4-27-2015

Variance Estimation for a Complex Life Table Quantity: Disease-free Life Expectancy

Ezra Fishman
University of Pennsylvania, fishmane@sas.upenn.edu

Variance Estimation for a Complex Life Table Quantity: Disease-free Life Expectancy

Abstract

Background: In the last decade, adult mortality in the United States has continued its long-run decline, while diabetes prevalence has increased. It is unknown whether the additional person-years lived in the adult population have mostly been spent in a diseased or a disease-free state. Furthermore, although illness and death are stochastic processes, little is known about the variance in diabetes-free life expectancy (DFLE) when compared across ages. More generally, methods of obtaining the variance of complex life table quantities are under-explored.

Objective: Estimate DFLE and its variance in the United States in 2000 and 2010.

Methods: Data on diabetes prevalence for ages 20+ come from the National Health and Nutrition Examination Surveys (NHANES), 1999-2000 (n=4,205) and 2009-2010 (n=5,752). Diabetes prevalence was defined as HbA1c at least 6.5% or taking diabetes medication. Deaths and population counts by age and sex come from the Human Mortality Database, covering the entire U.S. population. DFLE was estimated using Sullivan's method. Three methods of estimating variance in DFLE were explained and compared: the delta method, Monte Carlo simulation, and bootstrapping.

Results: Although life expectancy at age 20 rose by approximately 3 years for both males and females between 2000 and 2010, DFLE at age 20 did not change during this decade. At age 70, life expectancy rose by 2.5 years for males and 2.7 years for females, but DFLE rose only 0.7 years for males and 0.8 years for females. For all methods, both sexes and in both years, variance in DFLE was larger at younger ages (males, 2000, age 20, delta method: 0.020) than at older ages (males, 2000, age 70, delta method: 0.012). For any given age/sex/year, the delta method produced the smallest estimates of variance of DFLE, followed by Monte Carlo. Bootstrapping produced variance estimates that were by far the largest, often ten times larger than the Monte Carlo variances. Differences across methods in the variance in estimated diabetes prevalence accounted for most of the differences across methods in the variance of DFLE.

Conclusions: The vast majority of the person-years of life gained by the U.S. adult population between 2000 and 2010 were spent with diabetes. Variance in DFLE arises mostly from variance in estimated disease prevalence. The variance of life-table quantities can be obtained using multiple methods, and the appropriate method for a given research problem will vary.

Keywords

Diabetes, Life Expectancy, Sullivan Method, National Health and Nutrition Examination Survey (NHANES), Statistical Demography

Disciplines

Demography, Population, and Ecology | Social and Behavioral Sciences | Sociology

Comments

Fishman, Ezra. 2015. "Variance Estimation for a Complex Life Table Quantity: Disease-free Life Expectancy." PSC Working Paper Series, WPS 15-2, http://repository.upenn.edu/psc_working_papers/60.

Variance estimation for a complex life table quantity: Disease-free life expectancy

Ezra Fishman
May 2015

Contact information:
Ezra Fishman
Population Studies Center, University of Pennsylvania
3718 Locust Walk
239 McNeil Building
Philadelphia, PA 19104
fishmane@sas.upenn.edu
Tel: 215-898-6441

Abstract

Background: In the last decade, adult mortality in the United States has continued its long-run decline, while diabetes prevalence has increased. It is unknown whether the additional person-years lived in the adult population have mostly been spent in a diseased or a disease-free state. Furthermore, although illness and death are stochastic processes, little is known about the variance in diabetes-free life expectancy (DFLE) when compared across ages. More generally, methods of obtaining the variance of complex life table quantities are under-explored.

Objective: Estimate DFLE and its variance in the United States in 2000 and 2010.

Methods: Data on diabetes prevalence for ages 20+ come from the National Health and Nutrition Examination Surveys (NHANES), 1999-2000 (n=4,205) and 2009-2010 (n=5,752). Diabetes prevalence was defined as HbA1c at least 6.5% or taking diabetes medication. Deaths and population counts by age and sex come from the Human Mortality Database, covering the entire U.S. population. DFLE was estimated using Sullivan's method. Three methods of estimating variance in DFLE were explained and compared: the delta method, Monte Carlo simulation, and bootstrapping.

Results: Although life expectancy at age 20 rose by approximately 3 years for both males and females between 2000 and 2010, DFLE at age 20 did not change during this decade. At age 70, life expectancy rose by 2.5 years for males and 2.7 years for females, but DFLE rose only 0.7 years for males and 0.8 years for females. For all methods, both sexes and in both years, variance in DFLE was larger at younger ages (males, 2000, age 20, delta method: 0.020) than at older ages (males, 2000, age 70, delta method: 0.012). For any given age/sex/year, the delta method produced the smallest estimates of variance of DFLE, followed by Monte Carlo. Bootstrapping produced variance estimates that were by far the largest, often ten times larger than the Monte Carlo variances. Differences across methods in the variance in estimated diabetes prevalence accounted for most of the differences across methods in the variance of DFLE.

Conclusions: The vast majority of the person-years of life gained by the U.S. adult population between 2000 and 2010 were spent with diabetes. Variance in DFLE arises mostly from variance in estimated disease prevalence. The variance of life-table quantities can be obtained using multiple methods, and the appropriate method for a given research problem will vary.

Key words
Diabetes, Life Expectancy, Sullivan Method, National Health and Nutrition Examination Survey (NHANES), statistical demography

Introduction and Background

The most basic indicator of mortality in a population is life expectancy (LE), which is calculated from a set of age-specific mortality rates using a single-decrement life table [1]. When using current (cross-sectional) mortality rates, LE at birth measures the average length of time a newborn would live if he were subject to the current mortality risks for his entire life. For decades, population health researchers have been interested in methods to combine information on mortality with information on health into single indices to facilitate the comparison of health and mortality across populations and over time. Prominent among these indices is disease-free life expectancy (DFLE), also known as healthy life expectancy, which measures the average length of time a disease-free individual at a given age would live in a disease-free state, if she were subject to the current risks of mortality and disease for her remaining life. DFLE can refer to expected life without any disease or expected life without a particular disease, such as diabetes, the subject of the analysis in this paper. A fairly simple method of calculating DFLE was developed by Sullivan [2], in which the person-years lived at each age in the life table population are apportioned to diseased and disease-free states based on the age-specific prevalence of disease. Consistent with the basic life table, Sullivan's method uses a synthetic cohort approach, meaning it subjects the population at every age to current rates of disease and death, rather than observing the rates experienced by actual cohorts as they age. The synthetic cohort approach thus facilitates comparisons of the combined mortality-health states of a population at a point in time with the combined mortality-health states of the population at a different point in time.

Historically, demography has been more interested in measures of central tendency, such as means (of which LE and DFLE are examples), than in variances. Neither of the two major demography textbooks published in the last 15 years discusses the estimation of variances of life table parameters [1, 3]. And in fact, if one were to obtain information on disease presence from every member of the population, as well as complete records of mortality and age-specific population counts, one could calculate DFLE without consideration of sampling variation. However, in reality, disease prevalence in a population is almost always estimated from sample data in the form of queries from national health care databases with incomplete coverage [4] or from health surveys [5]. The variance of estimated DFLE, or some other measure of its uncertainty, such as a 95% confidence interval, is therefore of interest.

More fundamentally, even if the investigator had data on the entire population, one might consider the actual events that took place (deaths and disease cases) to be random draws from underlying, unobserved stochastic processes of mortality and disease. In this case, even quantities in the basic life table, such as LE, would have variances associated with them [6].

This paper will explore three methods of estimating variances of estimated DFLE. First, the delta method [7], so called because it involves approximating non-linear functions by taking partial derivatives and using Taylor Series linearization, will be employed as an analytic way to estimate the variance of non-linear functions of data. The delta method is widely used, but in many applications, the parameters estimated are functions too complex to obtain analytic confidence intervals; therefore, I will explore two simulation-based approaches as well. Monte Carlo simulation draws input parameters, such as regression coefficients, from an assumed distribution, and then calculates the output parameters of interest (such as DFLE) for each of *m* independent draws of inputs from this assumed distribution. The Monte Carlo variance is the variance of the output parameters from the *m* draws [8]. Finally, bootstrapping takes advantage of the property that a well-designed sample survey is representative of the population. By resampling observations from the sample subjects, bootstrapping recreates multiple simulated sample surveys and calculates DFLE for each. The bootstrap variance is again the variance of

the output parameters across all the resamples [9]. Both simulation-based approaches use the definition of the variance of an estimated parameter as the variation in the parameter over multiple independent, identically distributed samples.

Substantively, the paper compares diabetes-free life expectancy at various adult ages in the United States in 2000 versus 2010. As age-specific diabetes prevalence has risen [5] and adult mortality has declined [10], it is not obvious whether the additional years of life gained in the adult population over the last decade have mostly been spent in a diseased or a disease-free state. A recent paper on diabetes-free life expectancy in the United States for a somewhat earlier period did not report variances or confidence intervals [11].

Data

The basic data requirements for DFLE are a set of age- and sex-specific mortality rates and a set of age- and sex-specific disease prevalence rates. For mortality, I use sex-specific deaths and population counts for ages 20 to 85+, for years 1999, 2000, 2009, and 2010, from the Human Mortality Database [12], which organizes death data from the National Center for Health Statistics [13–16] and population counts from the U.S. Census. Death data are based primarily on death certificates and cover nearly 100% of deaths in the United States. Since each year's data are based on the mid-year population, averaging adjacent years' life tables centers the mortality schedules on January 1, 2000 and January 1, 2010.

Data on age-specific diabetes prevalence for ages 20 to 85+ come from the National Health and Nutrition Examination Surveys (NHANES), 1999 to 2000 and 2009 to 2010. Continuous NHANES began in 1999, for which data are released in two-year cycles, centered on January 1 of the second year of the release. NHANES is a complex, multi-stage probability sample of the U.S. civilian non-institutionalized population. Participants complete a home interview and are then examined in a mobile examination center, which includes sampling participants' blood for laboratory tests. Whenever possible, NHANES uses consistent laboratory procedures over time to facilitate analysis of trends in population health. Extensive documentation of NHANES survey, examination, and laboratory procedures and characteristics of the NHANES study sample are reported elsewhere [17–19].

I rely on laboratory results, rather than self-reported diagnoses, because the latter fails to capture the considerable number of individuals in the U.S. population with undiagnosed diabetes. A 2010 study estimated that 3.9 million individuals above age 20 had undiagnosed diabetes, representing 19% of the diabetic population [20]. Furthermore, intertemporal comparisons based on self-reported diagnosis are complicated by the fact that criteria for diagnosing diabetes in the clinical setting have changed over time [21]. Laboratory results are not available for children under age 12, and past studies have focused only on the population age 18 and above or 20 and above [11, 20], so I will also exclude children and begin with age 20. NHANES top-codes individuals aged 85 and above, and at those ages, prevalence barely rises at all with age [22], so I will use 85+ as the terminal age.

My definition of diabetes is based on HbA1C (glycated hemoglobin). This measure reflects average glycemia over a prolonged period and thus has more intra-subject stability than the leading alternative, a measure of fasting plasma glucose (FPG) [23]. Furthermore, HbA1c-based measures of diabetes are more strongly associated with cardiovascular disease and death than are FPG-based measures [24].

Several changes in laboratory measurement of HbA1C occurred over the course of Continuous NHANES (detailed elsewhere [18]), but I follow the NHANES recommendation and the methods of recent studies

and used HbA1C data without any corrections or adjustments [5, 18]. Individuals are considered diabetic if they had HbA1c ≥ 6.5% (48 mmol/mol) [25]. Because diabetes medication is expected to reduce glycemia, the HbA1c values of medicated persons might not capture their diabetes status correctly; therefore, all individuals who reported taking diabetes medication are considered diabetic.

Methods

Deaths and population counts from adjacent years (1999-2000 and 2009-2010) are added together to center the life tables on January 1 of the second year of each pair of years. Then, for each sex, standard life table columns – death rates ($m(x)$), death probabilities (q_x), survival probabilities ($p_x = 1 - q_x$, note lower-case *p*), cumulative survivors (l_x), person-years lived (L_x), and life expectancies – are calculated using standard methods [1]. One-year age intervals are used throughout, and individuals who die in a given interval are assumed to die, on average, halfway through that interval.

Diabetes prevalence among adults increases monotonically with age, but the increase is slow at the oldest ages [22]. There are not enough observations in NHANES to generate a smooth, monotonically increasing series of single-year age- and sex-specific prevalence estimates; this is a feature of almost any health survey, not a limitation specific to NHANES. Therefore, I model age- and sex-specific prevalence in a given year (2000 or 2010) using a logistic regression:

$$\Phi = log\left(\frac{P_x}{1 - P_x}\right) = \beta_0 + \beta_1 age + \beta_2 female. \quad \text{(Eq. 1)}$$

In **Equation 1**, P_x (note upper-case *P*) is the probability of having diabetes at age x; using the model it can be understood as the age-specific prevalence of diabetes. Φ is the logit of P_x, the linear combination of the predictors.
Prevalence is then estimated as

$$\hat{P}_x = 1/(1 + \exp(-(\widehat{\beta_0} + \widehat{\beta_1} age + \widehat{\beta_2} female))). \quad \text{(Eq. 2)}$$

The *logit* of prevalence is modeled as a linear combination of an intercept, age, and sex, but the prevalence is a non-linear combination of these.

Sullivan's DFLE at age x is

$$DFLE_x = \frac{\sum_{i=x}^{w} L_i\,(1 - P_i)}{l_x}, \quad \text{(Eq. 3)}$$

where *w* is the oldest age interval, L_i is person-years lived in the *i*th age interval, and l_x is the size of the life-table population at exact age x.

Delta Method

The delta method can be used to estimate variances and confidence intervals around life table-based estimated parameters [26]. Suppose l_0 is fixed (it is an arbitrary starting value fixed by the investigator), but L_x and P_x are random variables. L_x is a random variable if we consider mortality to be a stochastic process [6], and/or if mortality is estimated from sample data [27]. P_x is estimated from sample data (NHANES prevalence) and so is a random variable. Thus, DFLE is a non-linear function of random variables.

When using a period life table, as in the current study, deaths and person-years lived at a given age are uncorrelated with deaths and person-years at other ages [6, 28]. Furthermore, when age-specific prevalence is obtained from a separate cross-sectional survey, it can be considered independent of age-specific mortality [29]. Therefore, covariances can be ignored and the delta method yields

$$Var(DFLE_x) = \sum_{i=x}^{w} \left(\frac{\partial DFLE}{\partial L_i}\right)^2 Var(L_i) + \left(\frac{\partial DFLE}{\partial P_i}\right)^2 Var(P_i). \quad \text{(Eq. 4)}$$

The first term in **Equation 4** may be thought of as the contribution of the variance in mortality to the variance in DFLE, while the second term may be thought of as the contribution of the variance in disease prevalence to the variance in DFLE.

We have

$$\frac{\partial DFLE}{\partial P_x} = \frac{\sum_{i=x}^{w} L_i}{l_x}. \quad \text{(Eq. 5)}$$

Chiang [30] showed that L_x is a linear function of the cumulative probability of surviving to age *x*, which is itself a product of the probabilities p_i (note lower-case *p* here) of surviving each age group from 0 to *x*:

$$L_x = n_x l_{x+1} + n_x a_x (l_x - l_{x+1}), \quad \text{(Eq. 6a)}$$

$$l_x = l_0 \prod_{i=0}^{x-1} p_i. \quad \text{(Eq. 6b)}$$

In **Equation 6a**, n_x is the length of the age interval, and a_x is the average proportion of the interval lived by people who die in the interval. These are fixed quantities chosen by the investigator, not random variables [1, 30]. In the present analysis, n_x = 1 and a_x = 0.5 for every age, corresponding to the use of one-year age intervals and the assumption that individuals die, on average, halfway through each single-year age interval. Then, one can rewrite **Equation 3** using survival probabilities,

$$DFLE_x = \frac{\sum_{i=x}^{w} \left[n_i l_0 \prod_{j=0}^{x} p_j + n_i a_i (l_0 \prod_{j=0}^{x-1} p_j - l_0 \prod_{j=0}^{x} p_j)\right] (1 - P_i)}{l_0 \prod_{j=0}^{x-1} p_j}. \quad \text{(Eq. 7)}$$

Recalling that l_0, n_x and a_x are constants, **Equation 4** can thus be rewritten as

$$Var(DFLE_x) = \sum_{i=x}^{w} \left(\frac{\partial DFLE}{\partial p_i}\right)^2 Var(p_i) + \left(\frac{\partial DFLE}{\partial P_i}\right)^2 Var(P_i). \quad \text{(Eq. 8)}$$

The result is the following formula for the delta-method variance of disease-free life expectancy (see [29] and [30] for the complete derivation):

$$Var(DFLE_x) = \sum_{i=x}^{w-1} \left[\left(\frac{l_i}{l_x}\right)^2 ((1 - a_i n_i)(1 - P_i) + DFLE_{i+1})^2 Var(p_i)\right] + \sum_{i=x}^{w} \left[\left(\frac{L_i}{l_x}\right)^2 Var(P_i)\right]. \quad \text{(Eq. 9)}$$

The variance of the probability of survival within each interval, $Var(p_i)$, equals the variance of the probability of death within each interval,

$$Var(p_i) = Var(1 - q_i) = Var(q_i) = \frac{q_i^2 (1 - q_i)}{D_i}, \quad \text{(Eq. 10)}$$

with D_i the observed number of deaths in the interval, based on a manipulation of the formula for the variance of an estimated binomial probability [30].

The variance of the estimated age-specific prevalence, $Var(P_x)$, is also estimated using the delta method, since P_x is based on fitted values calculated from a non-linear function of the logistic regression parameter estimates (using **Equation 2**). Fitting the model in **Equation 1** to the data, I obtain a 3 x 1 vector of estimated coefficients,

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \widehat{\beta_2} \end{bmatrix}. \quad \text{(Eq. 11)}$$

The variance of the age-specific prevalence will be

$$\widehat{Var}(P_x) = \left[\frac{\partial \boldsymbol{P}}{\partial \widehat{\boldsymbol{\beta}}}\right]^{\boldsymbol{T}} \boldsymbol{VCOV}\left(\widehat{\boldsymbol{\beta}}\right) \left[\frac{\partial \boldsymbol{P}}{\partial \widehat{\boldsymbol{\beta}}}\right], \quad \text{(Eq. 12)}$$

where $\left[\frac{\partial \boldsymbol{P}}{\partial \widehat{\boldsymbol{\beta}}}\right]$ is the gradient vector of P_x, that is, a 3 x 1 vector of the partial derivatives of P_x with respect to each of the three estimated coefficients:

$$\left[\frac{\partial \boldsymbol{P}}{\partial \widehat{\boldsymbol{\beta}}}\right] = \begin{bmatrix} \dfrac{exp(\hat{\Phi})}{\left(1+\exp\left(\hat{\Phi}\right)\right)^2} \\ \dfrac{age * exp(\hat{\Phi})}{\left(1+\exp\left(\hat{\Phi}\right)\right)^2} \\ \dfrac{female * exp(\hat{\Phi})}{\left(1+\exp\left(\hat{\Phi}\right)\right)^2} \end{bmatrix}. \quad \text{(Eq. 13)}$$

The gradient vector is evaluated at each age and sex, with

$$\hat{\Phi} = \widehat{\beta_0} + \widehat{\beta_1} age + \widehat{\beta_2} female. \quad \text{(Eq. 14)}$$

$\boldsymbol{VCOV}\left(\widehat{\boldsymbol{\beta}}\right)$ is the 3 X 3 variance-covariance matrix of the estimated coefficients from the information matrix of the logistic model fitted to the data.

Despite Chiang's conceptualization of every life table as a realization of an underlying, unobserved stochastic process [6, 31, 32], it is common in studies of DFLE to consider basic life table quantities to have zero variance if they are calculated from deaths and population counts from the entire population [33, 34], as is the case with the data in this paper. Nevertheless, I will use **Equation 9** with a stochastic life table – that is, treating quantities associated with mortality, such as L_x, as random variables, following Chiang – for the purpose of elucidating the method.

Monte Carlo Simulation

The variance of an estimator is, by definition, the average of squared deviations from the mean of the estimator over many independent samples. Monte Carlo simulation (MC), commonly used in studies of DFLE [8, 35], uses this definition to estimate the variance of DFLE by simulating many DFLE calculations, each obtained from a simulated age schedule of mortality and an independent draw of the estimated coefficients (**Equation 11**), and then calculating the variance of the simulated DFLE calculations. The motivation for using MC is that in many applications, obtaining closed-form equation for the variance of

an estimator of interest, as was done in **Equation 9** above, is impractical or impossible, because the output parameter of interest is a non-differentiable function of the data. This is clearly not the case with Sullivan's DFLE, making DFLE a good opportunity to compare the variance estimated by the closed-form equation to that obtained via MC.

To generate simulated age schedules of mortality for each sex and year, I fit a Gompertz function [1, 36] to the observed age-specific mortality rates:

$$m(x) = \alpha * \exp(\beta x). \quad \text{(Eq. 15)}$$

Each observed mortality rate contributes D_x observations to the regression that fits the Gompertz function, where D_x is the number of deaths observed in the population at age *x*. This repeating of observations is sometimes called frequency weighting and reflects the fact that each mortality rate arises from a large number of observations in the population. The resulting vector of estimated parameters, $[\hat{\alpha}\ \hat{\beta}]$, and its associated variance-covariance matrix are then used as the mean and covariance matrix of a bivariate Normal distribution from which the elements of a new $[\hat{\alpha}\ \hat{\beta}]$ are drawn. Using the newly drawn $[\hat{\alpha}\ \hat{\beta}]$, age-specific mortality rates are calculated using **Equation 15**, and l_x and L_x are calculated using the standard procedures [1]. The process is repeated *m* independent times, resulting in *m* vectors of $[\hat{\alpha}\ \hat{\beta}]$ and thus *m* schedules of l_x and L_x.

Separately, the disease prevalence model in **Equation 1** is fitted to the data once. The resulting $\widehat{\boldsymbol{\beta}}$ vector and $\boldsymbol{VCOV}(\widehat{\boldsymbol{\beta}})$ matrix are then used as the mean and covariance matrix of a trivariate Normal distribution from which the elements of a new $\widehat{\boldsymbol{\beta}}$ vector are drawn. Using the newly drawn $\widehat{\boldsymbol{\beta}}$, age- and sex-specific fitted prevalence values are calculated using **Equation 2.** Then the elements of a new $\widehat{\boldsymbol{\beta}}$ vector are drawn and the process is repeated *m* independent times, resulting in *m* vectors of $\widehat{\boldsymbol{\beta}}$ and thus *m* age-specific prevalence schedules.

This simulation process mimics the observational data, whereby disease prevalence data is obtained separately from mortality data, and only then are the two inputs combined. Each of the *m* prevalence schedules is paired with one of the *m* L_x columns, and DFLE is calculated using Sullivan's formula for the *m* pairs,

$$\widehat{DFLE}_x^j = \frac{\sum_{i=x}^{w} L_i^j (1 - P_i^j)}{l_x^j};\ j = 1,2,\dots,m. \quad \text{(Eq. 16)}$$

The estimated DFLE in the population is the average DFLE from these *m* simulations, $\widehat{\overline{DFLE_x}}$. The Monte Carlo-based variance estimate is

$$\widehat{Var}(DFLE_x) = \frac{1}{m}\sum_{j=1}^{m}(\widehat{DFLE}_x^j - \widehat{\overline{DFLE_x}})^2, \quad \text{(Eq. 17)}$$

with j indexing each independent draw. I will use *m*=1,000, a common value in the literature [8].

Bootstrapping

Like MC, bootstrapping involves simulating many estimated values of DFLE and considering the variance of DFLE to be the average of squared deviations from the average simulated DFLE (**Equation 17**). However, the simulated DFLEs do not arise from independent draws of parameters that define regular

age patterns of mortality and disease, as in MC. Rather, in bootstrapping, the simulated DFLEs arise from independently drawn samples of data [9, 37].

Bootstrapping relies on sampling theory, whereby the subjects in a sample survey, when properly weighted, represent the population to which the parameters pertain. For disease prevalence, the *N* subjects that were actually sampled, and their characteristics – that is, the actual NHANES data set – are referred to as the *empirical sample*. Since the empirical sample represents the population, one can draw a large number of new samples from the empirical sample that will on average, based on the law of large numbers, also represent the population. One draws *N* subjects, with replacement[1], from the empirical sample, generating a *simulated sample*. Then one fits the model in **Equation 1** to the simulated sample, and the fitted prevalence values are calculated using **Equation 2**. The process is repeated *m* times, starting by drawing *N* subjects afresh, with replacement, from the empirical sample. The result is *m* independent estimates of prevalence for each age.

To generate simulated age schedules of mortality for each sex and year, I consider the empirical life table quantities for each age *x*: N_x, the size of the population age *x*, and q_x, the probability of death at age *x*. For each age, I draw deaths from a binomial distribution [6] with parameters $n = N_X$ and $p = q_X$. From the distribution of sampled deaths over all ages, I calculate the life table l_x and L_x in the usual way [1]. The process is repeated *m* independent times, resulting in *m* schedules of l_x and L_x.

Each of the *m* prevalence schedules is paired with one of the *m* life tables, and, using **Equation 16**, *m* values of DFLE are generated. The estimated DFLE in the population is the average DFLE from these *m* simulated samples, $\widehat{\overline{DFLE_x}}$. The bootstrapped variance of DFLE is that given in **Equation 17**. I will again use *m*=1,000 [9].

Since bootstrapping is based on the idea that the empirical sample represents the population, it is important that the procedure by which we generate the simulated samples mimic the sampling procedure in the original sample survey [38]. Due to concern for the confidentiality of subjects, NHANES does not publish sufficient information to reconstruct its sampling procedure fully. Instead, for each subject, it provides a sampling weight along with a "pseudo-stratum" and "pseudo-PSU (primary sampling unit)", or cluster, which, when used together, generate variance estimates that are approximately equal to those obtained when internal NHANES researchers use all sampling information [19, 39]. Since I am using public NHANES data, I will conduct the resampling as follows [40]. First, I divide the empirical sample into strata using the "pseudo-stratum" indicator for each observation. Then within each stratum, I sample *J* clusters with replacement, with *J* equal to the number of clusters ("pseudo-PSUs") in each stratum in the empirical sample. (In the NHANES public data, *J* is almost always 2 or 3.) Finally, I sample *n* observations (persons) with replacement from within each of the *J* chosen clusters,

[1] The justification for sampling with replacement is as follows. Suppose I take a simple random sample of 100 people from the population of the United States. The population is so big, and the sample so small, that after I have sampled the first of my 100 people, the probability of being chosen for the remaining individuals in the sample frame remains basically unchanged: it was 100/(population of the United States) and is now 99/(population of the United States minus 1). Now suppose I am trying to draw a subsample of 25 people from my original sample of 100 people. Initially, each sampled person's probability of being in my subsample is 25/100. After I have drawn one person, the probability of being in my subsample for everyone else is now 24/99, which is not the same as 25/100. Thus, in the subsampling case, a unit's probability of being chosen depends on whether another unit was chosen, and the observations are not independent. To avoid this problem, we sample with replacement, so that after the first person is chosen, the probability of being chosen in the next draw is still 25/100 for everyone.

with *n* equal to the number of observations in the given cluster. Thus, any given bootstrapped sample will likely have repeat clusters and repeat observations, just as a bootstrapped sample based on an empirical simple random sample will have some repeat observations. The number of (non-unique) observations in a given bootstrapped sample will be

$$N_b = \sum_{s=1}^{S} \sum_{j=1}^{J} n_{js}, \qquad \text{(Eq. 18)}$$

with *j* indexing clusters, *s* indexing strata, and *b* indexing bootstrap simulations. As with analytic calculation of variances, the clustering of observations in PSUs will increase the bootstrapped variance relative to resampling every individual observation independently.

A major advantage of bootstrapped variances is that they are nonparametric, meaning they do not rely on any assumption about the joint distribution of the elements of $\widehat{\boldsymbol{\beta}}$ and do not impose a functional form on the age schedule of mortality rates, as MC does. The advantage of MC is that if the assumed distributions of $\widehat{\boldsymbol{\beta}}$ and of the mortality rates is approximately correct, resulting variance estimates will be smaller than when using bootstrapping (i.e. an efficiency gain).

The methods described above are conducted separately for the 2000 and 2010 data. I used R (R core team, Vienna, Austria) for all computations. Computer code for all analysis is available upon request.

Results

Results based on the delta method are shown in **Table 1**. Each panel shows the following quantities for the given sex and year at selected ages: life expectancy (column labeled "LE"), diabetes-free life expectancy (column labeled "DFLE"), diabetes prevalence ("Prevalence"), estimated variance of DFLE ("Var.DFLE"), the value of the first summation term in **Equation 9** ("1st term"), estimated variance in diabetes prevalence ("Var.Prev"), and the estimated standard error of estimated DFLE ("SE.DFLE"). A comparison of panels A and B reveals important changes that occurred to U.S. males during the 2000s. First, life expectancy at adult ages rose by approximately 2.5 to 3 years (depending on the age) over the course of the decade. However, diabetes-free life expectancy barely rose at all. DFLE at age 20 rose by about 0.02 of a year, and DFLE at age 50 rose by about 0.33 of a year. The small gain in DFLE relative to LE indicates that, on average, the years of life gained over the decade were mostly spent with diabetes. This fact is reflected in the diabetes prevalence estimates, which are higher at every age in 2010 than in 2000. Finally, the small DFLE gains were smallest at young ages and largest at old ages, indicating that the increase in diabetes prevalence at the youngest ages played an especially important role in contributing to the smallness of the DFLE gain.

Turning to the columns showing the variance of DFLE and its components, one notes that the value of the first term in **Equation 9** is always extremely small. Because the life tables used the entire U.S. population, their effective "sample size" was extremely large and variances of life-table quantities were tiny. At every age, the vast majority of the variance in DFLE arose from variance in the estimated prevalence, represented by the second term in **Equation 9**. We also see that the variance of DFLE was higher in 2010 than in 2000. Variance was higher in 2010 because longer survivorship at every age in 2010 compared to 2000 raises the value of the $\left(\frac{L_i}{l_x}\right)^2$ factor in **Equation 9**, raising the value of the second term in the variance equation for 2010 relative to 2000. The $\left(\frac{l_i}{l_x}\right)^2$ part of the first term is also higher in

2010 due to better survivorship, but the variance in survivorship probability $Var(p_i)$ is so miniscule that it renders the changes in $\left(\frac{l_i}{l_x}\right)^2$ basically meaningless.

One finds similar results when comparing U.S. females in 2000 versus 2010. Although they enjoyed gains in life expectancy of 2 to 3.5 years over the course of the 2000s, the gains in DFLE were very small. Estimated DFLE at age 20 was actually lower in 2010 than 2000, though each year's estimate is within one standard error of the other year's estimate. On average, the additional years of life gained during the 2000s were spent almost entirely with diabetes, due to increased prevalence at all ages. As with men, the small increase in estimated DFLE was smallest at young ages (negative at age 20) and largest at old ages.

Patterns of variance are also similar for females as for males. Almost all the variance in estimated DFLE comes from variance in estimated prevalence, since the variances associated with mortality were so small. Variance of DFLE was larger in 2010 than in 2000.

Table 2 shows results when using Monte Carlo simulation. Differences in LE, DFLE, and estimated diabetes prevalence between 2000 and 2010 for men and women follow the same patterns here as when using the delta method. Specifically, large increases in LE occurred over the decade for both sexes, but DFLE barely changed, especially at younger ages, due to higher diabetes prevalence at all ages. Since prevalence is modeled the same way (**Equation 1**) in both tables, this similarity in results is also to be expected; however, because the estimated prevalence in **Table 2** is calculated as the mean of the simulations, some variation could arise due to differences in sampling error between the two methods. And in fact we see that for a given sex/year, DFLE at the oldest ages is smaller when using the delta method (**Table 1**) than when using MC (**Table 2**).

When one compares the variances shown in **Table 2** to those in **Table 1**, one finds that the variance of DFLE tends to be higher for a given sex/year/age when using MC than when using the delta method. The higher variances of DFLE appear to arise primarily because the variance in life expectancy is higher when using MC than when using the delta method. For almost all year/sex/age combinations, the delta-method variance from the life table ("1st term") is smaller than the MC-based variance in LE, but the delta method variance in prevalence is larger than the MC-based variance in prevalence. Also, whereas the delta method generated variances that were larger in 2010 than in 2000 for a given age/sex, MC generated variances that were larger in 2000 than in 2010 for a given age/sex.

Table 3 shows results when using bootstrapping. Bootstrapped estimates of LE, DFLE, and diabetes prevalence are very similar to MC-based estimates, leading to a now-familiar result: Between 2000 and 2010, LE increased by almost three years for both men and women, but DFLE barely increased at all, especially at the youngest ages. However, for any given year/sex/age combination, the bootstrapped variances of DFLE and of diabetes prevalence are about ten times the corresponding variances from MC. For example, the estimated variance in DFLE at age 20 for males in 2000 was 0.392 when using MC and 4.677 when using bootstrapping. This large increase in the variance of DFLE when moving from MC to the bootstrap is reflected in the variance in prevalence, which is about ten times larger for a given age/sex/year in **Table 3** than in **Table 2**. This result illustrates the efficiency disadvantage associated with the bootstrap.

Discussion

This paper explored three methods for estimation of variance of diabetes-free life expectancy (DFLE): the delta method, Monte Carlo simulation (MC), and bootstrapping. While all three methods produced similar point estimates of DFLE, the patterns of estimated variance differed somewhat. The delta method produced the lowest variances for any given year/sex/age combination, while bootstrapping produced the highest variances. A comparison of MC to bootstrapping also illustrated the loss of efficiency associated with using a nonparametric method, though the magnitude of the efficiency loss is still strikingly large. The variance of estimated age-specific disease prevalence was also surprisingly small when using the delta method and MC, given the relatively small sample size of NHANES.

A recent paper used the National Health Interview Survey (NHIS) to compare DFLE in 1980-89 to that in 2000-2004 [11]. The authors found that over the period in question, DFLE at age 18 declined for both men and women, and DFLE at age 60 remained essentially unchanged. My results paint a similarly grim picture, with nearly all mortality improvements balanced by increases in diabetes prevalence. Substantively, what this paper adds is an update based on more recent trends, the use of measured data to avoid the problem of changes in diagnosis standards, and estimates of variance and standard errors associated with DFLE. Methodologically, the contribution of the paper is to illustrate how three methods of variance estimation work and how results using each method compare to each other, using a simple but commonly used application in population health research. There have been recent advances in the study of stochastic life table quantities; most build on one or more of the methods described here [41–45].

The substantive results should be interpreted cautiously. Specifically, Sullivan's DFLE is a cross-sectional quantity: it illustrates the mortality and health status of the U.S. adult population at two points in time and quantifies the extent to which growth in diabetes-prevalence has, in a sense, "cancelled out" declines in mortality over the last decade. It thus relates to the major question in demography and epidemiology about whether population-level gains in survivorship are being enjoyed in a healthy state or an unhealthy state [46, 47]. But as a cross-sectional measure, it does not necessarily reflect the actual life-course experience of any specific cohort. It is entirely possible that any given cohort of Americans is living more years in a healthy state than the cohorts that preceded it.

Acknowledgements:
Special thanks to my colleagues Nikkil Sudharsanan and Arun Hendi, with whom I worked through the mathematics behind the paper and who provided valuable feedback. I also acknowledge the input of Professors Samuel H. Preston and Dylan Small of the University of Pennsylvania. An earlier version of this paper was submitted in partial fulfillment of the requirements for an MA degree in Statistics at the Wharton School, University of Pennsylvania.

Funding support:
Research was funded by Grant Number R01AG040212 from the National Institute on Aging at the National Institutes of Health. I declare no competing interests.

References

1. Preston SH, Heuveline P, Guillot M: *Demography: Measuring and Modeling Population Processes*. Oxford: Wiley-Blackwell; 2000.

2. Sullivan DF: **A single index of mortality and morbidity**. *HSMHA Health Rep* 1971, **86**:347–54.

3. Wachter KW: *Essential Demographic Methods*. Cambridge, MA: Harvard University Press; 2014.

4. Karpati T, Cohen CJ, Leibowitz M, Hoshen M, Feldman BS, Balicer RD: **Is the diabetes epidemic subsiding? Trends from a large population-based observational study in Israel**. *Popul Health Metr* 2014:1–8.

5. Cheng YJ, Imperatore G, Geiss LS, Wang J, Saydah SH, Cowie CC, Gregg EW: **Secular changes in the age-specific prevalence of diabetes among U.S. adults: 1988-2010.** *Diabetes Care* 2013, **36**:2690–6.

6. Chiang C: **A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions**. *Biometrics* 1960, **16**:618–635.

7. Hosmer DW, Lemeshow S, May S: **The Delta Method**. In *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Edited by Hosmer DW, Lemeshow S, May S. John Wiley & Sons, Inc.; 2008(X):355–358.

8. Salomon J, Mathers C, Murray C, Ferguson B: *Methods for Life Expectancy and Healthy Life Expectancy Uncertainty Analysis*. 2001. [*World Health Organization: Global Programme on Evidence for Health Policy*]

9. Efron B, Tibshirani RJ: *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC Press; 1993.

10. Murphy SL, Xu J, Kochanek KD: **Deaths: Final Data for 2010**. *Natl Vital Stat Reports* 2013, **61**.

11. Cunningham SA, Riosmena F, Wang J, Boyle JP, Rolka DB, Geiss LS: **Decreases in diabetes-free life expectancy in the U.S. and the role of obesity**. *Diabetes Care* 2011, **34**:2225–2230.

12. Shkolnikov V, Barbieri M, Wilmoth J: *Human Mortality Database*. 2014. [*Www.mortality.org*]

13. Anderson RN, DeTurk PB: *United States Life Tables, 1999. National Vital Statistics Reports.* Hyattsville, MD; 2002.

14. Arias E: *United States Life Tables , 2000. National Vital Statistics Reports. Volume 51*. Hyattsville, MD; 2002.

15. Arias E: *United States Life Tables , 2009. National Vital Statistics Reports. Volume 62*. Hyattsville, MD; 2014.

16. Arias E: *United States Life Tables, 2010. National Vital Statistics Reports.* Hyattsville, MD; 2014.

17. National Health and Nutrition Examination Survey: *National Health and Nutrition Examination Survey: Plan and Operations , 1999 – 2010*. Hyattsville, MD; 2013.

18. Bullard KM, Saydah SH, Imperatore G, Cowie CC, Gregg EW, Geiss LS, Cheng YJ, Rolka DB, Williams DE, Caspersen CJ: **Secular Changes in U.S. Prediabetes Prevalence Defined by Hemoglobin A1c and Fasting Plasma Glucose: National Health and Nutrition Examination Surveys, 1999-2010**. *Diabetes Care* 2013, **36**:2286–2293.

19. National Health and Nutrition Examination Survey: *National Health and Nutrition Examination Survey: Analytic Guidelines, 1999-2010*. Hyattsville, MD: Centers for Disease Control and Prevention; 2013.

20. Cowie CC, Rust KF, Byrd-Holt DD, Gregg EW, Ford ES, Geiss LS, Bainbridge KE, Fradkin JE: **Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006**. *Diabetes Care* 2010, **33**:562–568.

21. Stokes A, Mehta NK: **Mortality and excess risk in US adults with pre-diabetes and diabetes: a comparison of two nationally representative cohorts, 1988-2006**. *Popul Heal Metr* 2013, **11**:7.

22. Fishman EI, Stokes A, Preston SH: **The dynamics of diabetes among birth cohorts in the U.S.** *Diabetes Care* 2014, **37**:1052–9.

23. Bonora E, Tuomilehto J: **The pros and cons of diagnosing diabetes with A1C**. *Diabetes Care* 2011, **34 Suppl 2**:S184–90.

24. Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, Coresh J, Brancati FL: **Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults**. *N Engl J Med* 2010, **362**:800–811.

25. American Diabetes Association: **Diagnosis and classification of diabetes mellitus**. *Diabetes Care* 2010, **33 Suppl 1**:S62–9.

26. Preston SH, Glei D, Wilmoth J: **A New Method for Estimating Smoking-Attributable Mortality in High-Income Countries**. *Int J Epidemiol* 2010, **39**:430–438.

27. Keyfitz N: **Sampling Variance of Demographic Characteristics**. *Hum Biol* 1966, **38**:22–41.

28. Chiang C: *Life Table and Mortality Analysis*. Geneva: World Health Organization; 1979.

29. Mathers C: *Health Expectancies in Australia: 1981 and 1988*. Canberra: Australian Institute of Health; 1991.

30. Chiang C: *The Life Table and Its Applications*. Malabar, FL: Robert E. Krieger Publishing Company; 1984.

31. Chiang C: **A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions**. *Hum Biol* 1960, **32**:221–238.

32. Chiang C: **A stochastic study of the life table and its applications. III. The follow-up study with the consideration of competing risks**. *Biometrics* 1961, **17**:57–78.

33. Abatih E, Van Oyen H, Bossuyt N, Bruckers L: **Variance estimation methods for health expectancy by relative socio-economic status**. *Eur J Epidemiol* 2008, **23**:243–249.

34. Loukine L, Waters C, Choi BC, Ellison J: **Impact of diabetes mellitus on life expectancy and health-adjusted life expectancy in Canada**. *Popul Health Metr* 2012, **10**:7.

35. Mathers CD, Iburg KM, Salomon JA, Tandon A, Chatterji S, Ustün B, Murray CJL: **Global patterns of healthy life expectancy in the year 2002.** *BMC Public Health* 2004, **4**:66.

36. Gompertz BL: **On the Nature of the Function Expressive of the Law of Human Mortality**. *Philos Trans R Soc London* 1825, **115**:513–583.

37. Efron B: *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982.

38. Lumley T: *Complex Surveys: A Guide to Analysis Using R*. Hoboken, NJ: Wiley; 2010.

39. National Health and Nutrition Examination Survey: **Survey Methods and Analytic Guidelines**. *http://www.cdc.gov/nchs/nhanes/survey_methods.htm* 2014.

40. Alho JM, Spencer B: *Statistical Demography and Forecasting*. New York, NY: Springer-Verlag; 2005.

41. Schenker N, Parsons VL, Lochner K a., Wheatcroft G, Pamuk ER: **Estimating standard errors for life expectancies based on complex survey data with mortality follow-up: A case study using the National Health Interview Survey Linked Mortality Files**. *Stat Med* 2011, **30**(January):1302–1311.

42. Lynch SM, Brown JS: **A New Approach to Estimating Life Tables with Covariates and Constructing Interval Estimates of Life Table Quantities**. *Sociol Methodol* 2005, **35**:177–225.

43. Lynch SM, Brown JS: **Obtaining Multistate Life Table Distributions for Highly Refined Subpopulations from Cross-Sectional Data: A Bayesian Extension of Sullivan's Method**. *Demography* 2010, **47**:1053–1077.

44. Cai L, Hayward MD, Saito Y, Lubitz J, Hagedorn A, Crimmins E: **Estimation of multi-state life table functions and their variability from complex survey data using the SPACE Program.** *Demogr Res* 2010, **22**:129–158.

45. Willekens FJ, Putter H: **Software for multistate analysis**. *Demogr Res* 2014, **31**(August):381–420.

46. Crimmins E, Hayward M, Hagedorn A: **Change in disability-free life expectancy for Americans 70 years old and older**. *Demography* 2009, **46**:627–646.

47. Crimmins EM, Beltrán-Sánchez H: **Mortality and Morbidity Trends: Is There Compression of Morbidity?** *Journals Gerontol Ser B Psychol Sci Soc Sci* 2010, **66B**:75–86.

Notes for all tables:
LE = Life expectancy; DFLE = Diabetes-free life expectancy; Prevalence = prevalence of diabetes; Var.DFLE = estimated variance of DFLE; 1st term = first summation term in **Equation 9** in text; Var.LE = estimated variance of LE; Var.Prev = estimated variance of diabetes prevalence; SE.DFLE = estimated standard error of DFLE = square root of estimated variance. Source: Author's calculations from Human Mortality Database and National Health and Nutrition Examination Surveys, 1999-2000 and 2009-2010.

Table 1. Life expectancy and diabetes-free life expectancy at selected ages (Delta method)

A) U.S. Males, Year 2000

Age	LE	DFLE	Prevalence	Var.DFLE	1st term	Var.Prev	SE.DFLE
20	54.675	49.327	0.017	0.020	7.86E-07	0.00002	0.141
30	45.336	39.613	0.029	0.020	4.51E-07	0.00003	0.140
40	36.051	30.178	0.047	0.019	4.51E-07	0.00007	0.138
50	27.235	21.210	0.077	0.018	5.44E-07	0.00016	0.135
60	19.102	13.063	0.123	0.016	6.69E-07	0.00040	0.127
70	12.077	6.346	0.190	0.012	3.23E-07	0.00108	0.111
80	6.226	1.861	0.282	0.007	4.88E-08	0.00303	0.081
85	3.455	0.623	0.337	0.004	0	0.00482	0.065

B) U.S. Males, Year 2010

Age	LE	DFLE	Prevalence	Var.DFLE	1st term	Var.Prev	SE.DFLE
20	57.372	49.345	0.022	0.030	5.99E-07	0.00001	0.172
30	48.067	39.691	0.037	0.030	4.90E-07	0.00002	0.172
40	38.772	30.364	0.062	0.029	4.17E-07	0.00003	0.172
50	29.887	21.547	0.103	0.029	4.59E-07	0.00006	0.171
60	21.833	13.624	0.164	0.028	4.29E-07	0.00018	0.168
70	14.639	7.125	0.252	0.026	3.26E-07	0.00067	0.161
80	8.756	2.646	0.366	0.021	9.19E-08	0.00224	0.146
85	6.576	1.320	0.430	0.019	0	0.00349	0.137

C) U.S. Females, Year 2000

Age	LE	DFLE	Prevalence	Var.DFLE	1st term	Var.Prev	SE.DFLE
20	59.234	54.436	0.013	0.016	3.48E-07	0.00001	0.126
30	49.504	44.619	0.021	0.016	2.85E-07	0.00001	0.125
40	39.917	34.970	0.034	0.015	3.51E-07	0.00003	0.124
50	30.653	25.627	0.056	0.015	4.75E-07	0.00006	0.123
60	21.937	16.844	0.091	0.014	7.35E-07	0.00013	0.120
70	14.117	9.120	0.144	0.013	4.55E-07	0.00039	0.113
80	7.223	3.225	0.220	0.009	1.09E-07	0.00132	0.096
85	3.975	1.257	0.267	0.007	0	0.00239	0.084

D) U.S. Females, Year 2010

Age	LE	DFLE	Prevalence	Var.DFLE	1st term	Var.Prev	SE.DFLE
20	62.127	54.368	0.017	0.026	2.61E-07	0.00001	0.160
30	52.416	44.610	0.029	0.026	3.07E-07	0.00002	0.160
40	42.831	35.067	0.049	0.025	3.53E-07	0.00004	0.159
50	33.606	25.895	0.081	0.025	4.23E-07	0.00007	0.158
60	24.906	17.386	0.131	0.024	4.40E-07	0.00012	0.155
70	16.960	9.970	0.205	0.023	4.72E-07	0.00027	0.151
80	10.287	4.312	0.307	0.020	1.92E-07	0.00077	0.142
85	7.664	2.391	0.367	0.019	0	0.00131	0.137

Table 2. Life expectancy and diabetes-free life expectancy at selected ages (Monte Carlo simulations)

A) U.S. Males, 2000

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	54.064	48.932	0.018	0.392	5.15E-06	0.00002	0.62590
30	44.337	39.399	0.029	0.357	4.59E-06	0.00003	0.59734
40	34.895	30.254	0.047	0.310	3.77E-06	0.00006	0.55656
50	25.984	21.778	0.077	0.250	2.74E-06	0.00012	0.50049
60	17.975	14.366	0.123	0.182	1.69E-06	0.00025	0.42710
70	11.322	8.460	0.190	0.114	6.97E-07	0.00053	0.33720
80	6.403	4.382	0.282	0.056	1.75E-07	0.00111	0.23620
85	4.649	3.081	0.337	0.033	3.49E-06	0.00153	0.18278

B) U.S. Males, 2010

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	56.933	49.122	0.022	0.176	2.25E-06	0.00001	0.4195
30	47.246	39.672	0.037	0.173	1.95E-06	0.00001	0.4162
40	37.848	30.643	0.062	0.168	1.54E-06	0.00002	0.4097
50	28.963	22.311	0.102	0.157	1.07E-06	0.00004	0.3966
60	20.921	15.047	0.164	0.137	6.12E-07	0.00009	0.3707
70	14.151	9.273	0.252	0.105	2.15E-07	0.00025	0.3246
80	9.182	5.396	0.366	0.067	2.88E-07	0.00062	0.2593
85	7.679	4.370	0.431	0.052	2.45E-06	0.00087	0.2272

C) U.S. Females, 2000

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	58.441	53.854	0.013	0.224	5.73E-06	0.00001	0.4734
30	48.523	44.090	0.021	0.209	5.39E-06	0.00001	0.4572
40	38.733	34.544	0.035	0.188	4.72E-06	0.00003	0.4330
50	29.247	25.432	0.057	0.158	3.60E-06	0.00005	0.3977
60	20.419	17.141	0.091	0.121	2.13E-06	0.00010	0.3477
70	12.856	10.267	0.144	0.079	7.48E-07	0.00024	0.2819
80	7.435	5.563	0.221	0.044	3.68E-07	0.00059	0.2094
85	6.066	4.439	0.268	0.034	5.65E-06	0.00091	0.1831

D) U.S. Females, 2010

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	61.614	54.049	0.017	0.177	1.99E-06	0.00001	0.4210
30	51.729	44.370	0.029	0.160	1.82E-06	0.00002	0.4004
40	41.995	34.966	0.049	0.139	1.52E-06	0.00003	0.3727
50	32.584	26.066	0.081	0.114	1.08E-06	0.00005	0.3375
60	23.822	18.044	0.131	0.087	5.68E-07	0.00008	0.2956
70	16.254	11.441	0.206	0.062	1.66E-07	0.00013	0.2482
80	10.779	6.974	0.308	0.040	3.95E-07	0.00026	0.2002
85	9.446	5.969	0.368	0.034	2.76E-06	0.00038	0.1839

Table 3. Life expectancy and diabetes-free life expectancy at selected ages (Bootstrapping)

A) U.S. Males, 2000

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	54.675	49.332	0.020	4.677	5.21E-05	0.00030	2.1625
30	45.335	40.162	0.030	4.340	4.48E-05	0.00047	2.0832
40	36.051	31.157	0.047	3.863	4.02E-05	0.00077	1.9655
50	27.235	22.748	0.075	3.266	3.36E-05	0.00135	1.8071
60	19.102	15.205	0.120	2.507	2.59E-05	0.00271	1.5832
70	12.077	8.982	0.189	1.613	1.30E-05	0.00624	1.2702
80	6.227	4.241	0.286	0.664	5.01E-06	0.01402	0.8150
85	3.455	2.271	0.343	0.234	1.36E-05	0.01959	0.4836

B) U.S. Males, 2010

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	57.372	49.135	0.024	5.138	4.71E-05	0.00026	2.2667
30	48.067	40.021	0.039	4.897	4.13E-05	0.00044	2.2129
40	38.772	31.085	0.063	4.502	3.36E-05	0.00074	2.1219
50	29.887	22.752	0.103	3.950	2.71E-05	0.00130	1.9875
60	21.833	15.476	0.167	3.218	2.08E-05	0.00260	1.7940
70	14.639	9.429	0.259	2.197	1.25E-05	0.00583	1.4822
80	8.756	5.041	0.379	1.096	9.87E-06	0.01188	1.0470
85	6.576	3.654	0.444	0.667	3.63E-05	0.01544	0.8169

C) U.S. Females, 2000

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	59.234	54.274	0.014	4.276	3.89E-05	0.00016	2.0678
30	49.504	44.697	0.022	4.048	3.66E-05	0.00026	2.0120
40	39.917	35.341	0.035	3.724	3.25E-05	0.00043	1.9296
50	30.653	26.423	0.056	3.265	2.91E-05	0.00080	1.8071
60	21.937	18.217	0.092	2.620	2.13E-05	0.00171	1.6186
70	14.117	11.130	0.148	1.760	9.03E-06	0.00430	1.3266
80	7.224	5.327	0.231	0.727	3.62E-06	0.01080	0.8525
85	3.975	2.854	0.282	0.254	8.64E-06	0.01606	0.5040

D) U.S. Females, 2010

Age	LE	DFLE	Prevalence	Var.DFLE	Var.LE	Var.Prev	SE.DFLE
20	62.127	54.230	0.018	4.9330	3.02E-05	0.00016	2.2210
30	52.416	44.708	0.030	4.7547	2.61E-05	0.00027	2.1805
40	42.831	35.430	0.049	4.4777	2.31E-05	0.00046	2.1161
50	33.606	26.671	0.080	4.0655	1.94E-05	0.00082	2.0163
60	24.906	18.683	0.131	3.4209	1.60E-05	0.00168	1.8496
70	16.960	11.801	0.209	2.4493	9.35E-06	0.00409	1.5650
80	10.287	6.562	0.317	1.2931	7.89E-06	0.00951	1.1372
85	7.664	4.760	0.379	0.782	2.34E-05	0.01332	0.8844