# Frailty Modeling of Canadian and Swedish Mortality at Adult and Advanced Ages

Kirill F. Andreev[a] and Robert Bourbeau[b]

Revised 13 September, 2006

[a]*United Nations, Department of Economic and Social Affairs,*
*Population Division, Mortality Section*
*New York, United States*
*E-mail: andreev@un.org*

[b]*Department of Demography*
*University of Montreal,*
*Montreal, Quebec, Canada*
*E-mail: robert.bourbeau@umontreal.ca*

Abstract ..... - 4 -
Introduction ..... - 5 -
Aalen-Hougaard Model with Gompertz baseline ..... - 7 -
Data ..... - 12 -
Model Fit ..... - 13 -
Effect of Age Misreporting and Fit to the Cohort Life Tables of Canada ..... - 16 -
Lack of Fit to Female Period Data ..... - 20 -
Conclusions ..... - 22 -
Acknowledgments ..... - 24 -
Disclaimer ..... - 24 -
References ..... - 24 -

## Abstract

A newly established Canadian database provides an opportunity to study survival in Canada at adult and advanced ages on a cohort basis. Age-specific mortality patterns at ages 65 and over have been analyzed with a four parameter frailty model incorporating Gompertz baseline mortality and Aalen-Hougaard family of frailty distributions. For comparison purposes the analysis has been extended to recent period data and to that of Swedish population. A special attention has been paid to goodness of fit and possible bias in parameter estimates introduced by common deficiencies frequently found in demographic data. The results suggest the frailty model fits the data significantly better than Gompertz due to deceleration in the rate of mortality increase over age. However, no particular form of this frailty model including traditional gamma-Gompertz model has emerged as the best fitting model in all cases. A particular type of the best fitting model depends highly on population, starting age of life table, and whether the model is fitted to period or cohort data. Moreover, a thorough analysis

of female period data suggests that even this general model is not able to capture age pattern of mortality for entire age range properly because of acceleration in mortality increase in the age interval from 65 to 80. The observed acceleration might be a manifestation of emerging mortality pattern in low mortality countries that is incompatible both with Gompertz and with the frailty models incorporating Gompertz baseline which are frequently employed for modeling death rates at adult ages.

## Introduction

Far-sighted collection of statistical data – which has been carried on by government statisticians since 17th-century to the present day – makes it possible today to study different aspects of human life span. Observed patterns of age-specific death rates serve as a foundation for developing mathematical models of human mortality in attempt to provide possible explanations of the observed data.

Age-specific pattern of human mortality is known to take a very specific form. One of the best known models of human mortality was proposed by Gompertz (1825) suggesting an exponential form for the observed force of mortality:

$$\mu(x) = ae^{bx} \tag{1}$$

The model showed a remarkable performance in depicting pattern of human mortality at adult ages. As more data on human mortality at advanced ages had been accumulated it was realized that at the very old ages (>80) rate of mortality increase as measured by the parameter *b* of Gompertz distribution is not constant anymore but rather declines over age. To account for such deceleration in the rate of mortality increase Perks (1932) suggested a family of mortality curves with the simplest member to be a logistic:

$$\mu(x) = \frac{Bc^x}{1 + Dc^x} \tag{2}$$

In this model death rates initially increase exponentially but then they level off and approach an asymptote. This model has proved to provide significantly better approximation compared to Gompertz to the observed death rates at advanced ages.

In search for an explanation of the logistic form Beard (1959) proposed compositional interpretation of the observed death rates. He assumed that the population under study can be stratified into groups of individuals subject to the same force of mortality within each group: $\mu(x, z)$ ; the variable *z* is a stratification or frailty variable associated with each group of individuals. As individuals in the high mortality groups die earlier than individuals in the low mortality groups the composition of population changes over time, and the observed death rates generally follow a trajectory that is different from the trajectories in each of the

population groups. Therefore, one can observe a logistic curve even though the death rates follow Gompertz in each of the groups.

Formally, one can assume that *z* is continuously distributed with some density $\phi(z)$. In this case the initial number of individuals with mortality $\mu(x,z)$ is equal to $\phi(z)dz$. Only $s(x,z)\phi(z)dz$ of them will survive to age *x* from birth ($s(x,z)$ is a survival function corresponding to $\mu(x,z)$). By summing up survivors from all groups to age *x* survival function for the total population can be computed as:

$$s(x) = \int_0^\infty s(x,z)\phi(z)dz \tag{3}$$

Similarly, death rates in the total population are given by

$$\mu(x) = \frac{\int_0^\infty \mu(x,z)s(x,z)\phi(z)dz}{\int_0^\infty s(x,z)\phi(z)dz} \tag{4}$$

where denominator of Equ. (4) is simply a survival function given by Equ. (3), and numerator is sum of deaths occurred in each group at age *x*.

Equations (3) and (4) is essence of any fixed frailty model. The term "fixed" stems from observation that the variable *z* fixed at birth is an age independent attribute of survival of an individual. By selecting different parametric forms of $\mu(x,z)$, $s(x,z)$ and $\phi(z)$ different parametric models for $\mu(x)$ and $s(x)$ can be obtained.

Beard further assumed that the death rates vary proportionally between groups:

$$\mu(x,z) = z\lambda(x) \tag{5}$$

with $\lambda(x)$ often referred an individual mortality or baseline hazard. The coefficient of proportionality *z* now accounts for survival differences between sub-populations: if $z > 1$ mortality of this group is higher, if $z < 1$ mortality is lower and if $z = 1$ death rates in this group follows $\lambda(x)$ or baseline mortality.

Further assumptions have been made regarding parametric forms of $\lambda(x)$ and $\phi(z)$. As death rates in human populations increase almost exponentially after age about 35 it was reasonable to assume that $\lambda(x)$ follows Gompertz model. The frailty $z$ was assumed to be gamma distributed mostly on grounds of mathematical convenience (Beard, 1959). This leads to the following equation for force of mortality in the total population:

$$\mu(x) = \frac{(p+1)\beta\lambda e^{\lambda x}}{(\gamma\lambda - \beta) + \beta e^{\lambda x}} \tag{6}$$

where $p, \gamma$ and $\beta, \lambda$ are the original Beard's parameters of gamma and Gompertz distributions, respectively (Makeham term is not included here as it is close to zero in the contemporary low mortality populations).

It is easy to see that Equ. (6) follows a logistic form equivalent to Equ. (2) thus providing necessary flexibility to capture the deceleration in the rate of mortality increase at old ages. Moreover, the model provides also an explanation for the observed pattern of death rates which makes it especially attractive. In demography this model was further developed by Vaupel, Manton and Stallard (1979) and Vaupel and Yashin (1985) which led to extensive applications in mortality research e.g. Thatcher, Kannisto and Vaupel (1998). In biostatistics further generalizations of this model have been obtained by Hougaard (1986), Aalen (1988) and Aalen (1992) by introducing more flexibility in a choice of the frailty distribution.

As more data on human mortality become available it is important to existing mortality models on the new data as they emerge. Only few countries in the world have mortality series starting well back in time to re-construct survival experience of genuine cohorts. With establishment of Canadian Human Mortality Database (2005) (CHMD) exploration of the cohort patterns of survival becomes also possible for population of Canada. In this work we further elaborated a particular form of Aalen-Hougaard model with Gompertz baseline and applied it for exploring dynamics of age-specific death rates in Canadian cohort mortality. The analysis was also extended to recent period data and to that of Swedish population.

## Aalen-Hougaard Model with Gompertz baseline

If we denote cumulative hazard of the individual mortality $\lambda(x)$ by $\Lambda(x)$ Equ. (3) can be expressed as

$$s(x) = \int_0^\infty e^{-z\Lambda(x)} \phi(z) dz \tag{7}$$

It was noted by Hougaard (1986) that (7) is Laplace transform of $\Lambda(x)$ so it would be mathematically convenient to work with frailty distributions derived from the stable distributions for which Laplace transforms can be found analytically. In this case $s(x)$ can obtained in close form:

$$s(x) = \exp\left\{ \frac{\alpha}{(1-\alpha)\delta} \left[ 1 - \left( 1 + \frac{\delta\gamma}{\alpha} \Lambda(x) \right)^{1-\alpha} \right] \right\} \quad \text{if } \alpha \neq 1,\ \alpha > 0 \tag{8}$$

$$s(x) = \left\{ \frac{1}{1 + \delta\gamma\Lambda(x)} \right\}^{\frac{1}{\delta}} \quad \text{if } \alpha = 1 \tag{9}$$

where $\gamma$ is mean of the Hougaard's frailty distribution, $\delta$ is a squared coefficient of variation, and $\alpha$ is a shape parameter. As further noted by Aalen (1992) this model can be extended for the case $0 \leq \alpha < 1$. In this case the frailty distribution does not belong anymore to the Hougaard's family rather it is distributed according to the compound Poisson distribution. As the model is not identifiable for any $\gamma$ and the mean is usually set to $\gamma = 1$ and $\mu(x)$ derived from (8), (9) is simply:

$$\mu(x) = \frac{\lambda(x)}{\left\{1 + \alpha^{-1}\delta\Lambda(x)\right\}^{\alpha}} \quad \text{with } \alpha, \delta \geq 0 \tag{10}$$

By assuming that the individual hazard is Gompertz

$$\lambda(x) = ae^{bx} \tag{11}$$

we obtain the following Aalen-Hougaard model with Gompertz baseline:

$$\mu(x) = \frac{ae^{bx}}{\left\{1 + \alpha^{-1}\delta\frac{a}{b}\left[e^{bx} - 1\right]\right\}^{\alpha}} \tag{12}$$

It is easy to see that the model (12) indeed describes selection in a heterogeneous population. As individuals with higher frailties die earlier the population survived to a certain age consist of more robust individuals, and the observed death rates $\mu(x)$ have to be lower than the individual mortality $\lambda(x)$. It follows from (12) that the ratio of total and individual mortality $\mu(x)/\lambda(x)$ declines over age from one to zero because the denominator of (12) is positive and increasing ($\lambda(x) > 0$). Formally, it can be shown that this ratio is mean of frailty distribution among survivors to age $x$: $E(z \mid X > x)$. As selection is operating in the population the mean declines from initial value of $E(z) = 1$ to zero following age trajectory given by denominators of (10) and (12).

A particular shape that $\mu(x)$ takes depends both on type of frailty distribution as measured by $\alpha$ and on amount of heterogeneity as measured by $\delta$. In Fig. 1 we show typical mortality curves implied by (12) for various $\alpha$ and $\delta$. Gompertz parameters of $\lambda(x)$

($a$= 0.025; $b$ = 0.09) have been chosen to be approximately equal to values of such parameters if the model (12) is fitted to Canadian mortality data for age 65 and over.

The panel a) of Fig. 1 shows mortality curves for different values of $\alpha$ ($\delta$ was held constant at 0.1). This is the most typical case if the model is fitted to the real data. If $\alpha = 0$ then $\mu(x) = \lambda(x)$ (it can be shown by taking limit of (10) with $\alpha \to 0$) and the curve is simply Gompertz which appears as a straight line on the log scale. In this case the frailty distribution is degenerate and population is homogeneous. The case $\alpha = 0.5$ shows typical mortality curves if $0 < \alpha < 1$. As follows from Fig. 1 death rates are increasing for the entire age range. Asymptotically the death rates continue to increase exponentially but the rate of increase is lower than at younger ages. By taking limit $x \to \infty$ we can find that

$$\lim_{x\to\infty} \mu(x) \approx \frac{a}{\left(\alpha^{-1}\delta\frac{a}{b}\right)^{\alpha}} \exp\{b(1-\alpha)x\} \tag{13}$$

and the asymptotic rate of increase: $b(1-\alpha)$. The rate appears to depend only on the rate of increase of the baseline hazard $b$ and on type of frailty distribution $\alpha$. The frailty distribution in this case belongs to Hougaard's family derived from the stable distributions. For $\alpha = 0.5$ the distribution is inverse Gaussian while for the other values of $\alpha$ no close forms are known.

For $\alpha = 1$ as it follows from (13) the asymptotic rate of increase is zero and we arrive at the traditional gamma-Gompertz model:

$$\mu(x) = \frac{ae^{bx}}{1+\delta\frac{a}{b}\left[e^{bx}-1\right]} \tag{14}$$

In this case the frailty z is Gamma distributed and $\mu(x)$ follows a logistic curve which approaches an asymptote $b/\delta$ at higher ages.

For $a > 1$ after initial increase $\mu(x)$ reaches maximum and then declines. As it follows from (13) the decline is exponential having a consequence that the cumulative mortality stops increasing at higher ages. As a result the survival function stops to decline reaching a value above zero as $x \to \infty$. Such survival functions are called defective as they imply that certain proportion of initial population lives forever. As it discussed by Aalen (1992) the frailty distribution in this case is compound Poisson which consists of two parts: one is continuous for $z > 0$ and one is discrete with a single atom at $z = 0$ with positive

probability $P(z=0)>0$. Individuals with frailty $z=0$ have zero risk of death that makes the survival function defective. By taking the following limit we can find $P(z=0)$: .

$$P(z=0)=\lim_{x\to\infty} s(x)=\exp\left\{\frac{1}{\delta}\frac{\alpha}{1-\alpha}\right\}$$

It follows from this equation that $P(z=0)$ depends entirely on the parameters of frailty distribution but not on parameters of individual mortality $\lambda(x)$. This implies that such behavior of a survival function for total population will be the case for any regular individual hazard introduced into Equ. (10). Indeed, all individuals with non-zero mortality will eventually die out leaving only proportion $P(z=0)$ of immortal individuals.

It is also easy to compute age at maximum of $\mu(x)$:

$$x_{\max}=\frac{1}{b}\ln\left[1+\frac{\alpha}{\alpha-1}\left(\frac{b}{a\delta}-1\right)\right]$$

The maximum does not always exist. Its existence depends on whether $\mu(x)$ is initially increasing. By examining rate of mortality of mortality increase $r(x)=\dfrac{d\ln\mu(x)}{dx}$ on can find that the initial rate of increase is equal to $r(0)=b-\delta a$. If this quantity is negative the death rates decline for the entire age range (recall that $\alpha>1$). This case is shown in Fig. 1, panel d) with $\delta=b/a$ and $\delta=10$.

It can be also shown, by taking the second derivative of $\mu(x)$ and observing that the derivative does not change sign that $\mu(x)$ can have at most one maximum. It implies that the only way for $\mu(x)$ to decline over age is by having parameter $\alpha$ to be greater than unity resulting in the defective survival function. For analysis of the real data it implies that if the declines in death rates at the highest ages can be reliably detected then the frailty interpretation based on the model (12) has to be dismissed as inconsistent due to the defectiveness of the survival function.

For large values of $\alpha$ there exists a limit of Equ. (10) which can be found in a general form by letting $\alpha\to\infty$:

$$\mu(x)=\lambda(x)\exp\left[-\delta\Lambda(x)\right] \tag{15}$$

$$S(x)=\exp\left[\frac{1}{\delta}\left(\exp\left[-\delta\Lambda(x)\right]-1\right)\right] \tag{16}$$

For this special case of (10) the frailty distribution converges to Poisson distribution—instead of partly continuous it becomes discrete (Aalen, 1992). The distribution has an atom at z = 0,

the proportion of individuals with frailty zero, with probability $P(z=0)=\exp(-1/\delta)$. The limiting form of Equ. (12) can be obtained by substituting Gompertz baseline into (15):

$$\mu(x)=ae^{bx}e^{-\delta\frac{a}{b}\left[e^{bx}-1\right]} \tag{17}$$

For this special case (12) the decline in death rates is both profound and early occurring as compared to the general model with $\alpha<+\infty$.

Effect of amount of heterogeneity in the population on death rates as measured by the parameter $\delta$ is shown in Fig. 1, panels b)-d). If $\delta\to 0$ the total mortality $\mu(x)$ converges to $\lambda(x)$ because the parameter $\delta$ is squared coefficient of variation of frailty distribution at age 0 or, in our case, simply the variance (recall that $\gamma=1$). By letting $\delta\to 0$ the frailty distribution degenerates. For $\delta>0$ the larger is the value of $\delta$ the larger is the effect on $\mu(x)$, consequently, the larger is deviation of $\mu(x)$ from $\lambda(x)$. The overall shape of mortality curves is nevertheless retained for different values of $\delta$.

A qualitative change in dynamics of mortality occurs at $\delta=b/a$. At this value of $\delta$ the initial rate of mortality increase $r(0)$ changes sign. If, for example, $\delta>b/a$ then $r(0)$ is negative and initially mortality declines. Following this decline the death rates either start

a) to increase ($a<1$, Fig. 1, panel b) or
b) reach an asymptote which might be higher or lower than the initial mortality value $\mu(0)$ or simply equal to $\mu(0)$, the case of constant mortality, if $\delta=b/a$ ($\alpha=1$, panel c) or
c) decline monotonically for all ages ($\alpha>1$, panel d)

The cases $\delta=b/a$ or $\delta<b/a$ can be described in a similar way by recalling that asymptotically $\mu(x)$ changes exponentially at a rate given by (13). The rate is independent of $\delta$, and its sign determined entirely by type of frailty distribution $\alpha$.

In a nutshell, all mortality shapes shown in Fig. 1 could be thought as transitions from one exponential curve to another—it appears visually as if $\mu(x)$ is following initially one Gompertz curve but over time it switches to Gompertz curve with different parameters and stays with it thereafter. The time of transition is governed by the amount of heterogeneity, $\delta$. If $\delta$ is small $\mu(x)$ stays on the first curve longer and if $\delta$ is large the transition to the second curve occurs earlier. In the most typical case—human populations with low initial mortality and $\delta\approx 0.1$ (Fig. 1, panel a) —the initial rate of increase is close to the rate of increase in the baseline $r(0)\approx b$ and $\mu(x)$ follows $\lambda(x)$ at ages close to zero. Then, as the

selection takes toll, $\mu(x)$ either switches to increasing Gompertz curve or to constant mortality or to declining Gompertz curve depending on the type of frailty distribution, $\alpha$ .

## Data

Canadian Human Mortality Database (2005) includes deaths, population and death rates for all years starting with 1921, and for all ages from birth to the highest age attained. The data are stratified by sex and by province. All data on deaths are custom tabulations from individual death certificates prepared by Health Statistics Division, Statistics Canada. The data on deaths have been tabulated either by single year of age and by year of birth or by single year of age only. Before 1950 death counts at age 100 and over are available only aggregated in open age group 100+. Starting with 1950 deaths are available for all ages making possible direct estimation of death rates above 100 by single year of age.

Canadian population estimates are published by Demography Division, Statistics Canada. The data are published at middle of each calendar year and by single year of age up to age 90+. The estimates originate from population censuses which are held in Canada every five years. Postcensal estimates are frequently revised until data from a new census becomes available; whenever it was possible, the latest population figures have been incorporated in the database. As the published population data are not detail enough for producing estimates of death rates at the highest ages, the population estimates above 80 have been computed by method of almost extinct generations (Thatcher, Kannisto and Andreev, 2002). The method takes only data on deaths supplemented by population estimates in the *last* year only as inputs for producing population estimates above 80. The death rates above 80 are based then almost entirely on deaths rather on published population estimates.

We downloaded the data from Human Mortality Database (HMD) (www.mortality.org) which includes a copy of mortality database for Canada together with similar data for more than 30 national populations. Data for each country are supplemented with originally published data on deaths and population used for mortality estimation. Complete reference information is available as well. For each country there are also so-called background and documentation files, also available online, which include detail review of available population and vital statistics, information on completeness of vital statistics and highlights of data quality checks. Methods employed for computing life tables and death rates is completely documented in Method Protocol also published online. Before conducting our analysis we have also analyzed internal data quality report made available to us by the

HMD staff. The report includes several hundreds graphs highlighting different aspects of the data: availability charts of original data, time trends and age-specific schedules of death rates, Lexis maps of age heaping tests among others.

For analysis proposed here we computed life tables for cohorts 1856–1893 for males and 1856–1889 for females, and for ages 65 and over. All cohorts are extinct so death rates can be computed up to the highest age attained by the members of respective cohorts. We decided to exclude the earlier cohorts incorporating deaths in open age groups 100+. Therefore the death rates in the cohort life tables are based entirely on population and vital statistics data available by single year of age.

The life tables have been constructed by the cohort method: for each age interval we computed population at risk and deaths aggregated over all cohorts. Necessary migration corrections for population at risk have been also applied. In addition we included life tables based on the most recent period data, 1990–2001, as their quality is higher than that of cohort life tables. We also included life tables for the same cohorts and periods of Swedish population commonly recognized for their quality. There is some evidence of age misreporting in the cohort life tables for Canada (see details below) so it is important to compare results of model fitting to that based on good quality data.

## Model Fit

Five models have been selected to be fitted to the real data: the general four-parameter Aalen-Hougaard model (12) (labeled as AH) and four special cases of this model. The first special case is the traditional gamma-Gompertz (GG) model (14). The next special case is further simplification of (14) obtained by equating $b = \delta$ which results in logistic curve for $\mu(x)$ with asymptote equal one. This model is also known as Kannisto model of mortality (Thatcher, 1999) and it was labeled accordingly as KN. The next special case considered here is (17) obtained from (12) by letting $\alpha \to \infty$ (labeled as AI), and finally we fitted Gompertz model (GO) (1) which is obtained from (12) by $\delta \to 0$ that is baseline mortality hazard.

If fit of the model (12) results in estimate of $\hat{\alpha} > 1$ frailty interpretation of the observed death rates becomes impossible as individuals with $z = 0$ are immortal in mortality context. Nevertheless, this case was included in our analysis as it permits to detect lack of fit of the model (12) if values $\alpha$ are limited to be less than one in order to obtain regular

survival functions. This case is also useful for testing a hypothesis whether death rates are declining at advanced ages (one need to ignore frailty interpretation of $\mu(x)$ in this case).

The models have been fitted to the eight cohort and period life tables described in the previous section. For a single life table the fit has been carried out to a progressively increasing sequence of ages: first the current model has been fitted to death rates at ages 65 and over, then to death rates at ages 66 and over and so on until age 85 and over. The selected scheme permits to check whether or not a single model is applicable for any age above 65 or only for older ages.Parameter estimates have been obtained by maximizing binomial log-likelihood function:

$$L = \sum_x D_x \ln q_x + (N_x - D_x)\ln(1 - q_x) \tag{18}$$

where $N_x$ is population at risk at age $x$, $D_x$ is number of deaths in age interval $[x, x+1]$, $q_x = 1 - e^{\int_x^{x+1} \mu(u)du}$ is age-specific probability of dying, and $\mu(x)$ given by (12). In the case of the AH model, for example, the $q_x$ is a function of four parameters: $a, b, \alpha, \delta$. Their values maximizing (18) are parameter estimates of this model, $\hat{a}, \hat{b}, \hat{\alpha}, \hat{\delta}$, fitted to the current life table. In this method of estimation ages with larger number of deaths receive more weight and have larger influence on parameter estimates. Visually, the observed $q_x^{obs} = D_x / N_x$ at such ages will be closer approximated by the model than that with smaller number of deaths.

In total the selected five models have been fitted to 8 * 21 = 168 datasets (8 life tables and 21 starting ages: from 65 to 85). For each dataset the best fitting model has been selected by the likelihood ratio test (LR) with 1% significance level if the models to be compared are nested. In our case the following comparisons can be made with the LR test: AH→GG→KN, AH→GG→GO, AH→AI. If the models to be compared are not nested as, for instance, GG and AI, we used Akaike Information Criterion (AIC) for selection of the best fitting model. This situation arises if the AH model is rejected by the LR test in favor of both GG and AI models.

Results are summarized in Table 1. Each cell of the table shows the best fitting model for a particular life table and age range. For facilitating comprehension of information in this table each model has been also associated with a specific background color. For example, the first cell of the table (AI with yellow background) indicates that the best fitting model for male population of Canada if fitted to ages 65 and over is the AI model. By fitting the models to ages 66 and over the best fitting model is gamma-Gompertz (GG), and by fitting

the models to age 67 and over the best fitting model is general four-parameter Aalen-Hougaard model (AH) with $\hat{\alpha} > 1$. For the AH model we also distinguish between values of $\hat{\alpha} < 1$ by light green color and $\hat{\alpha} > 1$ by dark green background color.

For Canadian cohort life tables, both males and females, the results of the fit are irregular. Gamma-Gompertz model dominates for ages from 66 to 71 while for ages from 79 to 82 the best fit provided either by Kannisto or Aalen-Hougaard model with $\hat{\alpha} < 1$. For all other ages (65, 72–78, above 82) the best fitting model is the AI model of mortality. Such irregular pattern is likely due to inaccuracies in age reporting of Canadian data on death. As it is shown in the following section age misreporting can introduce a serious bias into the parameter estimates making results of fit to such data less credible. To obtain more credible results the model should be fitted to the good quality data. This is the reason why we computed the same cohort life tables for Swedish population: Swedish population data are considered to be of exceptional quality and they are used frequently in demographic research for illustration and modeling purposes.

Fit to the Swedish cohort life tables resulted in more stable results. For males, ages 65 through 67, the best fitting models are the models with defective frailty distributions: AI and AH with $\hat{\alpha} > 1$. Therefore, the frailty interpretation is inappropriate for explaining mortality pattern for the entire age range. For ages 68 and over the best fitting model is of a logistic type: gamma-Gompertz up to age 70 and Kannisto model for the higher ages. In both cases frailty distribution is gamma. For females the best fitting model appears to be the Aalen-Hougaard model with $\alpha < 1$ for all ages except 65 where the GG model provides the best fit. This is the only life table where set of the best fitting models do not include any model with defective survivorship for any starting age within the analyzed age interval. For ages 65–85 the average value of estimate of $\hat{\alpha}$ is 0.62. As this value is closer to 1/2 than to one the form of frailty distribution is closer to inverse Gaussian distribution than to gamma. The principle difference between fit to males and females for ages above 68 is that for males the best fitting models are of the logistic type: initially death rates are increasing exponentially but over age the rate of increase falls to zero and death rates reach a plateau. For females the rate of increase does not fall entirely to zero and death rates continue to increase without reaching any constant mortality level.

As quality of Canadian data is improving over time the death rates in the latest period might reflect the true pattern more closely than death rates in the cohort life tables based on deficient data. This observation was main incentive for including recent period data

in the analysis. Moreover, period life tables might manifest emerging patterns of death rates so their analysis might prove useful for better understanding of underlying demographic processes—indeed if mortality conditions are unchanged there will be no difference between period and cohort life table for people born in the same year. On the other hand one should be aware that period life tables are not based on survival experience of the real cohorts while model (12) was developed for cohort data.

For the period data analyzed here the overall fit is dominated by the AI model: the model with defective survival function and Poisson distribution of frailty. It has been selected as the best fitting model for the entire age range for Swedish males, for both female populations up to 76, and for Canadian males for 65–71 and 84–85. Gamma-Gompertz model has been selected as the best fitting one for Canadian males for starting age within interval 72–83, and Kannisto model provided the best fit for both female populations above 78.

In sum, Table 1 conveys several important messages. First, Gompertz model has been never selected as the best fitting model. This observation entirely supports earlier findings that this model is inappropriate for modeling mortality at adult and old ages. Second, there is no single model that dominates as the best fitting one. Results depend highly on population, starting age, and whether we are fitting to period or cohort data. Third, the models selected for period and cohort data are different implying that age-specific patterns of mortality are principally different within this modeling framework. Therefore, inference made on fitting frailty models to period data is not readily applicable to cohort mortality. Finally, the models with defective survival functions (AI, AH with $\hat{\alpha} > 1$) turned out to perform equally well as the models with regular survivorships (GG, KN, AH with $\hat{\alpha} < 1$) providing better fit in about 48% of all cases. If we consider both period and cohort datasets the AI model with discrete Poisson distribution of frailty emerges as a clear winner among all models. The only notable exception from this pattern is Swedish cohort data: almost all fits resulted in the best fitting models with regular survival functions.

## Effect of Age Misreporting and Fit to the Cohort Life Tables of Canada

Frailty models have been developed for explaining age-specific patterns of death rates observed in life tables for national populations. On applying the models to such data one should always keep in mind that underlying data are not necessary completely accurate. The

Canadian life tables analyzed here are based, for example, on about 2.63 millions deaths above age 65 and 1.22 millions above age 80. It would be naïve to assume that age reporting on death certificates and on census returns both providing basis for estimates of death rates is completely accurate. Therefore, it is important to understand what kind of bias is introduced in the estimates by deficiencies commonly found in demographic data.

Figure 2 shows observed age-specific probabilities of dying in the Canadian cohort life tables. As highlighted by dots in this figure $q_x$ at ages ending in 5 or 10 are either higher or equal to $q_{x+1}$ at the next age: $q_{65} > q_{66}$, $q_{70} > q_{71}$ etc. Such irregularities are attributed to age heaping in the data on deaths. If age is not known accurately there is a propensity to report age at convenient numbers ending in 5 or 10. How parameter estimates of the AH model will be biased if the model is fitted to such data?

Figure 3a shows fit of the Aalen-Hougaard model and its special case, gamma-Gompertz model, to female life table starting at age 81. Visually, gamma-Gompertz model approximates observed death rates better, especially at ages 100 and over. Nevertheless, this impression is misleading. As we can see from this figure the death rate at age 81 is substantially lowered by age heaping thus producing striking but artificial increase in mortality from age 81 to 82. Because of large number of deaths reported at age 81 such increase is captured well by the AH model consequently providing significantly better fit to the observed data than gamma-Gompertz model.

Formally, we can analyze reduction in deviance ($-2\ln L$, see also Equ. 18) achieved by the AH model as compared to the GG model. As the GG model is nested into the AH model it cannot provide a better fit than the AH model and the reduction in deviance is always positive. Indeed as shown in Fig. 3 the reduction in deviance by fitting the AH model is 165.74 which is highly statistically significant as suggested by the LR test with one degree of freedom. Fit of a more general model does not necessarily provide better approximation of observed death rates in a uniform way—some death rates might be approximated better and some of them worse than by a less general model. This statistical test does not provide any information where the fit was improved and where it deteriorated. An answer to this question can be provided by exploring distribution of difference in deviance by age as shown in Fig. 2, panel b.

Positive values of this bar chart show the ages where the fit was improved by the AH model and the negative values shows the ages where it was worsen as compared to the GG model. Sum of all bar values over age is equal to the total deviance between two models:

165.74. Clearly, as follows from Fig. 2, the greatest contribution is provided by age 81: more than 120. Contribution is also large for other ages affected by age heaping such as 86 and 91. This figure illustrates that the AH model provides better fit due to better approximation of death rates at ages with age heaping. Moreover, death rates that are corrupted by age heaping appear to have dramatic influence at parameter estimates. As shown in the footnote of Fig. 2 estimate of parameter *a* of Gompertz baseline drops from *2.46E-5* for the GG model to *6.94E-17* for the AH model, parameters *b* and $\delta$ are inflated by factor of 4: from 0.1 to 0.44 and from 0.13 to 4.62, and estimate of $\alpha$ drops to 0.84 from 1 (recall that $\alpha$ equal to 1 corresponds to gamma distribution of frailty).

Obviously, if data suffer from age heaping, which is easily detectable visually, other forms of age misreporting are also likely to be present in the data. Age at death, for example, might be either exaggerated or understated on average. Alternatively, it might be simply misstated by some random amount. Empirical observations show that at older ages death rates are generally understated if age misreporting is present in the data. It is often thought that age exaggeration affects older ages more because elderly people usually takes proud in reporting higher ages. Overstating true ages on the census returns results in inflation of the tails of population distributions. Data on deaths are not self reported so they are thought to be less affected by age misreporting. As a result death rates are biased downwards because of the inflated denominators.

Even if there is no deliberate age exaggeration and death rates are computed entirely from vital statistics data—e.g. by producing estimates of population at risk by extinct cohort method as here—the general bias is still downwards if the data are affected by age misreporting. Suppose that equal proportions of deaths are reallocated into two neighboring ages from a certain age. In absolute terms impact on deaths at lower age will be significantly less than that at higher age because of the very steep decline of the tail of the death distribution (Preston, Elo and Stewart; 1999). Resulting inflation of the tail of the death distribution leads to downward bias in the raw death rates.

To gain more insights into effect of the age misreporting on the observed death rates and on the parameter estimates of frailty models we conducted the following simulation study. Assume that the reported age at death $\widetilde{X}$ is sum of two random variables: true age at death $X$ and amount of displacement or misreporting *Y*. Further assume that age given at a single death certificate can be misreported with probability *p(x)*. Age that is reported at death certificate is then:

$$\widetilde{X} = \begin{cases} X + Y & with \quad p(x) \\ X & with \quad 1 - p(x) \end{cases}$$

By selecting particular distributions for $X, Y$ and $p(x)$ effect of various patterns of age misreporting on parameter estimates can be studied.

In the first simulation study we conducted (Table 2) the true mortality follows Gompertz model, *Y* is normally distributed with mean zero and $\sigma$=3 and probability of misreporting *p(x)* increases logistically from 1% a age 0 to 99% at age 35 (for the selected parameter values it corresponds to ages from 65 to 100 on human age scale, see notes to Table 2). This simulation scheme implies that population is homogenous and there is no deliberate age exaggeration or understatement—age misreporting is completely random as $E(Y) = 0$.

Plots of true and simulated death rates, and fit of Gompertz, gamma-Gompertz and Aalen-Hougaard mortality models are presented in Fig. 4. The parameter estimates are included in Table 2. As follows from Fig. 4. the simulated death rates deviate downwards at older ages as compared with the true values. On fitting Gompertz model to the simulated data parameter $\hat{b}$ falls to 0.08 from the true value of 0.09 and parameter $\hat{a}$ increases to 0.0315 from 0.0290 in order to capture such deviation. The GG and AH model provide significantly better fit as they are suited to capture such patterns of death rates. All parameters associated with frailty distributions are statistically significant and of a sizeable magnitude (Table 2). As expected estimate $\hat{\alpha} = 0.45$ of the AH model implies a regular survival function because the simulated death rates are not declining at higher ages.

In the second simulation study the true mortality follows Kannisto distribution. Population is assumed to be heterogeneous and the true death rates approache an asymptote equal to unity over age. Parameters describing age misreporting are the same as in the first simulation; also as before there is only random but not deliberate age exaggeration. In this case (Fig. 4) the simulated death rates not only lower than the true values but they are also declining after age 9. On fitting Kannisto model to the simulated data parameter $\hat{a}$ is overestimated (0.0304 vs. 0.0282) and parameter $\hat{b}$ is underestimated (0.48 vs. 0.50). Gamma-Gompertz model captures the original values of $a, b$ quite well but $\hat{\delta}$=0.62 is higher than $\hat{b}$=0.52 implying that the fitted mortality curve tends to an asymptote equal to 0.84 instead of one as for the true death rates. Therefore, age misreporting has an effect of lowering asymptotic mortality level of the logistic curve by inflating parameter $\delta$ responsible for amount of population heterogeneity. Further improvements in fitting the data

are obtained by Aalen-Hougaard model which captures decline in death rates at high ages. As expected estimate of $\hat{\alpha}$ =1.18 is higher than one implying that the survival function is defective. Similar to the first simulation all parameter estimates are statistically significant.

The simulation studies conducted here demonstrate that the parameter estimates describing frailty distribution ($\alpha$ , $\delta$ ; model 12) are inflated if the model is fitted to mortality data affected by age misreporting. In severe cases of age misreporting death rates might even decline at advanced ages. If the AH model captures such decline the estimated value of parameter $\alpha$ —type of frailty distribution—will be greater than one implying defective survival function. As the cohort life tables for Canada are apparently affected by age misreporting the parameter estimates included in Table 1 should be taken cautiously because of likely inflation of their values.

## Lack of Fit to Female Period Data

Despite that the frailty models have been developed modeling survival experience of the cohort data they are frequently applied to the period life tables. Mortality patterns observed in the period data closely resemble mortality patterns in the genuine cohorts so it is tempting to expect that the models perform equally well both for the cohort and period data. Analysis of the most recent data is even more intriguing as the latest life tables reveal record high life expectancy levels attributed to the exceptionally low levels of death rates, especially those for female populations of the developed countries. Observed mortality patterns in low mortality populations might manifest new patterns of human survival brining up a fascinating question whether the existing models are capable to capture and explain the emerging mortality patterns.

To illuminate this perplexing question we conducted a thorough analysis of goodness of fit of Aalen-Hougaard model to two female period life tables of Canada and Sweden used in our analysis. Firstly, we plotted observed and fitted age-specific probabilities of dying: on a first glance the model fits data quite well (Fig. 5a). Secondly, we examined ratio of fitted and observed death rates (Fig. 5b). If a model fits data well one would except to observe the ratio bouncing up and down around one in random way. The pattern in Fig. 5b is clearly not random: the fitted death rates are lower than observed up to age 71, then higher up to age 85, then again lower up to 95, and higher afterwards, both for Sweden and Canada.

Yet another way to look at the goodness of fit is to plot distribution of deviance between fitted model and saturated models (Fig. 5c). A saturated model is a model that fits

observed data exactly thus providing the lowest boundary of deviance for any model that could be applied to the data. By maximizing the log-likelihood function (18) we obtain parameter estimates that minimize deviance of the fitted model. Contribution of a single age to the total deviance is interplay between two factors: closeness of approximation of observed death rate at that age by mortality model and number of deaths recorded at that age interval. If the fitted death rate is almost the same as the observed this age provides virtually no contribution to the deviance. If the fitted and observed death rates are quite different and if the number of deaths is large the contribution will be substantial—ages close to 65, 75 and 90, Fig. 5c. For a good fitting model one's expectation is to observed more or less flat pattern in Fig. 5c. Obviously, the pattern of deviance distribution—similar to Fig. 5b—is not random thus providing additional evidence that the model does not fit the data in an adequate way.

Choice of a particular model that provides the best approximation to the female period mortality clearly depends on the starting age of life table (Table 1). If the starting age is less than 77 the best fit is provided by the AI model—the model that asymptotically predicts the steepest declines in the force of mortality. On the other hand if the starting age is higher than 77 the best approximation is that of Kannisto model—the model of a logistic type that asymptotically predicts a constant level of mortality. Such dramatic shift from one model to another of a fairly different nature is a manifestation of structural changes in the age-specific pattern of death rates. In other words, if we consider the entire age interval the pattern of mortality is radically different as if we consider only ages above 77.

Further insights into the pattern of female period mortality can be provided by examining age-specific rates of increase of $q_x$: $k_x = \ln q_{x+1}/q_x$ (Fig. 6). Direct examination of $k_x$—especially at ages with small number of deaths—is hampered by high degree of stochastic noise. To reveal the underlying pattern we also produced smoothed curves of this quantity by applying cubic smoothing spline procedure to the raw estimates if $k_x$ (de Boor, 2001). Finally, we have added series of age-specific rates of increase computed from age-specific probabilities of dying, $q_x$ predicted by the fitted AI model. The fitted $q_x$ are exactly the same as those given in Fig. 5; the observed $q_x$ have been computed for an extended age interval starting with age 35 for illustration findings discussed below.

As it follows from Fig. 6 the observed $k_x$ is approximately constant from age 40 to 65—this is what one would expect if mortality follows Gompertz model. From 65 to 80 it

increases sharply, then reaches maximum and finally declines. On the other hand, the *r(x)* of the fitted model declines uniformly for the entire age interval. Therefore, the fitted model does not capture the pattern persisting in the real data.

Failure to capture such patterns is inherent to this model because Equ. (12) describes selection in heterogeneous population. If death rates are initially increasing then due to process of selection the rate of increase decelerates as compared with homogeneous population because the frail individuals die out and the share of robust individuals increases. Formally, by examining second derivative of $\mu(x)$ (Equ. 12) it can be shown that the derivative can be either positive or negative depending on parameter values, and its sign does not change with age. Typically for human populations, the second derivative will be negative corresponding to the case of deceleration in mortality increase (Fig 1a). As the derivative is negative no acceleration in mortality increase is possible and there is no flexibility built in the model to produce the bell-shaped pattern observed in the real data (Fig. 6). The age-specific schedule of female period life tables is therefore more complicated as those implied by the Aalen-Hougaard model and any of its special cases.

## Conclusions

Investigation of age-specific mortality in Canada on a cohort basis has been made possible by establishment of Canadian Human Mortality Database. Such studies of human mortality are possible only for a limited number of countries in the world with long-standing tradition of collecting demographic data. Canadian database is a very welcome addition to this collection.

For modeling purposes we employed a four-parameter frailty model with Gompertz baseline mortality. Different from other traditional frailty models the model incorporates more general class of frailty distributions described by a single parameter $\alpha$. Consequently, the model produces more flexible set of mortality curves for the total population and permits making formal inference about the type of frailty distribution by the standard statistical techniques. By estimating parameter $\alpha$ one can test, for example, whether frailty is gamma distributed or whether another distribution of frailty is more appropriate for the life table.

Application of this model to Canadian cohort data suggests that Gompertz model is inappropriate for approximation Canadian mortality at adult and advanced ages due to lack of ability to capture deceleration in the rate of mortality increase. Canadian pattern of age-specific mortality appears to be similar to patterns found in other populations and the frailty

models with their ability to capture such patterns provide significantly better approximation to the observed death rates than Gompertz. No firm conclusions could be made, however, about a specific form of frailty distribution. In the earlier years Canadian data on deaths are affected by age misreporting which introduces a bias in parameter estimates. Generally, age misreporting inflates parameter estimates related to frailty distributions, both type and variance, as suggested by the simulation studies conducted here. Moreover, the estimates appear to be quite sensitive to the age heaping if it takes place at ages with large number of deaths. Revealing pattern of true mortality in the data affected by age misreporting requires introduction of additional assumptions about age misreporting though feasible but it has not been undertaken here.

Fitting Aalen-Hougaard frailty model to Swedish cohort data and to recent period life tables both of Canada and Sweden had a motivation to apply the model to good quality data. As before the application has not returned a superior frailty model as well. For Swedish male data, for example, the best fitting model (ages > 70) appeared to be the two-parameter Kannisto model while for females it was four-parameter Aalen-Hougaard with estimate of $\hat{\alpha}$ close to 0.5 and frailty distribution close to inverse Gaussian, consequently. The principle difference between two models is that Kannisto model is of a logistic type with force of mortality approaching a constant level over age while the Aalen-Hougaard model with $\alpha < 1$ predicts indefinitely increasing death rates without approaching any plateau. The period life tables for Canada were best approximated by gamma-Gompertz model if starting age of the life table is between 72 and 83, and by the special case of Aalen-Hougaard model with $\alpha \to \infty$ if outside of this age range. The AI model emerged also the best fitting model for Swedish males for the entire age range, and for both female populations up to age 76. Above 76 the period female death rates follows Kannisto model.

Emergence of the AI model as the best fitting model in so many cases came as a surprise because this model is characterized by discrete Poisson distribution of frailty, declining death rates at advanced ages, and defective survival function—survivorship that reaches a positive limit instead of zero. In mortality research this limit has interpretation initial proportion of individuals with zero frailty or proportion of immortal individuals. Such interpretation cannot be accepted as a valid one but it does not mean that we should limit our parameter space only by values of $\alpha$ corresponding to the regular survival functions. Strictly speaking, frailty models based on notion of unobserved heterogeneity are not identifiable from univariate life table data because one can always argue that population is homogeneous with mortality for each individual to be given by Equ. (12). This weakens interpretation

power of such models. An estimate of $\hat{\alpha}$ greater than one do lead to rejection of hypothesis that population is heterogeneous with Gompertz baseline and with frailty distribution belonging to Hougaard's family. Further insights into mortality pattern can be obtained by analyzing age-specific rates of mortality increase and comparing them with those implied by the fitted model. In case of female period data it was found that the pattern of mortality is more intricate that it is implied by the general Aalen-Hougaard model. In other words there is a lack of fit of this model to the data. Instead of decelerating as one would expect if selection process takes place in the population the rate of mortality increase *accelerates* from age 65 to 80. This is both incompatible with Gompertz distribution and with Aalen-Hougaard model.

A long-standing paradigm is that human mortality increases approximately on a constant rate from age 35 to about 80 where the rate of increase drops down and death rates eventually reach a plateau at the highest ages. Generally, the results in this work do not support this standpoint—analysis of cohort Swedish female life table suggests that death rates are better approximated by a model without plateau, and analysis of female period life table reveals rapid acceleration in death rates after 65. As death rates in these life tables among the lowest recorded in human populations, the observed patterns might be a manifestation of a new emerging pattern of human survival which would require reconsideration of the prevailing viewpoints. Direct testing of this hypothesis can be made once mortality levels in cohort life tables reach the levels of the current period life tables.

## Acknowledgments

The main part of this project was carried out at Queen's University, Kingston, Canada. This research was supported by grants from the U.S. National Institute on Aging (R01 AG18444 and R01 AG11552).

## Disclaimer

The views expressed in the paper do not imply the expression of any opinion on the part of the United Nations Secretariat.

## References

1. Aalen OO. Heterogeneity in Survival Analysis. Statistics in Medicine 1988;

7(11):1121-37.

2. Aalen OO. Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. The Annals of Applied Probability 1992; 2(4):951-72

3. Beard RE. Note on Some Mathematical Mortality Models. G.E.W. Wolstenholme and M. O'Connor (eds.). The Lifespan of Animals. Ciba Foundation Colloquium on Ageing. Little, Brown, Boston . 1959

4. Canadian Human Mortality Database. Available online at http://www.bdlc.umontreal.ca/chmd/index.htm: University of Montreal, 2005

5. de Boor C. A practical guide to splines. Revised Edition, Applied Mathematical Sciences Vol 27. New York: Springer, 2001

6. Gompertz B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Philosophical Transactions of the Royal Society of London 1825; 115 :513-83

7. Hougaard P. Survival Models for Heterogeneous Populations Derived from Stable Distributions. Biometrika 1986; 73(2):387-96 (Correction vol 75 395)

8. Human Mortality Database. Available online at www.mortality.org: University of California, Berkeley, United States and Max Planck Institute for Demographic Research, Germany, 2004

9. Perks W. On Some Experiments in the Graduation of Mortality Statistics. Journal of the Institute of Actuaries, 1932; 63:12-40

10. Preston SH, Elo IT, Stewart Q. Effects of age misreporting on mortality estimates at older ages. Population Studies 1999; (53):165-77

11. Thatcher RA. The Long-Term Pattern of Adult Mortality and the Highest Attained Age. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1999; 162(1):1

12. Thatcher RA, Kannisto V, Vaupel JW. The Force of Mortality at Ages 80 to 120. Odense, Denmark: Odense University Press, 1998

13. Thatcher RA, Kannisto V, Andreev KF. The Survivor Ratio Method for Estimating Numbers at High Ages. Demographic Research 2002; 6(1)

14. Vaupel JW, Manton KG, Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography 1979; 16(3):439-54

15. Vaupel JW, Yashin AI. The Deviant Dynamics of Death in Heterogeneous Populations. Sociological Methodology 1985; 15:179-211

Table 1 The Best Fitting Model by Starting Age of Life Tables

| | Cohort life tables | | | | Period Life Tables | | | |
|---|---|---|---|---|---|---|---|---|
| | Males | | Females | | Males | | Females | |
| Age\Country | Canada | Sweden | Canada | Sweden | Canada | Sweden | Canada | Sweden |
| 65 | AI | AI | AI | GG | AI | AI | AI | AI |
| 66 | GG | AH | GG | AH | AI | AI | AI | AI |
| 67 | AH | AH | GG | AH | AI | AI | AI | AI |
| 68 | AI | GG | GG | AH | AI | AI | AI | AI |
| 69 | GG | GG | GG | AH | AI | AI | AI | AI |
| 70 | GG | GG | GG | AH | AI | AI | AI | AI |
| 71 | KN | KN | GG | AH | AI | AI | AI | AI |
| 72 | AI | KN | AI | AH | GG | AI | AI | AI |
| 73 | AI | KN | AI | AH | GG | AI | AI | AI |
| 74 | AI | KN | AI | AH | GG | AI | AI | AI |
| 75 | AI | KN | AI | AH | GG | AI | AI | AI |
| 76 | AI | KN | AI | AH | GG | AI | AI | AI |
| 77 | AI | KN | AI | AH | GG | AI | AI | KN |
| 78 | AI | KN | AI | AH | GG | AI | GG | KN |
| 79 | KN | KN | KN | AH | GG | AI | KN | KN |
| 80 | AH | KN | AH | AH | GG | AI | KN | KN |
| 81 | AH | KN | AH | AH | GG | AI | KN | KN |
| 82 | KN | KN | AH | AH | GG | AI | KN | KN |
| 83 | AI | KN | KN | AH | GG | AI | KN | KN |
| 84 | AI | KN | AI | AH | AI | AI | KN | KN |
| 85 | AI | KN | AI | AH | AI | AI | KN | GG |

AH—four parameter Aalen-Hougaard model
GG—gamma-Gompertz model
KN—Kannisto model
AI—Aalen-Hougaard model with $\alpha \to \infty$
The bright green background of the AH model with corresponds to the case of parameter estimate $\hat{\alpha} > 1$ which implies defective survival function and the pale green background to the case of $\hat{\alpha} \le 1$.

Table 2 Fit of Frailty Models to the Data with Simulated Age Misreporting

| Simulation 1 | | | | | |
|---|---|---|---|---|---|
| Age | $a$ | $b$ | $\alpha$ | $\delta$ | -2ln(L) |
| Sim. | 0.0290 | 0.09 | | | |
| Sim. Est. | 0.0315 | 0.08 | | | 676,962 |
| GG | 0.0295 | 0.10 | 1.00*) | 0.11 | 676,810 |
| AH | 0.0286 | 0.12 | 0.45 | 0.48 | 676,786 |
| Simulation 2 | | | | | |
| Sim. | 0.0282 | 0.50 | 1.00 | 0.50 | |
| Sim. Est. | 0.0304 | 0.48 | 1.00*) | 0.48**) | 469,328 |
| GG | 0.0280 | 0.52 | 1.00*) | 0.62 | 469,138 |
| AH | 0.0293 | 0.49 | 1.18 | 0.48 | 469,065 |

*) parameter was fixed at this level
**) parameter was fixed equal to *b*

Sim. — parameters of distribution of true age at death, *X*
Sim. Est. — fit of the simulated model to the data with age misreporting
GG — fit of gamma-Gompertz model to the data with age misreporting
AH — fit of Aalen-Hougaard model to the data with age misreporting

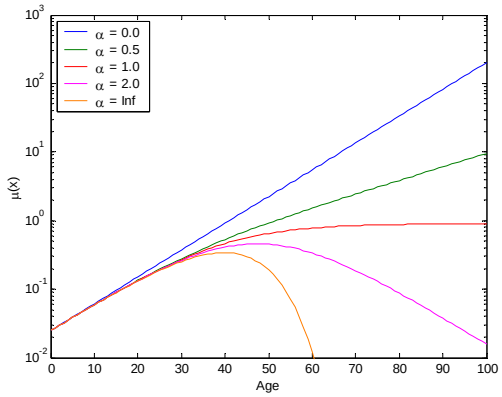
All parameter estimates of Aalen-Hougaard and gamma-Gompertz models are significant at 0.001 level.
Notes on simulation:

- Simulated sample size N = 100,000;
- For Simulation 1 the true age at death *X* is distributed according to the Gompertz model. Parameters of this model have been selected approximately to be equal to the parameters estimated from the fit to the ages 65 and over. The simulated age range 0–35 corresponds to ages 65–100 on human age scale;
- For Simulation 2 the true age at death $X$ is distributed according to Kannisto model of mortality. The mortality curve is logistic with asymptote equal to one. The large value of parameter *b* was selected for illustration purposes.
- For all simulations probability of misreporting is logistic $p(x) = \left[1 + \exp(4.5951 - 0.2626x)\right]^{-1}$. The value of $p(0) = 0.01$ and $p(35) = 0.99$ corresponding to 1% of age misreporting at age 0 and 99% at age 35
- The age misreporting variable *Y* is normally distributed with mean 0 and $\sigma = 3$. If the reported age at death $\widetilde{X} = X + Y$ is negative during the simulations it was replaced with the true age at death. This adjustment affects only a few cases as initial probability of misreporting only 1%.

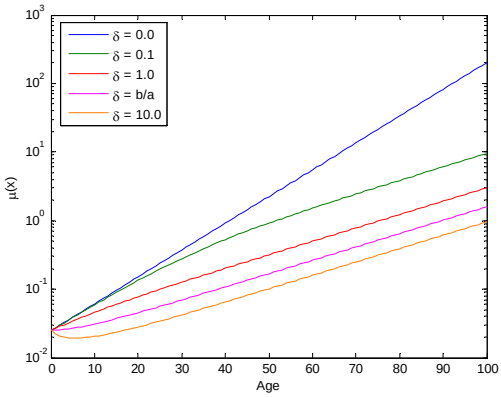
Figure 1. Typical mortality curves of Aalen- Hougaard frailty model with Gompertz baseline, Equ. (12).

Gompertz parameters are the same for all graphs and equal to $a$=0.025, $b$=0.09

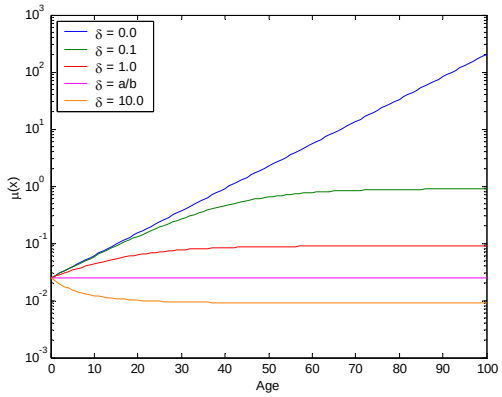
a) $\delta$ =0.1




b) $\alpha$ =0.5




c) $\alpha$ =1




d) $\alpha$ =2

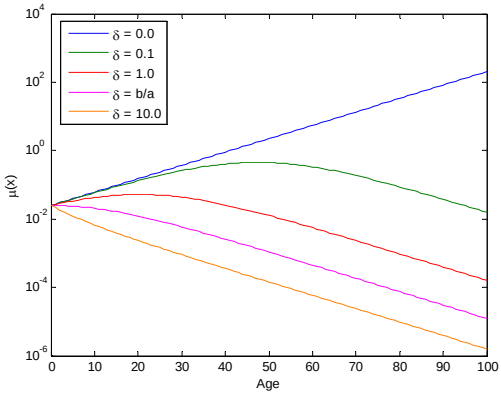


Figure 2. Age Heaping in Canadian Cohort Life Tables

Cohorts 1856–1893 for males and 1856–1889 for females.

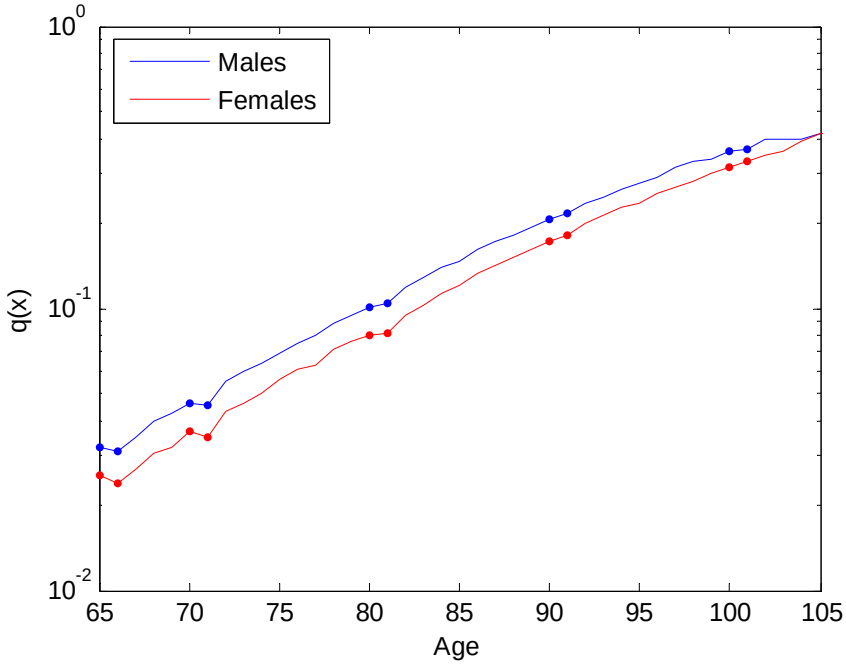
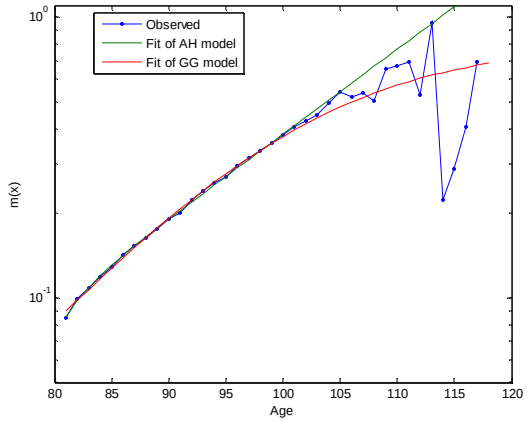
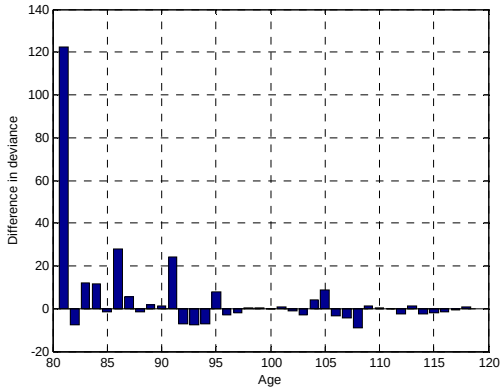


Figure 3 Capture of Age Heaping by Aalen-Hougaard Model

a) Fit of Aalen-Hougaard and gamma-Gompertz models to Canadian female life table, cohorts 1856–1889




b) Deviance difference by age between Aalen-Hougaard and gamma-Gompertz models




Parameter estimates of Aalen-Hougaard Model:

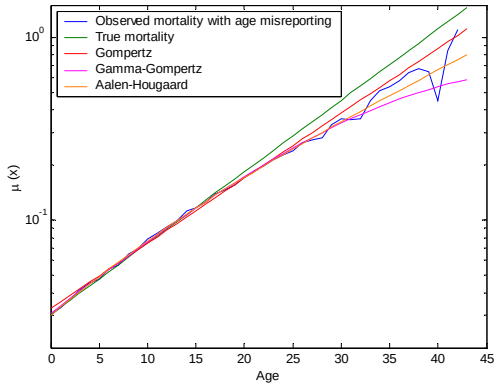
$\hat{a}$ = 6.94E-17; $\hat{b} = 0.44$; $\hat{\alpha} = 0.84$; $\hat{\delta} = 4.62$; -2LnL = 3,166,107

Parameter estimates of gamma-Gompertz Model:

$\hat{a}$ = 2.46E-05; $\hat{b}$ = 0.10; $\hat{\delta}$ = 0.13; -2LnL = 3,166,272

Figure 4 Fit of Mortality Models to the Data with Simulated Age Misreporting

Simulation 1. True mortality is Gompertz model




Simulation 2. True mortality is Kannisto model

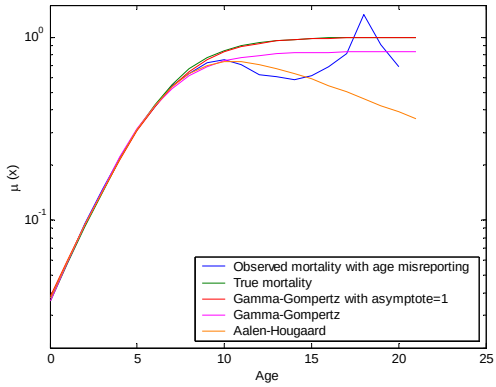
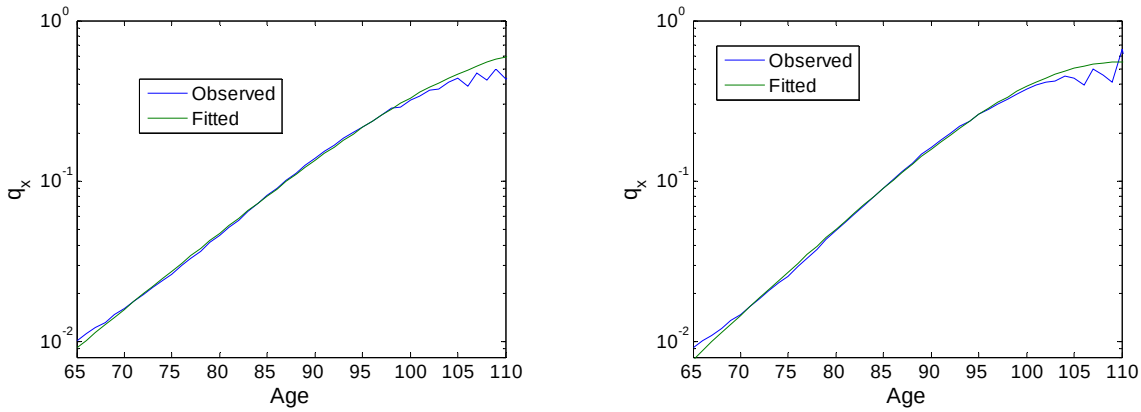


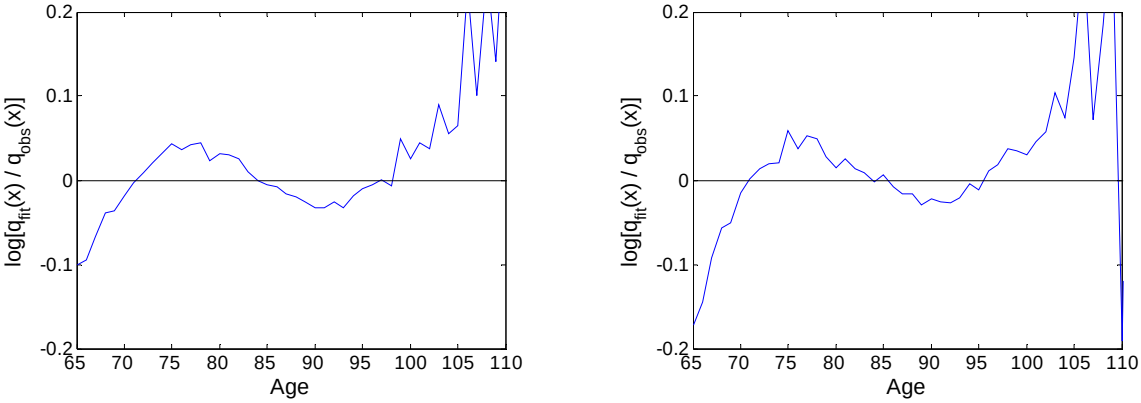
Figure 5. Goodness of Fit of the AI Model to the Period Female Life Tables, Years 1990–2001, Ages 65 and over

Canada Sweden

a) Age-specific probabilities of dying




b) Ratio of fitted to observed age-specific probabilities of dying




c) Distribution of difference of deviance by age between the AI and the saturated model

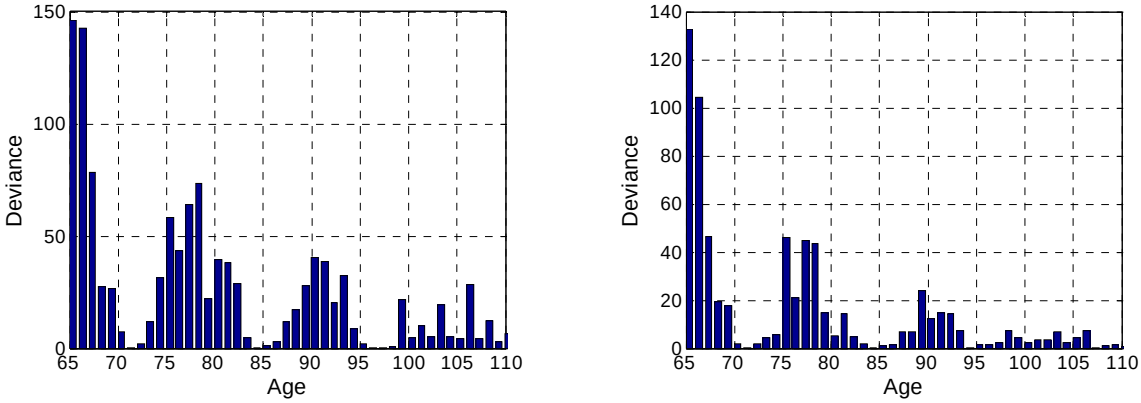
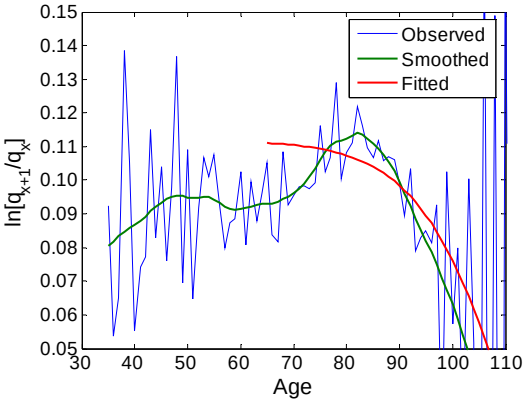
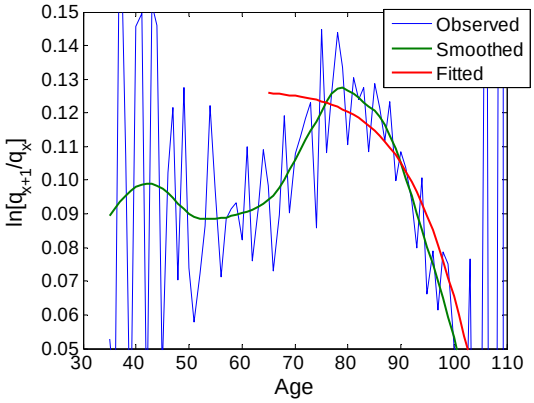


Figure 6 Age-Specific Rate of Mortality Increase in Female Period Life Tables, Years 1990–2001

a) Canada




b) Sweden




The AI model has been fitted to ages to ages 65 and over.