Human Mortality and Chronic Disease Incidence at Extreme Ages: New Data and Analysis

Igor Akushevich, Ph.D., Kenneth G. Manton, Ph.D., Alexander Kulminski, Ph.D.
Center for Demographic Studies, Duke University
Durham, North Carolina 27708-0408, USA

We analyze age pattern of mortality and incidence of major age-associated diseases (e.g., cardiovascular diseases, cancer, diabetes etc.) in a population of the U.S. elderly with special emphasis on individuals above age 85 when a strong deviation from the Gompertz mortality function has been observed. We study the survival and health outcomes using nationally representative National Long Term Care Survey (NLTCS) longitudinal data collected from 1982 to 1999 because of the high quality of age reporting and over samples of persons aged 95+ in 1994 and 1999 (to be done again in the 2004 NLTCS). Diagnoses and onsets are identified using ICD-9-CM codes from Medicare service use files linked to NLTCS. To model these complex mortality and incidence patterns, we develop a new model with three parameters describing shape, location of maximum and random effect heterogeneity. This model provides excellent fits to the mortality and disease incidence patterns.

I.INTRODUCTION

Analyzing mortality rates at advanced ages in human populations, and inferring their relation to the biological processes that determine them is difficult because of errors in recording date of death and birth, incompleteness of human population vital registry systems, the biological complexity of the human organism and the heterogeneity of the distribution of those complex organisms in national populations. The importance of solving these problems has been emphasized by the frequent observation of human mortality trajectories at advanced ages that strongly deviate from commonly used models of adult mortality above, say, age 85 (e.g., the Gompertz (Lew and Garfinkel 1984, 1987; SOA 2000; Whittemore 1977)).

One strategy to deal with mortality data is to “smooth” data using a standard hazard model. This method is used by SSA actuaries when they fit a Gompertz hazard function to Medicare data at late ages (with gender differences in the shape parameter, i.e. 5% for males and 6% for females) who then report only “fitted” or “modeled” parameters in life tables at ages 95+. This makes those tables useless for demographic analysis of the biological basis of mortality at extreme ages.

An alternate strategy is to take what is viewed as the most reliable country data at late ages (e.g., Sweden) and then forcing the data from the other countries to fit models of that data using by statistical filters e.g., model life tables. This again makes the data artificial and inappropriate for describing the biological basis of late age mortality – all we can extract is the biological basis of mortality in Sweden.

These procedures only make sense when, because of the poor quality of data at advanced ages in many developing countries, there is little reliable information on which to base the selection of models more complex than the Gompertz or Weibull hazard. In

the U.S. at least, this limitation no longer applies due to a number of factors. One is the increasing proportion of births officially recorded – a process started at the beginning of the 20th century. Second, is the introduction of the Social Security program in 1933-1935. Persons entering the system used three items of self report information recorded at a reported age 65+ in 1933. Since this recording was 70 years ago, adults physically appearing to be, and reporting as being, age 65 in 1933, would be age 135 in 2004. Thus it seems implausible that a person who modestly misreported his age (say being age 55 instead of the reported age 65) would still be alive in 2004. If a person reporting being age 65 in 1933 had actually been age 45 he would now be age 115. Thus the room for error in the initial age reporting in 1933 having an effect at extreme ages in 2004 is becoming vanishingly small.

The National Vital Statistics Act was passed in 1907 though it was not universally applied then. What stimulated the universal coding of birth and death certificates in the U.S. was the effect of federal welfare programs developed after the great depression (Hetzel 1997). By the time of Medicare's introduction in 1965, thirty years had elapsed. Persons reporting being age 95 in 1995 would have been 65 in 1965. Thus, there is little chance that a person age 65 in 1965, who was aged 35 in 1933-1935 could have truly been less than 65. Persons aged 105 or less in 2004 are unlikely to have significantly inflated ages (i.e. persons aged 34 in 1933 and 65 in 1965). The accuracy of the mortality reporting system was also likely improved by Medicare's introduction in 1965 with underlying cause of death computer records available back to 1962 (multiple cause coding by computer was tested 1966-1967 and fully implemented in 1968). In addition, the U.S. has implemented a number of data quality studies (e.g., follow – back surveys) in a program likely to be more comprehensive than in a smaller country with a national registry system, like Sweden. Since Rosenwaike and Logues's (1985) book showed relatively little bias in age reporting up to age 100, as did Kestenbaum in 1992, we should expect, 12 years later, to have reasonably accurate data up to ages 112.

An alternate source of U.S. data comes from the private sector where corporate profits from life insurance depend on an accurate age reporting. Indeed any motivation for biases in age reporting in the private sector balances out with annuity programs being hurt by too high a life expectancy and life insurance hurt by too low a life expectancy. In estimating costs of annuities, annuitant experience for U.S. populations is used. Rates from insured lives after age 95 are assumed to reach a high constant value of 0.40 for males at age 106+, and for females at age 115+, based on eleven actuarial data sets of insured lives (SOA 2000). This is lower than the mortality rates in the SSA life tables at those ages, i.e., the two sets of life table mortality rates, one federal and one private, differ in level and shape. As the U.S. population ages, especially as the number of extreme elderly (e.g. 95+) grows, differences will grow in significance.

To age successfully, a person must not only survive but also have a good health. The issue of future trends in health and vital statuses in the national population with growing proportion of the elderly is a major governmental concern. To better address the health demands of the elderly and to reduce economical burden on society it is of importance to understand forces governing onsets of diseases and death. Key quantities addressing such issues are incidences of age-associated diseases, mortality rates and their age patterns.

We study human mortality and age-associated disease incidence above age 65, to extreme ages, 100+ in the U.S. population. The used dataset is the 1982 to 1999 National Long Term Care Survey where individuals, especially at late ages, have been followed up to 21 years (1982 to 2003) with continuous recording of health service use, age of death in Medicare records and detailed reassessment of health status by survey every five years. Persons in NLTCS samples were recruited into the Medicare system at age 65 for persons up to age 103 (i.e. in 1965 when Medicare was established). Older person's records were likely transferred from Social Security files with that system starting in 1933. The sample was enhanced in the 1994 NLTCS with 540 persons over age 95. The supplement in the 1999 NLTCS was 600 cases (planned in 2004; 1500+ cases). We will use, as a referent, SSA constructed life tables (SSA 2003), life tables for the experience of annuitants produced by the Society of Actuaries (SOA 2000), and U.S. mortality data from 1982 to 1999 from the Human Mortality Database (Wilmoth, 2005).

II.MODEL

Perhaps the most frequently used model for studying the age trajectories of adult human failure processes is the Gompertz function, (Strehler 1977) or,

$$\mu_G(t) = \alpha e^{\theta t}, \qquad (1)$$

which can be estimated from the time to failure event distribution for a cohort population. In (1) α (the scale parameter) is the mortality rate at t = 0 and θ (the shape parameter) is the percentage of per year increase in mortality rates. Also used is the Weibull function:

$$\mu_W(t) = \beta t^{m-1}. \qquad (2)$$

Parameters are: β, scale parameter; *m,* shape parameter which is related to the number of mutations in a multi-hit model of carcinogenesis (Armitage and Doll 1961). Rosenberg et al (1973) applied (2) to human mortality data and interpreted its coefficient *m*, not as the number of genetic "hits" or genetic errors, as in cancer mortality (Armitage and Doll 1954, 1961), but as related to protein thermodynamics, i.e. energy necessary to de-nature proteins critical to survival.

There have been various attempts to rationalize (1) as a biological theory, e.g. Sacher and Trucco (1962) and Strehler and Mildvan **(**1960). Rosenberg et al. (1973) compared the empirical behavior of both the Weibull and Gompertz functions assuming mortality was due to the thermodynamics governing protein denaturation induced by heat stress. Heat stress, related to core body temperature and basal metabolism, is also related to caloric restriction and endocrine (thyroid) function. Unfortunately, those models require the organism to be homogeneous with no internal structure – obviously inappropriate for humans.

Specifically, the Strehler-Mildvan theory does not describe heterogeneity (e.g., individual genetic differences) in risk nor does it show how such physiological mechanisms could be modeled. One approach to describing heterogeneity is to assume α follows a theoretically specified distribution such as the gamma or inverse Gaussian distribution (Manton, Stallard and Vaupel 1986). Though empirically better behaved at late ages such a function is still limited to extracting information from the age at failure event distribution under the assumption of a static heterogeneity

distribution. This can be constructed with models describing processes of health change prior to event. Functional forms have to be rationalized from other data.

We use a more general model of the effects of population risk heterogeneity (Vaupel et al 1979, Manton et al., 1986, Manton, Lowrimore and Yashin 1993),

$$\mu(t,\beta,m,n,\gamma)=\frac{\mu_0(t)}{\left[1+n\gamma\int_0^t du\cdot\mu_0(u)\right]^{\frac{1}{n}}} \tag{3}$$

The denominator in (3) shows the age increase in mortality as the most susceptible person dies (or become ill) first changing the distribution functions by systematic selection of fixed traits.

The additional parameters are γ, the squared coefficient of variation of the distribution of individual frailty with *n* controling the shape of the distribution e.g. $n=1$ or $n=2$ correspond to the gamma distribution or the inverse Gaussian distribution. $\mu_0(t)$ can be modeled as Gompertz $\mu_G(t)$ or Weibull $\mu_W(t)$ distributions. Thus, this function (3) can be used to model a wide range of frailty distributions.

In our analyses, NLTCS data are used to calculate one-year hazard rates, μ. To model the hazard rate for *k* years we integrate (3) over ages *t* to *t+k*,

$$h(t,\beta,m,n,\gamma)=\int_t^{t+k} du\,\mu(u,\beta,m,n,\gamma)\,. \tag{4}$$

Explicitly, we have for $\mu_0(t)=\mu_W(t)$

$$h_W=\frac{1}{(n-1)\gamma}\left[\left(1+n\beta(t+k)^m\right)^{1-\frac{1}{n}}-\left(1+n\beta t^m\right)^{1-\frac{1}{n}}\right], \tag{5}$$

where

$$\beta=t_0^{-m}\frac{m-1}{m(1-n)+n}\,. \tag{6}$$

Formula (6) links β with a new parameter t_0 – age at which the hazard reaches a maximum. It is more convenient to analyze the function (and calculate the age for the maximum hazard) in the space defined by t_0 rather than β. The maximum does not always exist – it may exist only with constraints on other parameters. This follows from β in the RHS of (6) being positive.

We analyzed NLTCS data using fixed and variable *n*. The fit of the model was improved as *n* declines. The best fit was achieved when $n\to 0$. This limit can be written explicitly. For a Weibull hazard, and the event probability, respectively, we have,

$$h_W(t,t_0,m,\gamma)=\frac{1}{\gamma}\exp\left(-\left(\frac{t}{t_0}\right)^m\frac{m-1}{m}\right)\left[1-\exp\left(-\frac{(t+k)^m-t^m}{t_0^m}\frac{m-1}{m}\right)\right], \tag{7}$$

$$\mu_W(t,t_0,m,\gamma)=\frac{(m-1)}{t_0\gamma}\exp\left(-\left(\frac{t}{t_0}\right)^m\frac{m-1}{m}\right)\left(\frac{t}{t_0}\right)^{m-1}. \tag{8}$$

A similar approach starts with (3). In this case,

$$h_G(t,t_0,\theta,\gamma)=\frac{e^b}{\gamma}\left(\exp\left[-be^{\theta t}\right]-\exp\left[-be^{\theta(t+k)}\right]\right), \quad (9)$$

$$\mu_G(t,t_0,\theta,\gamma)=\frac{b\theta}{\gamma}\exp(b(1-e^{\theta t})+\theta t), \quad (10)$$

where *b* is related to the Gompertz scale parameter α and can be equated (as for the case of Weibull) to the age at which the maximal hazard value t_0 is reached,

$$b=e^{-\theta t_0}\frac{\theta}{e^{\theta}-1}. \quad (11)$$

Does a distribution corresponding to *n=0* exist? It does. The general form of a model (3) with Gompertz $\mu_0(t)=\mu_G(t)$ can be written as $\mu=\overline{z}\,\mu_G(t)$, where the mean frailty,

$$\overline{z}=\frac{\int_0^\infty dzf(z)ze^{-zH}}{\int_0^\infty dzf(z)e^{-zH}}=e^{-\gamma H}, \quad (12)$$

has to be satisfied for any cumulative hazard function *H.* Expanding $\int dzf(z)(ze^{-zH}-e^{-(z+\gamma)H})$ into a series over *H* and combining coefficients for H^n produces a recurrence relation for moment calculations:

$$M_{n+1}=\sum_{k=0}^{n}C_i^n\gamma^{n-k}M_k. \quad (13)$$

with initial conditions $M_0=M_1=1$ (C_i^n — binomial coefficients). The second central moment is equal to γ as for the gamma model (for *n=1*). We can illustrate by expanding,

$$\int dzf(z)(ze^{-zH}-(1+n\gamma H)^{(-1/n)}e^{-zH}) \quad (14)$$

for $n\to 0$. $M_0=M_1=1$ and $M_2=1+\gamma$ independent of *n*. This dependence starts from moments of third order:

$$M_3=1+3\gamma+\gamma^2(n+1), \quad (15)$$

with the relation to fourth and higher order moments being derivable.

III.RESULTS

The NLTCS was started in 1982 with roughly 20,000 individuals examined in 1982, 1984, 1989, 1994, and 1999. Sample size was maintained by sampling roughly 5,000 persons passing age 65 between survey dates to replace persons lost due to death. In 1994 540, and in 1999, 600 persons over age 95 were drawn to improve disability and mortality estimates at late ages. In all five NLTCS roughly 42,000 individuals were examined with roughly 25,000 deaths recorded to 2003, i.e. about 400,000 person years of exposure over age 65 and over 100,000 person years of exposure over age 85.

The NLTCS database includes individuals who entered into NLTCS at different times and at different ages; some of them did not fail during the survey; the information of several samples is still under investigation, so different censoring schemes have to be considered. Therefore, hazard rates can be conveniently calculated using the Kaplan-Meier approach and taking into account a right-censoring scheme and left truncation. For our problem this means that we have to calculate the individual duration

of observation in the NLTCS and Medicare rather than to use life tables. There are two advantages of this approach. First, we avoid making assumptions about the size of the steps what is necessary in the life tables; and, second, the calculation will be accurate to within one day. The latter fact is important because our data allow us to have this accuracy. Specifically, date of birth, onset, death, enrollment/disenrollment into the NLTCS and Medicare are known with one-day accuracy. If an individual did not fail and a censoring scheme is used instead of vital statistics data, the date of the last day under observation (e.g., date of last record in NLTCS or Medicare files) is also known with one-day accuracy.

Direct projection of estimates from the NLTCS data to the U.S. population of the elderly (65+) is biased by the design effect. To have estimates that are representative of the U.S. elderly, sample design effects are taken into account using CDS screener weights (Manton and Gu, 2001). The sample weight function for each individual can undergo jumps when, for example, a new wave starts. The date when the sample weight function jumps, is known with one-day accuracy. Therefore, each individual in the NLTCS can be associated with precise time interval under observation $[t_{i1};t_{f1}]$ with corresponding weight function and life history extracted from Medicare files. Hazard (e.g., mortality and incidence) rates will be assessed by stratifying the sample relevant age categories (one month, year, or several years). Cumulatively, empirical risks will be calculated as follows: $h_1 = \sum_i n_i \Big/ \sum_i PYR$, where n_i is the number of cases, PYR_i is the number of person-years. Standard error (SE) will be estimated as SE for binomial distribution with adjusting for sample design effect (Manton et al., 1997).

Important subtask of our analysis will be careful investigation of the censoring effects. It is known that risk calculated using Medicare files can be underestimated. This can be, for example, because of uncontrolled death or disease onset occurrence abroad. As discussed by Kestenbaum (1992) such underestimation can strongly bias estimates for oldest-old (85+) and centenarians (100+). Although the Kestenbaum results are 12 years old and the data has further improved, we will pay special attention to this problem because of focus of this analysis on extreme ages.

We will consider two alternative censoring schemes. One is when the final day is defined from the Medicare vital statistics file (August 6, 2003 at the moment), and the other when this day corresponds to the last record either in Medicare claim or in NLTCS files.

III.a.Mortality

The contradictions in the description of national mortality projections provided by SoA and SSA, and disagreement with U.S. mortality data, as modeled by NCHS, is a sufficient motivation for verification of U.S. mortality trends using alternative data (Manton 2004). Such data must be nationally representative and followed for a long time. The NLTCS has 5 waves, i.e., 1982, 1984, 1989, 1994 and 1999 (a sixth is in process for 2004 with complete mortality for 2004 available in late 2005 or early 2006). 20,000 individuals are in the sample at each wave. The 2004 NLTCS can be used to confirm results of this analysis.

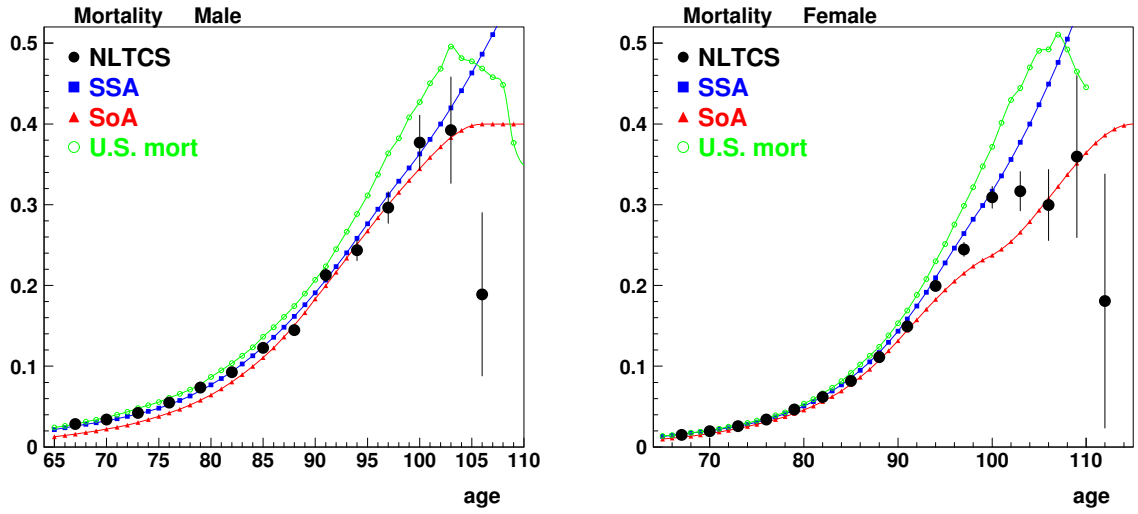

Figure 1. *Hazard rates calculated for SSA, SoA, U.S. and NLTCS mortality data.*

Figure 1 reproduces mortality results obtained from U.S. mortality data, SSA, and SoA as well as calculations of NLTCS hazard rates. We used two different techniques for censoring events described above but found no significant differences in the results.

The hazard rates calculated from SoA reach a plateau for both males and females. This is not the case for SSA projections which use a Gompertz function at ages 95+. NLTCS life tables also show a plateau. NLTCS data closely coincide with SSA projections in the age range 65 to 97. After age 95 the data for the SSA life tables are generated by a Gompertz. Since this essentially differ from the NLTCS predictions, we speculate SSA hazards are overestimated. NLTCS mortality rate is systematically lower than for the U.S. mortality data starting from age 85. Peak for mortality risk is about age 105, what is in agreement with U.S. mortality data.

Then we model mortality trends in the NLTCS using the models presented above. A model capable of describing plateau effects, and even declines of the hazard rate at advanced ages, was proposed by Manton et al (1993). This is a gamma model with the Weibull or Gompertz describing age dependence of the mortality rate. Parameter n in the model corresponds to different distributions of the unobservable frailty variable. For $n=1$ and 2 we have the gamma and inverse Gaussian distributions, respectively. Value $n=1/2$ was used by Manton et al (1993) to fit mortality data caused by lung cancer. Hence, we begin with model (7) and (9) with three values of parameter n to fit NLTCS data. These models did not fit the data with the worse fit for $n=2$. It improved for smaller n. Consequently, we made n a free parameter to generalize the set of frailty distributions we can represent. Our analysis reveals model fit improved as $n \to 0$. This limiting case is useful for finding analytical expressions for hazard and recurrence formulae for all moments of the corresponding frailty distribution as was done in previous section.

The results of fitting this model with three free parameters (gamma and two parameters adjusting age distribution) are in Figures 2. Using the new generalized distribution, fits are improved for the Gompertz. Moreover, t_0 appears more realistic for

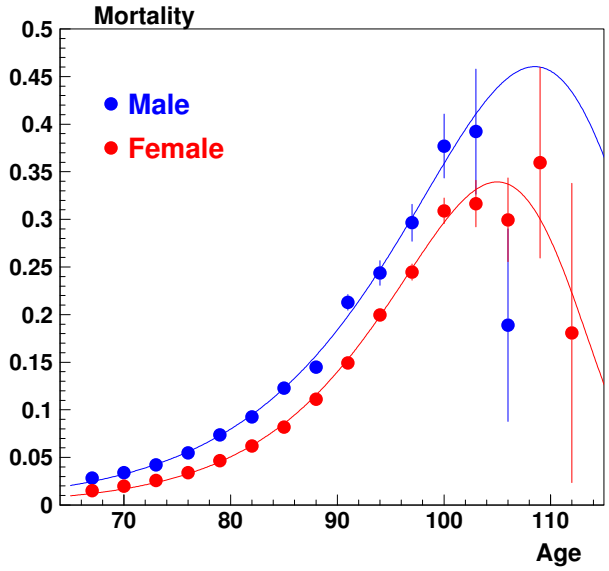

Figure 2. *Age specific mortality rates for males (blue) and females (red) with fitting by generalized hazard Gompertz model (9).*

the Gompertz. Fitting coefficients in the model show several effects. First, values of the Gompertz shape parameter (i.e. 0.10 to 0.12) are larger than for the usual fits of the Gompertz up to age 85 – but which underestimate mortality at early ages. This suggests the rate of aging of people dying in this age range is higher than for people dying later. γ represents heterogeneity (selection, as might be expected, reduces this). The issue is whether the greater heterogeneity is due to measured variables or is theoretically justified. This suggests peak mortality risk occurs about age 105 for female and age 108 for males.

III.b.Chronic Disease Incidence

Calculation of an incidence rate requires dates of individual disease onset, which will be extracted from 1984-2001 Medicare files linked with NLTCS. An individual will premably have an onset of certain disease in his/her observation period if there is at least one record with ICD-9-CM code corresponding to the disease on a single institutional claim (inpatient, skilled nursing facility, home health care, hospice, and outpatient) or non-institutional claim (carrier/physician supplier/Part B (1991-2001 only), durable medical equipment, clinical labs) on service for which beneficiary received medical care. Actually, Medicare data do not contain information whether appearance of ICD-9 code is an onset or not. Therefore, to determine a date of onset we will assume, that beneficiary with chronic condition certainly receives medical care at least once within first 6 months since his/her enrollment into Medicare program. Therefore, if certain diagnosis appears in Medicare files within an initial 6-month period, such individual will be considered as already been chronically impaired at the time of his enrollment in Medicare. Otherwise the date of first appearance of the corresponding diagnosis will be considered as the date of onset.

In this work we focus on the following age-associated diseases:

- ***Cancer,*** which includes all malignant neoplasms (140-208);
- ***Diabetes,*** mellitus which includes both insulin dependent and non-insulin dependent forms (250);
- ***Cardiovascular diseases*** (CVD), which include acute rheumatic fever (390-392), chronic rheumatic heart disease (393-398), hypertensive heart disease (402), hypertensive heart and renal disease (404), ischemic heart disease (410-

414), diseases of pulmonary circulation (415-417), and other forms of heart diseases (420-429);

- ***Cerebro-vascular disease*** (430-438);
- ***Neurodegenerative disorders*** (NDD), which include psychoses (290-299), nonpsychotic mental disorders (300-316), and hereditary and degenerative diseases of the central nervous system (330-337).

For calculation and modeling of incidence rate we use the same approach as for analysis of mortality risk. This approach includes Kaplan-Meier estimation for empiric incidence rate and the model developed above for fitting.

Medicare claim data have certain limitations which concern with determination of the diagnoses. One of the most serious issues is associated with lack of information on diagnoses (ICD codes) from Physician/Supplier/Part B source before 1991. Naïve usage of the data in this time domain can result to systematical underestimating the number of diagnoses. On the other hand, there are reasons for systematical overestimation of the number of diagnoses. In particular, this can occur when diagnoses are registered not within the observation period, so first record might be a false date of disease onset. Other reasons concern with efficiency of registration related, for example, with age dependent refusing to pass annual medical exam by elderly. Such sources of uncertainties have to be estimated using additional experimental information. Therefore we begin checking the order of such uncertainties.

Evident sources for uncertainties related with overestimation and shift of dates of onsets are enrollment of new beneficiaries and alteration of coverage by Medicare program of certain beneficiaries within the observation period because of legal (eligibility) or personal (enrollment under another health insurance) reasons. Actually, enrollment of new beneficiaries does not lead to the overestimation in this analysis because of a) the 6 month cut used to determine onsets and b) the fact that Medicare data cover larger time domain than domain of individuals' observation for the last two waves. Medicare data are available from 1984 for first three waves and from 1991 for last two. To project of estimates of incidence rates to the U.S. population we use the design weights, which become nonzero only after first survey in which individual participates. Thus, participants of fourth wave (1994) contribute to the incidence pattern only beginning from 1994 (when their weights become nonzero), but their diagnoses are analyzed from 1991, that provides sufficient time period to avoid such bias. Problem with 1984 year is also not so important because of relatively small amount of retained onsets when we apply the 6 month cut. This was directly tested by estimating of differences between incidence distributions calculated from 1984, 1985, and 1986 years, which appear to be negligible. Another source for uncertainties related with partial coverage does not provide essential bias, because of relatively small fraction of individuals who are not under part A&B coverage. For example, in March 2001 only 1.2% of surveyors were not entitled under Medicare, 3.4% were under Part A only, 0.21% were under Part B only, while 84.75% had Part A and Part B coverage, and additionally 9.9% were in the state "pay-in" of Part A and Part B. About 15% of all surveyors were under HMO.

We continue with estimating the effect of lack of medical information in Physician/Supplier/Part B source for the period from 1984 to 1990. Missing this information might result in loosing diagnoses or their shift to later ages. We tested

significance of this effect by comparison of the incidence patterns for the period from 1992 to 2001 with and without information from this source. Result of such estimation is that physician diagnoses account about 30-40% of the total diagnoses. The differences between incidence rates calculated for different diseases for the period 1984-1990 (where Physician/Supplier/Part B information is not available) and 1992-2000 are similar and attributed to the same reason. The physician-associated loss of diagnoses is practically disease independent. It is slightly larger for NDD, and slightly smaller for diabetes. It is also independent of age with exception of very advanced ages (about 100) when the difference practically vanishes. However, this effect is sex-dependent with larger loss of the diagnoses for females practically for all diseases and age regions. This conclusion on essential loss of diagnoses due to lack of medical information from Physician/Supplier/Part B source is also confirmed by analysis of time distribution of new diagnoses, in which an abnormally large peak appears in 1991 for participants of the whole NLTCS cohort and of the first three (82, 84 and 89) waves. Consequently, we will focus in our analysis on data from 1992 to 2001.

The results of estimation of age specific disease incidence rates are presented in Figure 3. It is seen that incidence rates of CVD and NDD monotonically increase with age. This occurs because CVD and NDD groups cover a wide range of specific diseases. Consequently, relatively few individuals at advanced ages do not have diagnoses corresponding to diseases from these groups that significantly decreases total number of person-years. Focusing on specific diseases from these groups we observe less pronounced increase or even decline of incidence rates at advanced ages that results in a peak in the age pattern. In contrast, the incidence rates associated with diabetes remain practically constant with age.

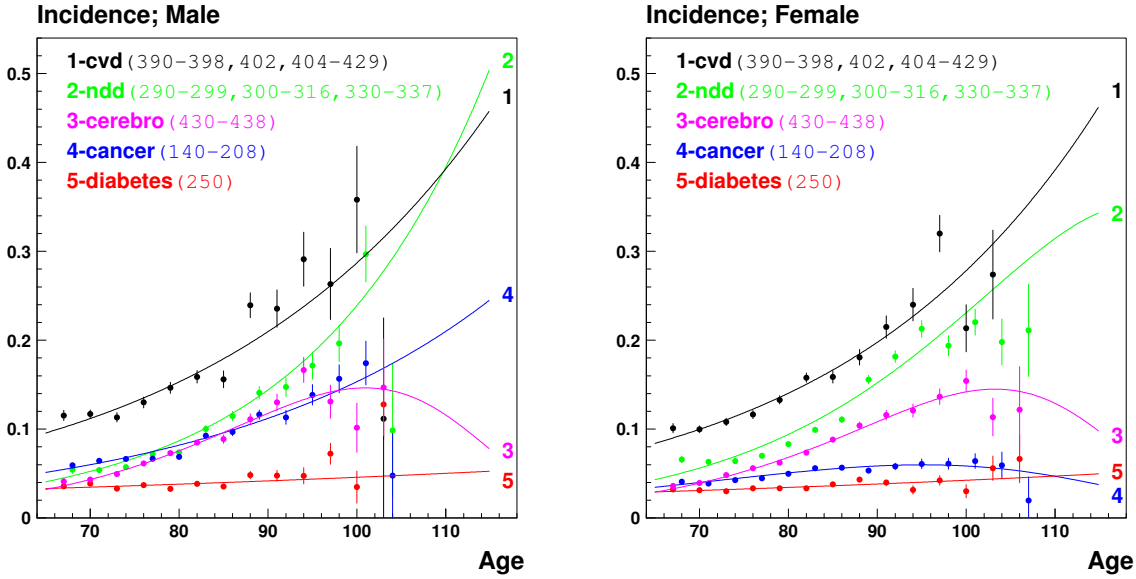

Figure 3*. Age specific disease incidence: dots denote Kaplan-Meier estimations (means and SE's) of NLTCS/Medicare data for 1992-2001, lines are the generalized Gompertz model (9).*

To estimate time trend from 1984 to 2001 we evaluated incidence rates for 1984-1990 and for 1992-2001 (no Physician/Supplier/Part B). Their comparison shows that

there is no statistically significant difference between the rates for these two periods with important exception of NDD. For NDD incidence in 1992-2001 is found to be as twice as larger than that for whole age region and for both sexes.

To compare our results with those known in the literature we will focus on each group of diseases emphasizing i) shape of the incidence age pattern, (e.g., appearance of a peak in the pattern and corresponding age), ii) absolute incidence level, and iii) sex differences.

Search of the literature reveals lack of nationally representative data on incidence age patterns on many diseases. The most reliable data on CVD were recently published by Arnold et al (2005). They used Cardiovascular Heart Study data to estimate incidence rates of major CVD in older Americans. Our estimates will be compared with those focusing on myocardial infarction (410) and stroke (436). The most detailed data on cancer come from SEER Register for invasive cancers. Although we did not distinguish between total malignant cancer incidence and its invasive forms, we hypothesize that both forms are adequately addressed by considering onsets for inpatients only. Therefore, for the purpose of comparison we have calculated age patterns of the incidence rates using all sources for onsets and inpatient only subset. The results are presented in Figure 4. As we can see there is an excellent agreement for cancer when we take into account only inpatient onsets. The rates with all onsets are higher that can be attributed to the contribution of non-invasive cancer forms. Reasonably good agreement is also seen for incidence rate of myocardial infarction for males considering all sources of onsets. For females the results are in better agreement with inpatient rates. Incidence rates for stroke are in pure agreement with those provided by Arnold et al (2005). This can be attributed to uncertainties in both studies related with inadequate quality of self-report data and mismatch in used ICD codes.

Incidence rates of diabetes were in the focus of many studies. However there are only few results showing their age distribution for 65 and older. The most comprehensive results are provided by McBean et al., (2004) who examined diabetes prevalence, incidence, and mortality from 1993 to 2001 among fee-for-service Medicare beneficiaries aged 67+ using 5% random sample. Shape and absolute level of incidence rates are in agreement with those by McBean et al., (2004) (not shown because their results are not sex-specific).

Since not for all diseases there is clearly expressed peak the incidence patterns associated with certain age, the model fitting these data has to be quite flexible to capture both monotonic growth and non-monotonic change. Both these situations are captured by our model when we fit these patterns. For those fits which reproduce peaks the parameter t_0 describing age at maximum of incidence rates are 95 for cancer for females; 101 and 103 for male and female cerebro-vascular disease. Peaks become more pronounced if we choose more specific diseases as shown in Figure 4. General conclusion from this figure is that peaks appear at significantly younger ages than peaks for mortality. One possible reason for such behavior can be attributed to the effect of selection (Vaupel et al., 1998), when frail individuals do not survive by these advanced ages. Another explanation is under-registration of the diagnoses at advanced ages. Other possible biologically and physiologically motivated explanations are discussed by Ukraintseva and Yashin (2001).

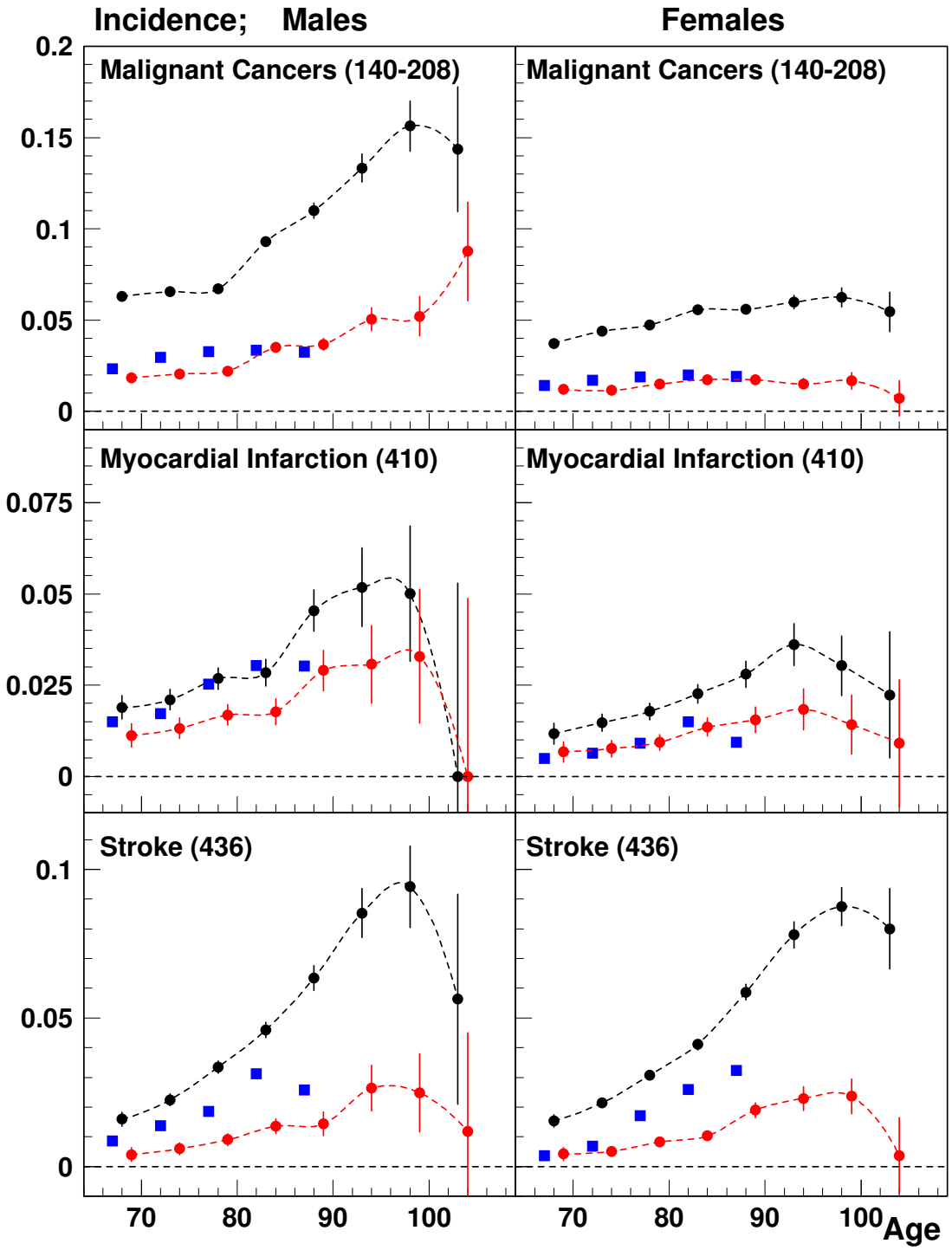

Figure 4. *Age pattern of the incidence rates for selected age-associated diseases. Black (red) dots correspond to the rate obtained using all sources (inpatient only). Blue squares denote SEER based incidence rates (Ries et al., 2004) for invasive cancers and the incidence of myocardial infarction and stroke from analysis by Arnold et al. (2005).*

IV. CONCLUSION

We present a more general model for heterogeneity hazard rates which provides a better fit to mortality above age 95 in the NLTCS data. The data are consistent with, not only a plateau effect, but with declines in the per annum hazard rate among survivors to ages 100+. We developed a model which further generalizes a concept of frailty and allows us to capture effect of decline of hazard rate with age. Mortality data are consistent with SSA projection up to ages 95-100 and then demonstrate platue and decline behavior, what is closely to SoA predictions. NLTCS mortality rate is systematically lower than for the U.S. mortality data starting from age 85. Peak for mortality risk found by the generalized Gompertz model is 105 for female and 108 for males, what is in agreement with U.S. mortality data. Age pattern of incidences of age specific diseases was calculated for CVD, NDD, cancer, diabetes and cerebro-vascular diseases. Possible sources for systematical uncertainties were analysed and their contribution to age pattern of incidence was estimated. Comparison of age pattern with those known in the literature shows good agreement for cancer and myocardial infarction.

REFERENCES

Armitage, P. and R. Doll. 1954. "The Age Distribution of Cancer and a Multistage Theory of Carcinogenesis. *British Journal of Cancer* 8:12.

---------. 1961. "*Stochastic Models for Carcinogenesis.*" Conference. *Proceedings* of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.

Arnold A.M., B.M. Psaty, L.H. Kuller, G.L. Burke, T.A. Manolio, L.P. Fried, J.A. Robbins, R.A. Kronmal. 2005. Incidence of Cardiovascular Disease in Older Americans: The Cardiovascular Health Study. Journal of the American Geriatrics Society 2005 53:2 p. 211-8

Hebert, P. L., L. S. Geiss, E. F. Tierney, M. M. Engelgau, B. P. Yawn, and A. M. McBean. 1999. "Identifying Person with Diabetes Using Medicare Claims Data." American Journal of Medical Quality 14 (6): 270277.

Hetzel, A. 1997. "History and Organization of the Vital Statistics System." Pp. 1-75 in *U.S. Vital Statistics System* 1997, Bethesda, MD: NCHS.

Kestenbaum B. 1992. "A Description of Extreme Aged Population Based on Improved Medicare Enrollment Data." *Demography* 29(4):565-80.

Lew, E. and L. Garfinkel. 1984. "Mortality at Ages 65 and Over in a Middle-Class Population." *Transactions of the Society of Actuaries* 36: 257-95.

Lew, E. and L. Garfinkel. 1987. "Differences in Mortality and Longevity by Sex, Smoking Habits and Health Status." *Transactions of the Society of Actuaries* 39: 107-25.

Manton, K.G., E. Stallard and J.W. Vaupel. 1986. "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged." *Journal of the American Statistical Association*. 81(395):635-44.

Manton, K.G., G. Lowrimore and A. Yashin. 1993. "Methods for Combining Ancillary Data in Stochastic Compartment Models of Cancer Mortality: Generalization of Heterogeneity Models." *Mathematical Population Studies* 4(2)133-47.

Manton KG, Corder L, Stallard E. (1997). Chronic disability trends in the U.S. elderly populations 1982 to 1994. *Proceedings of the National Academy of Sciences* 94:2593-2598.

Manton KG, Gu X. (2001). Changes in the prevalence of chronic disability in the United States black and nonblack population above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences, U.S.A*. 98(11):6354-9.

Manton, K.G. 2004. "The Demography of Aging." Forthcoming. In *Principles and Practice of Geriatric Medicine Fourth Edition* 2004, edited by M.S.J. Pathy, J.E. Morley and A. Sinclair. England: John Wiley & Sons, Ltd.

McBean AM, Li S, Gilbertson DT, Collins AJ. (2004) Differences in diabetes prevalence, incidence, and mortality among the elderly of four racial/ethnic groups: whites, blacks, hispanics, and asians. Diabetes Care. 27(10):2317-24.

Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2001, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2001/, 2004

Rosenberg, B., G. Kemeny, L. Smith, I. Skurnick and M. Bandurski. 1973. "The Kinetics and Thermodynamics of Death in Multicellular Organisms." *Mechanisms of Ageing and Development* 2:275-93.

Rosenwaike, I. and B. Logue. 1985. *The Extreme Aged in America.* Westport, CT:

Greenwood Press.
Sacher, G. and E. Trucco. 1962. "The Stochastic Theory of Mortality." *Annals of the New York Academy of Sciences* 96: 985.
Singpurwalla, N.D. 1999. "*Statistical Methods in Software Engineering*." New York: Springer.
Social Security Association. 2003. "Life Tables for the United States Social Security Area 1900-2100." Actuarial Study No. 116. Distributed by Social Security Association.
Society of Actuaries. 2000. RP-2000 Mortality Tables. In:
http://www.soa.org/ccm/content/research-publications/experience-studies-tools/the-rp-2000-mortality-tables.
Strehler, B. 1977. "*Time, Cells and Aging.*" New York: Academic Press.
Strehler, B. and A. Mildvan. 1960. "General Theory of Mortality and Aging." *Science* 132: 14-21.
Ukraintseva, S.V., Yashin, A.I. (2001) How individual aging may influence human morbidity and mortality patterns, *Mech. Aging & Development* 122: 1447-1460
Vaupel, J.W., K.G. Manton and E. Stallard. (1979) "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16: 439-454.\
Vaupel JW, Carey JR, Christensen K, Johnson TE, Yashin AI, Holm NV, Iachine IA, Khazaeli AA, Liedo P, Longo VD, Zeng Yi, Manton KG and Curtsinger JW. Biodemographic trajectories of longevity. /Science./ 280(1998)5365: 855-860.
Whittemore A. 1977. "The Age Distribution of Human Cancer for Carcinogenic Exposures of Varying Intensity." *American Journal of Epidemiology* 106:418-32.
Wilmoth, JR. 2005. The Human Mortality Database**. http://www.mortality.org**