to receive 12 interviews over a four-year period. Health status information is gathered once each year in the Fall. The 1998-2002 panels contain 14,892 elderly beneficiaries. We excluded 1,017 persons of Hispanic origin or other racial/ethnic groups to focus on white and black non-Hispanics only, since the IMaCh program cannot accept more than two categories at a time for any covariate. The full analysis sample contains 50,830 person-year observations for 13,875 persons of age 65 and older. We use a dichotomous measure of health based only on the presence of limitations in activities of daily living (ADLs). A person is considered disabled if he or she either responds “yes” to having difficulty with one or more of the six ADLs – bathing, dressing, eating, transferring, walking, and using the toilet, or responds “does not do the activity because of a health or physical problem.” Otherwise, this person is considered non-disabled or active. Survey respondents potentially “move” between the disabled and non-disabled states over time, and death is the third and the absorbing state.

Table 1 shows several characteristics of sampled persons in the study population. The majority of sampled persons are women, white non-Hispanics, between ages of 65 and 74 and free of IADL and ADL limitations. The educational achievement of the incoming panels shows substantial improvement over the five-year period.

Between 1998 and 2002, the proportion with less than 11 years of education dropped from 30 percent to 24 percent, while the proportion of those with at least some college education increased from 34 percent to 43 percent. On the other hand, the prevalence of active health in the incoming panels dropped slightly from 62 percent in 1998 to 59 percent in 2002, while the prevalence of both IADL and ADL limitation increased slightly.

THE SPACE PROGRAM

The SPACE program has two components. The main component controls the generation of bootstrap samples, calls the statistical module to perform MSLT calculations and collect the bootstrap estimates, as well as the original point estimates from the full sample, into a small data set for further analysis of sampling variability. The statistical module estimates the observed prevalence of health states, performs the logistic regression estimation, and using these two pieces of input to estimate health expectancy or other statistics of interest. It is written in PC SAS 9.1 and requires SAS/BASE, SAS/STAT and SAS/IML (Interactive Matrix Language). The “users manual” in the zip file contains detailed instructions on how to use this program.

Methods

1. Model Estimation

The MSLT estimation is performed in the main statistical module of the SPACE program. It takes the traditional event history approach by assuming at most one event between two successive observations: if one has the same health status in both occasions, then it is assumed that no event has occurred between the two dates; if they are different, it is then assumed that only one event has occurred. Following this assumption, the SPACE program examines the occurrence of event (or lack of it) between pairs of successive interviews and fits a multinomial logistic regression of the following basic form:

$$\log(p_{ij} / p_{ii}) = \alpha_{ij} + \beta_{ij} age, \qquad (1)$$

where p_{ij} is the transition probability from the current state i to state j ($i \neq j$) over the observation interval.[2] Additional covariates (e.g., gender and race/ethnicity) can be added to help evaluate the differences in MSLT function among population subgroups. The expanded regression takes the following basic form:

$$\log(p_{ij} / p_{ii}) = \alpha_{ij} + \beta_{1ij} age + \beta_{2ij} gender + \beta_{3ij} race / ethnicity. \qquad (2)$$

Although these basic forms are identical to those in the IMaCh program, users will have substantial flexibility to modify their specifications to find the best model for their data. SAS offers powerful modeling capabilities in its procedures for categorical variable analysis. Researchers can relax the Gompertz assumption of the age function to test other forms of age dependence (e.g., logarithm or polynomial functions). They can use model selection procedures to quickly evaluate the main and interaction effects of covariates on either the full sample or its subsets. They can also evaluate different forms of the link function (e.g., cumulative, multinomial or complementary log-log) to reflect the characteristics of the underlying stochastic process. This flexibility of

[2] The SPACE program also provides the option to fit a set of constant hazard functions as in Hayward (1999) and Crimmins (1996), and to use the radix population method to calculate health expectancy.

modeling is unavailable in the IMaCh program.

It is worth noting that the traditional approach does not take into account the variation in actual interval between interviews. It simply arranges the data into pairs of observations and models the conditional event probabilities directly. In the case of MCBS, although it is *designed* with 12-month intervals, the actual gap ranges from 8 or 9 months to 16 months (Table 1). This variation of time interval is ignored by the traditional approach, but not by the eMC method. As a result, the SPACE coefficient estimates for the "annual" interval are slightly different from the IMaCh estimates with 12-month interval (results not shown here).

The SPACE program has a lower burden of computation than the IMaCh program. The logistic regression estimation is very quick in the SPACE program because it directly estimates the conditional probability between pairs of observations. Under the assumption of eMC, "the probability of making a transition from state i to state j over an interval of w months is the (i, j) entry in the matrix P^w" (Laditka and Wolf 1998:224). Given that only a fraction of the states a person occupies over the course of study is actually observed at scheduled follow-ups, the eMC approach used by the IMaCh program estimates the short-interval transition probabilities through an iterative algorithm that maximize the likelihood of "producing" these partially observed data. The convergence of iterative algorithm, if it exists, is typically time-consuming and depends on the starting values assumed for the parameters and characteristics of the data sample.

2. Calculation of the MSLT Functions

A distinct feature of the SPACE program is its use of microsimulation as the primary means of computation of MSLT functions. During the last two decades, microsimulation has emerged as a promising computational tool to analyze population health in demography. By tabulating the life history for each simulated individual, researchers can examine a broad array of summary statistics.

The simulation procedure is also performed in the main statistical module. Since it has been used in many previous studies (e.g., Cai and Lubitz 2007; Laditka and Wolf 1998; Wolf 1986), we will only provide a brief overview here. Suppose we want simulate the life histories of a 100,000-person cohort of 65-year olds. We first convert the logistic regression coefficient estimates into age-specific multi-state transition probabilities. Then for each member of the 100,000-person cohort, we evaluate his or her possible health changes by comparing a uniform random number with the transition probabilities, conditional on the current status. We perform this comparison one age at a time until his or her death. After we repeat this process for the entire cohort, we have a large collection of individual life histories from which various summary measures such as health expectancy can be derived.

There is a minor difference between the SPACE and the IMaCh program regarding the estimation of population-based health expectancy. Population-based health expectancy at age 65 is a weighted average of status-based expectancy. To make the estimates representative of the current population, researchers typically use as weight the observed prevalence of health status at 65. The SPACE program follows this tradition, although in this present study we used smoothed prevalence estimates since the number of 65-year olds in the MCBS sample is very small. The IMaCh program, on the other hand, uses the period prevalence (Lièvre, Brouard and Heathcote 2003). The period prevalence is an equilibrium prevalence achieved in a stable population, assuming all current conditions are to remain constant. During times of improving health, this period prevalence of disability is expected to be lower than the observed prevalence, as is shown in Table 2B. The effect of prevalence on population-based health expectancy is generally very small, relative to the regression coefficients.

3. The Bootstrap Procedure

Variance estimation for a complex survey needs to consider additional sources of variability that are not present in a SRS. A complex survey usually includes stratification and multi-stage clustering. Treating a stratified sample as a SRS usually overestimates the variance, while treating a clustered sample as SRS usually underestimates the variance. Although the net effect is often not straightforward, it is nonetheless clear that ignoring the complex sampling design can lead to incorrect statistical inference (Lohr 1999).

Given the sampling design of the MCBS, we apply a version of the rescaling bootstrap of Rao and Wu (1988), which has been described in detail in Lohr (1999:307), and has also been used in two recent studies (Cai and Lubitz 2007; Cai et al. 2006). We first sample $n_h - 1$ PSUs with replacement from each of the 112 strata, where n_h is the number of PSUs in stratum h. For each PSU$_i$ sampled from stratum h, we multiply the sample weight by $\frac{n_h}{n_h - 1} * m_i$, where m_i is the number of times the PSU$_i$ is selected. It is worth noting that this particular procedure has two potentially offsetting sources of bias.

First, this procedure resamples only at the PSU level and thus will underestimate the variance for a multistage survey. This source of bias is not likely to be significant, however, since the additional variability due to subsampling at later stages is usually negligible compared to variability at the PSU level (Lohr 1999). Second, this procedure draws the bootstrap samples with replacement, which may lead to overestimation of the variance for data sampled without replacement. This second source of bias may be negligible if the first-stage sampling fraction is small (Rao 1988). If not, then alternative procedures specifically developed for without-replacement samples (e.g., Bickel and Freedman 1984; Sitter 1992b) can be considered. But these procedures are more difficult to implement and require knowledge of the sampling fraction, which we often do not know and which is typically not available to researchers using the publicly released versions of the survey data.

To be sure, the bootstrap method is not the only approach to variance estimation in complex surveys. Other alternatives include the traditional linearization approach and two other data resampling methods – the balanced repeated replication (BRR) method (McCarthy 1969), and the Jackknife method (Quenouille 1949; Tukey 1958). The linearization approach has been used extensively in statistics. It can produce a variance estimate for any statistic if the variance formula is known and continuously differentiable. But not all statistics fall into that category (e.g., the median and other quantiles), and for those that do, derivation of the formula may sometimes be too complicated to be useful (Shao and Tu 1995). In addition, a separate variance formula must be derived for each statistic, which can be “difficult and tedious” (Shao and Tu

1995:4). The BRR method is often used for surveys with $n_h = 2$ primary sampling units (PSU) per stratum h, where BRR samples are created by sampling one PSU from each stratum. The BRR variance estimators are consistent if the statistic is a smooth (i.e., continuously differentiable) function of the weighted sample mean; for sample quantiles, the BRR estimators are still consistent under some weak conditions (Shao and Tu 1995). Although the BRR method can be generalized to handle $n_h \geq 2$ PSUs per stratum (Valliant 1987), the application "... is not easy when the n_h are not equal", and "may require separate software or involve nontrivial mathematical developments" (Shao and Tu 1995:244). It may even overestimate the variance if the number of PSUs in stratum h in the population is small (Lohr 1999). The Jackknife samples are formed by deleting one PSU from each stratum at a time. Like the BRR estimators, its variance estimator is also consistent if the statistic is a smooth function of the weighted sample mean, but the Jackknife method is not applicable to quantile estimators (Shao and Tu 1995). Its property in unequal-probability, without-replacement survey is generally unknown (Lohr 1999).

Compared with these alternatives, the bootstrap method usually requires more computation, and its theoretical properties in complex surveys are not as fully studied as the other methods (Lohr 1999). There is also evidence from simulation studies that the bootstrap results do not outperform the Jackknife and the BRR results in the case of stratified one-stage simple random sampling with replacement (Kovar, Rao and Wu 1988). But the bootstrap method also has a number of advantages. It can be used to estimate variance for a broader class of statistics, including sample quantiles or even the entire sampling distribution. It can also provide consistent variance estimators for surveys with imputed data, and has a “higher potential to be extended to other complex problems” than the BRR and jackknife approaches (Shao and Tu 1995:280). From a practitioner’s perspective, it therefore seems reasonable to conclude that the bootstrap method is a suitable all-purpose variance estimator for MSLT functions.

RESULTS

The main results of this study are presented in Tables 2-4. Tables 2A and 2B present the coefficient estimates of the logistic regressions as well as the prevalence of disability as input to the calculation of MSLT health expectancy estimates at age 65, which are presented in Tables 3 and 4. We fit the same logistic regressions of the form of eq. (2) for both IMaCh and SPACE programs to facilitate comparisons. In Table 2A, the IMaCh coefficients are estimated with one-month transition interval and the SPACE coefficients are estimated with annual interval. All of the coefficient estimates are statistically significant. The gender coefficients indicate that elderly women are more likely to become disabled, while less likely to recover and to die, than elderly men. The race coefficients indicate that elderly blacks are more likely to become disabled and die, while less likely to recover, than elderly whites. Table 2B shows the period prevalence of health states used in IMaCh health expectancy estimates as well as the predicted or smoothed prevalence for SPACE estimates. The SPACE prevalence estimates are similar to the observed prevalence, while the IMaCh prevalence estimates of disability are noticeably lower.

Table 3 presents the IMaCh estimates of population-based average life expectancy at age 65 by gender and race/ethnicity, and compares two sets of variance estimates – the bootstrap estimates that consider survey design and the IMaCh estimates that do not. The point estimates of health expectancy in this table are based on the IMaCh one-month coefficient estimates in Table 2A. The bootstrap variance estimates are obtained in the following manner. We first randomly select 250 bootstrap samples from the full analysis sample, then use them as input data set to the IMaCh program to derive 250 sets of IMaCh point estimates. The variances of these 250 estimates are used as the bootstrap variance estimates for the original IMaCh point estimates and are compared to the IMaCh variance estimates that do not reflect the complex sampling design of MCBS.

The design effect (DEFF) is measured by the ratio of the bootstrap variance estimates to the IMaCh estimates to evaluate the degree of bias in variance estimates by treating a complex survey as SRS. Since stratification and clustering have opposing effect on sampling variability, the value of the ratio may suggest the relative size of these design factors: if the ratio is greater than one then the clustering effect may be stronger; if the ratio is less than one then the stratification effect may be stronger. Table 3 shows that the bootstrap variance estimates are larger than the IMaCh estimates in all cases – an indication of the clustering effect being stronger than the stratification effect in MCBS. In some cases the bootstrap estimates are much larger. For example, the variance of ALE for all 65-year old is 67% larger and the variance of

DLE for white non-Hispanic female is 75% larger. The differences in variance estimates are not large enough to change the results of hypothesis tests, however.

Table 4 presents the SPACE estimates of the average, median and the 25th and 75th percentiles of total, active and disabled life expectancy at age 65, by gender and race/ethnicity. We also include the 2000 Vital Statistics as a comparison (Arias 2002). The SPACE estimates of health expectancy at age 65 are calculated using the regression coefficient estimates in Table 2A and the smoothed prevalence in Table 2B.

The point estimates of average TLE in Table 4 indicate some small differences with the 2000 Vital Statistics - the largest is 4.7 percent for white women. They are caused mostly by the lack of control for the variations of actual time interval between interviews in the SPACE program. We verified this source of difference by manually calculating the transition probabilities using the SPACE coefficient estimates in Table 2A and the IMaCh coefficient estimates with 12-month transition interval (not shown here). The IMaCh estimates of average TLE assuming a 12-month interval is closer to the 2000 Vital Statistics than the SPACE estimates.

The bootstrap variance estimates in Table 4 are also derived from 250 bootstrap samples, and are mostly smaller than the bootstrap variance estimates in Table 3. There may be an intuitive explanation for this: as one treats the partially-observed MCBS data as complete data and limits the number of events between successive observations, estimates of the average number and duration of health events become less variable. As a result, the sampling variability of health expectancy becomes smaller.

Estimates of median health expectancy at age 65 suggest that the distributions for TLE and ALE are generally symmetric, while the distributions of DLE have a longer tail on the right. Due to space constraints, we present the histograms of ALE and DLE at 65 for black men only in Figures 1 and 2. Figure 2 indicates the distribution of DLE is highly skewed to the left. About half of all black men are expected to have 2 or fewer disabled years; a insignificant percent (25%) will spend five or more years in disability stage after age 65.

CONCLUSION

This paper introduces the SPACE program, which provides demographers and social scientists with an alternate, and we believe, improved, means to model the dynamics of complex events and to draw statistical inferences. There are two major differences between the SPACE program and the other programs: the use of the bootstrap method to obtain consistent variance estimates from complex survey samples and the use of microsimulation to tabulate individual life histories. The bootstrap variance estimates are larger than the IMaCh estimates that do not consider sampling design in all the subgroups of the elderly population that we examine in this study, which suggests the importance of measuring sampling variability correctly in studies of MSLT functions. The use of microsimulation also provides a wide range of measures to characterize the dynamics of the aging process. In this study we examined the average values as well as the distributions of health expectancy to highlight the usefulness of microsimulation.

The high degree of asymmetry of the DLE distribution would have been hidden from analysts using deterministic approaches such as the radix population.

A practical consideration of the bootstrap procedure is the number of bootstrap samples to use. Unfortunately, there is no definitive answer to this question, given the many factors that may need to be considered. Our advice is to first select the best model from the full analysis sample, and then to run a large number of bootstrap samples to check if variance estimates display any noticeable pattern of variation for different sizes of the samples. If the variance estimates do not appear to stabilize , then more samples need to be drawn. In the present study, the variance estimates already appear stable at 250 samples; using a larger size such as 500 samples does not seem to affect these estimates at all.

Arguably the most controversial assumption used by the methods in the SPACE program is about the number of transitions between two successive observations, which is either zero or one. This assumption is conceptually weaker than that used in the eMC method. The differences between the IMaCh and SPACE estimates are small in our study, likely because of the frequent interviews in MCBS. As the length of interval between interviews increases, the MSLT parameter estimates in the SPACE program will likely become more biased. As a result, the simulated life histories are likely to be incorrect also.

From a practical perspective, however, there can be difficulties to implement the conceptually superior eMC method. A problem with the IMaCh program is the nonconvergence of the iterative likelihood maximization algorithm. While this is not necessarily an indication of any problem with the eMC method itself, it makes the use of eMC method more difficult. Craig and Sendi (2002) suggest using the deterministic expectation-maximization (EM) algorithm to avoid the convergence problem. The E-step calculates the probability of *every* possible path one can take between state i and j over the w months between two successive observations. This approach requires a large amount of computation that can quickly become unmanageable. For example, suppose there are five health states (excellent, very good, good, fair and poor), which can change once per month. If two interviews are made on the 65th and 66th birthdays, then the number of possible trajectories one can take during the 10 intervening months will be 5^{10}=9,765,625. If the interviews are made every 24 months, then the number of possible paths becomes $5^{22}=2.384\text{x}10^{15}$. Such a large number of possibilities may mean that there is no possibility to obtain exact solutions. Alternatively, one can use the regularization techniques (Charitos, de Waal and van der Gaag 2007). It replaces negative entries on each row of the short-interval matrix with zeros and adjusts the non-negative entries on the same row based on a distance criterion. Using simulated data, the authors show that their method outperforms the EM approach when the number of health states is small.

Both solutions mentioned above assume homogeneity within an annual age intervals (i.e., constant transition probabilities between two successive birthdays) as in the eMC method. This assumption is intended to simplify the computation rather than to be a realistic description of the underlying stochastic process. In our opinion, a more general solution should allow nonhomogeneous short-interval transition matrices that change over every short-interval. Since it is not clear how the two solutions mentioned above can be extended to this more general case, a new approach to estimation is therefore needed.

In future works, semi-Markov model of the form in Cai et al. (2006) should also be incorporated into the SPACE program. A semi-Markov model is likely to be more appropriate for analysis of disability transitions given the evidence of duration dependence (e.g., Crimmins and Saito 1993; Hardy et al. 2005; Hardy and Gill 2004). It will also help reduce the problem of unobserved heterogeneity in the current choice of event history models. The regression models in both the SPACE and the IMaCh programs treat the observations within a person as conditionally independent of each other, given the covariates in the regression. If the model is misspecified, this would overestimate the sampling variability and conservative results of hypothesis testing. By including duration dependence in the model, the bias in variance estimates may be reduces, although the issue of unobserved heterogeneity may still remain.

Another area that awaits further research is the way sample weights are used in MSLT calculations. The IMaCh program uses only a single weight across all monthly intervals within a person to estimate transition probabilities, effectively equalizing the sample representativeness of these “observations” at different time points. The SPACE program, on the other hand, uses multiple weights, one for each time interval. In the current study, these are the cross-sectional weights in MCBS that correspond to the year when current health status is observed. Although the MCBS provides longitudinal weights to analyze persons across waves of observations, they are designed only for survivors of each panel (Ferraro and Liu 2005), and not appropriate for analysis of the event of death. Based on our own calculation, health expectancy estimates do not appear to be affected by the approaches to sample weights. But it is still desirable to devote more research to this issue.

We hope the SPACE program will become a useful analytic tool for researchers. It can be obtained from ???, or from the corresponding author.

REFERENCES

Al Snih, S., K.J. Ottenbacher, K.S. Markides, Y.-F. Kuo, K. Eschbach, and J.S. Goodwin. 2007. "The Effect of Obesity on Disability vs Mortality in Older Americans." *Arch Intern Med* 167(8):774-780.

Arias, E. 2002. "*United States abridged life tables, 2000*." in *National Vital Statistics Reports*. Hyattsville: National Center for Health Statistics.

Bickel, P.J.and D.A. Freedman. 1984. "Asymptotic Normality and the Bootstrap in Stratified Sampling." *The Annals of Statistics* 12(2):470-482.

Bumpass, L.and H.H. Lu. 2000. "Trends in cohabitation and implications for children's family contexts in the United States." *Population Studies-a Journal of Demography* 54(1):29-41.

Cai, L.and J. Lubitz. 2007. "Was There Compression of Disability for Older Americans From 1992 to 2003?" *Demography* 44(3):479-495.

Cai, L., N. Schenker, and J. Lubitz. 2006. "Analysis of functional status transitions by using a semi-Markov process model in the presence of left-censored spells." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 55(4):477-491.

Charitos, T., P.R. de Waal, and L.C. van der Gaag. 2007. "Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data." *Statistics in Medicine* 9999(9999):n/a.

Craig, B.A.and P.P. Sendi. 2002. "Estimation of the transition matrix of a discrete-time Markov chain." *Health Economics* 11(1):33-42.

Crimmins, E.and Y. Saito. 1993. "Getting better and getting worse." *Journal of Aging and Health* 5(1):3-36.

Crimmins, E.M., M.D. Hayward, and Y. Saito. 1994. "Changing mortality and morbidity rates and the health status and life expectancy of the older population." *Demography* 31(1):159-175.

—. 1996. "Differentials in active life expectancy in the older population of the United States." *J Gerontol B Psychol Sci Soc Sci* 51(3):S111-120.

Duncan, G.J.and W.L. Rodgers. 1988. "Longitudinal Aspects of Childhood Poverty." *Journal of Marriage and the Family* 50(4):1007-1021.

Espenshade, T.J.and R.E. Braun. 1982. "Life Course Analysis and Multistate Demography - an Application to Marriage, Divorce, and Remarriage." *Journal of Marriage and the Family* 44(4):1025-1036.

Ferraro, D.and H. Liu. 2005. "Uses of the Medicare Current Beneficiary Survey for Analysis Across Time." in *Joint Statistical Meeting of the American Statistical Association*. Minneapolis, MN.

Gill, T.M., H. Allore, S.E. Hardy, T.R. Holford, and L. Han. 2005. "Estimates of Active and Disabled Life Expectancy Based on Different Assessment Intervals." *J Gerontol A Biol Sci Med Sci* 60(8):1013-1016.

Goldman, D.P., B. Shang, J. Bhattacharya, A.M. Garber, M. Hurd, G.F. Joyce, D.N. Lakdawalla, C. Panis, and P.G. Shekelle. 2005. "Consequences Of Health Trends And Medical Innovation For The Future Elderly." *Health Aff*:hlthaff.w5.r5.

Guralnik, J.M., K.C. Land, D. Blazer, G.G. Fillenbaum, and L.G. Branch. 1993. "Educational Status and Active Life Expectancy among Older Blacks and Whites." *N Engl J Med* 329(2):110-116.

Hardy, S.E., J.A. Dubin, T.R. Holford, and T.M. Gill. 2005. "Transitions between States of Disability and Independence among Older Persons." *Am. J. Epidemiol.* 161(6):575-584.

Hardy, S.E.and T.M. Gill. 2004. "Recovery from disability among community-dwelling older persons." *JAMA* 291(13):1596-1602.

Hayward, M.D.and W.R. Grady. 1990. "Work and Retirement among a Cohort of Older Men in the United-States, 1966-1983." *Demography* 27(3):337-356.

Hayward, M.D., W.R. Grady, and S.D. McLaughlin. 1988. "Recent Changes in Mortality and Labor-Force Behavior among Older Americans - Consequences for Nonworking Life Expectancy." *Journals of Gerontology* 43(6):S194-S199.

Hofferth, S.L. 1985. "UPDATING CHILDRENS LIFE COURSE." *Journal of Marriage and the Family* 47(1):93-115.

Jagger, C., R. Matthews, D. Melzer, F. Matthews, C. Brayne, and C. Mrc. 2007. "Educational differences in the dynamics of disability incidence, recovery and mortality: Findings from the MRC Cognitive Function and Ageing Study (MRC CFAS)." *Int. J. Epidemiol.* 36(2):358-365.

Kovar, J.G., J.N.K. Rao, and C.F.J. Wu. 1988. "BOOTSTRAP AND OTHER METHODS TO MEASURE ERRORS IN SURVEY ESTIMATES." *Canadian Journal of Statistics-Revue Canadienne De Statistique* 16:25-45.

Laditka, S.B.and D.A. Wolf. 1998. "New methods for analyzing active life expectancy." *Journal of Aging and Health* 10:214-241.

Land, K.C., J.M. Guralnik, and D.G. Blazer. 1994. "Estimating increment-decrement life tables with multiple covariates from panel data: The case of active life expectancy." *Demography* 31(2):297.

Lièvre, A., N. Brouard, and C. Heathcote. 2003. "THE ESTIMATION OF HEALTH EXPECTANCIES FROM CROSS-LONGITUDINAL SURVEYS." *Mathematical Population Studies* 10(4):211 - 248.

Lohr, S.L. 1999. *Sampling : design and analysis*. Pacific Grove, CA: Duxbury Press.

Lubitz, J., L. Cai, E. Kramarow, and H. Lentzner. 2003. "Health, Life Expectancy, and Health Care Spending among the Elderly." *N Engl J Med* 349(11):1048-1055.

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:S153-166.

McCarthy, P.J. 1969. "Pseudo-replication: Half samples." *Rev. Internat. Statist. Inst.* 37:239-264.

Palloni, A. 1996. "Demography of HIV/AIDS." *Population Index* 62(4):601-652.

Quenouille, M.H. 1949. "Approximate Tests of Correlation in Time-Series." *Journal of the Royal Statistical Society. Series B (Methodological)* 11(1):68-84.

Quillian, L. 2003. "How long are exposures to poor neighborhoods? The long-term dynamics of entry and exit from poor neighborhoods." *Population Research and Policy Review* 22(3):221-249.

Rao, J.N.K. 1988. "Variance Estimation in Sample Surveys." Pp. 427-447 in *Handbook of Statistics*, edited by P.R. Krishnaiah and C.R. Rao. Amsterdam: Elsevier Science Publishers.

Rao, J.N.K.and C.F.J. Wu. 1988. "Resampling Inference with Complex Survey Data." *Journal of the American Statistical Association* 83(401):231-241.

Reynolds, S.L., Y. Saito, and E.M. Crimmins. 2005. "The Impact of Obesity on Active Life Expectancy in Older American Men and Women." *Gerontologist* 45(4):438-444.

Rogers, A., R.G. Rogers, and L.G. Branch. 1989. "A multistate analysis of active life expectancy." *Public Health Rep* 104(3):222-226.

Rogers, R.G., A. Rogers, and A. Belanger. 1989. "Active Life among the Elderly in the United-States - Multistate Life-Table Estimates and Population Projections." *Milbank Quarterly* 67(3-4):370-411.

Schoen, R. 1988. *Modeling Multigroup Populations*. New York: Plenum Press.

Schoen, R.and K.C. Land. 1979. "GENERAL ALGORITHM FOR ESTIMATING A MARKOV-GENERATED INCREMENT-DECREMENT LIFE TABLE WITH APPLICATIONS TO MARITAL-STATUS PATTERNS." *Journal of the American Statistical Association* 74(368):761-776.

Shao, J.and D. Tu. 1995. *The jackknife and bootstrap*. New York, NY, USA: Springer Verlag.

Sitter, R.R. 1992a. "Comparing 3 Bootstrap Methods for Survey Data." *Canadian Journal of Statistics-Revue Canadienne De Statistique* 20(2):135-154.

—. 1992b. "A Resampling Procedure for Complex Survey Data." *Journal of the American Statistical Association* 87(419):755-765.

Tukey, J.W. 1958. "Bias and Confidence in not quite large samples." *The Annals of Mathematical Statistics* 29(2):614.

Valliant, R. 1987. "Some Prediction Properties of Balanced Half-Sample Variance Estimators in Single-Stage Sampling." *Journal of the Royal Statistical Society Series B-Methodological* 49(1):68-81.

Wolf, D.and T.M. Gill. 2008. "Modeling Transition Rates Using Panel Current-Status Data: How Serious is the Bias?"

Wolf, D.A. 1986. "Simulation methods for analyzing continuous-time event-history models." Pp. 283-308 in *Sociological Methodology*, edited by N.B. Tuma. New York: Hossey-Bass Publications.

Wolf, D.A.and S.B. Laditka. 1997. "Stochastic Modeling of Active Life and Its Expectancy." in *Papers in Microsimulation Series Paper*. Syracuse: Center for Policy Research.

Table 1. Characteristics of the Analysis Sample of 1998-2002 Panels in MCBS

	1998	1999	2000	2001	2002
Sample size (N)	2577	2724	2861	2882	2831
Time between interviews (in months)					
Min	9	9	9	8	8
Mean	12	12	12	12	12
Max	16	16	16	16	16
			(in weighted percents of sample size)		
Gender					
Male	40.3	40.4	42.3	41.4	41.8
Female	59.7	59.7	57.7	58.7	58.2
Race/Ethnicity					
White non-Hispanic	92.3	91.6	91.3	91.5	92.3
Black non-Hispanic	7.7	8.4	8.7	8.5	7.7
Age					
65-74	57.9	56.5	54.7	53.0	53.7
75-84	32.4	33.8	35.4	35.4	34.4
85+	9.7	9.7	9.9	11.7	11.9
Education					
0-11 yrs	30.3	29.0	27.7	26.9	23.6
High school graduates (12 yrs)	35.6	29.7	29.4	30.9	33.8
At least some colleges (13+ yrs)	34.2	41.3	42.9	42.2	42.7
Functional health status at first interview					
Active (no IADL/ADL limitations)	62.0	58.8	59.7	59.2	59.2
IADL limitations only	11.5	13.2	11.3	12.8	12.8
1+ ADL limitations	26.5	28.0	29.1	28.0	28.0

Table 2A. Logistic Regression Coeffcients as Input to Multi-State Life Table Calculation in the IMaCh and SPACE Program

		IMaCh - Monthly Interval				SPACE - Annual Interval			
Current State	Destination State	Intercept	Age	Gender	Race/Ethnicity	Intercept	Age	Gender	Race/Ethnicity
Active	Disabled	-9.1406	0.0619	0.1795	0.1893	-7.3928	0.0694	0.2512	0.2111
Active	Death	-12.8755	0.0856	-0.5982	0.2458	-11.2588	0.1057	-0.4657	0.3014
Disabled	Active	0.8304	-0.0550	-0.1152	-0.0885	4.3649	-0.0661	-0.1971	-0.1360
Disabled	Death	-9.3780	0.0677	-0.5077	0.0399	-6.7215	0.0702	-0.5842	0.0208

Source: The 1998-2002 Medicare Current Beneficiary Survey.

Note: Men and whites are the reference categories of gender and race, respectively.

Table 2B. Observed and Estimated Prevalence of Health Status at Age 65 as Imput to the IMaCh and SPACE Program, in percents

Gender	Race/Ethnicity	Current State	Observed Prevalence	IMaCh - Period Prevalence	SPACE - Predicted Prevalence
Male	White non-Hispanic	Active	85.84	92.48	89.16
		Disabled	14.16	7.52	10.84
Male	Black non-Hispanic	Active	87.96	90.13	86.27
		Disabled	12.04	9.87	13.73
Female	White non-Hispanic	Active	83.12	90.47	85.73
		Disabled	16.88	9.53	14.27
Female	Black non-Hispanic	Active	79.39	87.55	77.95
		Disabled	20.61	12.45	22.05

Source: The 1998-2002 Medicare Current Beneficiary Survey.
Note: The IMaCh period prevalence is calculated with one-month transition interval. Due to the unique sampling design of MCBS, the incoming cohort of 65-year olds are small. So the observed prevalence are the average of 65- and 66-year olds.

Table 3. IMaCh Estimates of Average Health Expectancy At Age 65 and Their Standard Errors

			Standard Errors				Standard Errors				Standard Errors		
Gender	Race/Ethnicity	Total LE	IMaCh	Bootstrap	$DEFF = \frac{\text{var}(Bootstrap)}{\text{var}(IMaCh)}$	Active LE	IMaCh	Bootstrap	$DEFF = \frac{\text{var}(Bootstrap)}{\text{var}(IMaCh)}$	Disabled LE	IMaCh	Bootstrap	$DEFF = \frac{\text{var}(Bootstrap)}{\text{var}(IMaCh)}$
Male	White non-Hispanics	16.52	0.248	0.285	1.32	13.35	0.213	0.244	1.31	3.17	0.100	0.122	1.49
	Black non-Hispanics	15.01	0.589	0.678	1.33	11.70	0.476	0.591	1.54	3.31	0.244	0.259	1.13
Female	White non-Hispanics	19.04	0.236	0.248	1.10	13.81	0.186	0.229	1.52	5.23	0.127	0.168	1.75
	Black non-Hispanics	17.58	0.546	0.641	1.38	12.11	0.434	0.550	1.61	5.47	0.324	0.327	1.02

Source: IMaCh estimates from the 1998-2002 Medicare Current Beneficiary Survey, assuming monthly transition interval. Bootstrap standard errors are the standard deviations of 250 bootstrap estimates.

Table 4. SPACE Estimates of Health Expectancy At Age 65 and 2000 Vital Statistics, by Gender and Race/Ethnicity

			Average			25th Percentile			Median			75th Percentile		
Gender	Race/Ethnicity	2000 VS	Total	Active	Disabled	Total	Active	Disabled	Total	Active	Disabled	Total	Active	Disabled
Male	White non-Hispanic	16.3	15.9	12.9	3.0	10.5	8.5	1.0	16.5	13.0	2.0	21.5	17.5	4.5
	Std. Err.	-	0.13	0.16	0.10	0.43	0.25	0.00	0.43	0.41	0.00	0.00	0.00	0.25
	Black non-Hispanic	14.5	14.5	11.3	3.2	8.5	6.5	1.0	14.5	11.5	2.5	19.5	15.5	5.0
	Std. Err.	-	0.63	0.66	0.26	0.50	0.44	0.00	0.50	0.71	0.25	0.50	0.74	0.43
Female	White non-Hispanic	19.2	18.3	13.4	5.0	12.5	8.5	2.0	18.5	13.5	4.0	24.5	17.5	7.0
	Std. Err.	-	0.22	0.16	0.18	0.43	0.22	0.00	0.43	0.00	0.41	0.43	0.00	0.41
	Black non-Hispanic	17.4	17.0	11.5	5.5	11.5	7.5	2.0	17.5	11.5	4.5	22.5	15.5	8.0
	Std. Err.	-	0.33	0.37	0.14	0.83	1.12	0.22	0.83	0.55	0.00	0.43	0.25	0.22

Source: SPACE estimates are derived from microsimulation using the 1998-2002 Medicare Current Beneficiary Survey. Bootstrap standard errors are the standard deviations of 250 bootstrap estimates.

Figure 1. Histogram of Active LE at age 65 for Black Men

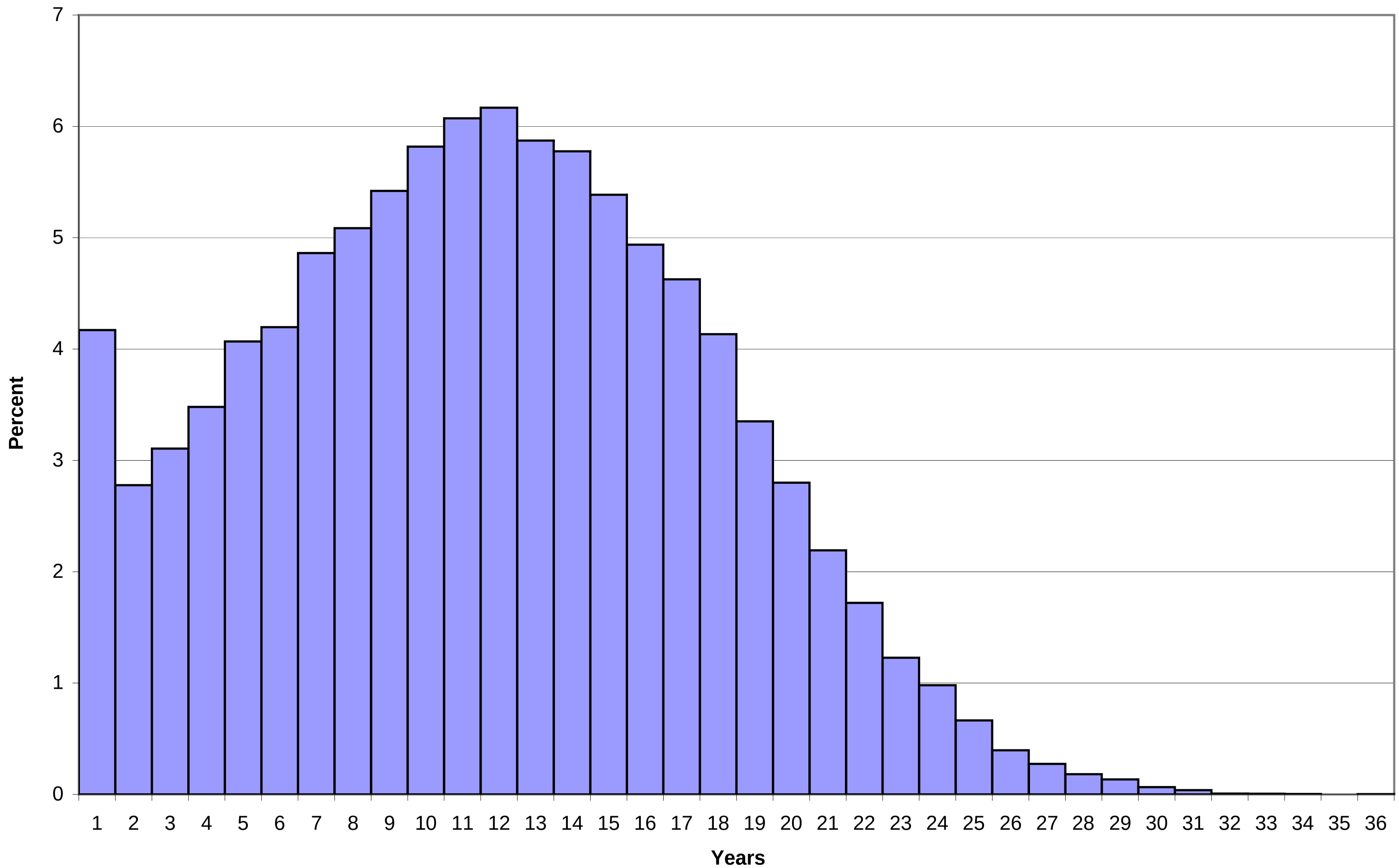

Figure 2. Histogram of Disabled LE at 65 for Black Men