Understanding human mortality and demographic heterogeneity through a two-process vitality model

Ting Li[1]

Quantitative Ecology and Resource Management

University of Washington

Box 352182

Seattle, WA 98195

James J. Anderson[2]

School of Aquatic & Fishery Sciences and

Quantitative Ecology and Resource Management

University of Washington

Box 358218

Seattle, WA 98195

1. E-mail: ltkitty@u.washington.edu

2. E-mail: jjand@u.washington.edu

Abstract

We examine population mortality through two interactive stochastic processes: an intrinsic process defining the survival capacity (vitality) of an organism declines stochastically to a zero-boundary representing intrinsic death and an extrinsic mortality that occurs when external challenges exceed the vitality. With the framework, the model fits mortality data for the entire human life span. With the construction of two processes as well as the heterogeneity structure, the model is capable of explaining the perplexing patterns of mortality parameters that emerge in the Strehler and Mildvan (SM) theory. The deviation of survival patterns from SM correlation suggest that human beings have enter a new period where environment conditions have stabilized and improvements in longevity are the result a decrease in the intrinsic rate of senesces.

1. Introduction

For many years, understanding the mysterious causes of death as well as the mechanism under the senescence process has attracted scientists in almost every field to dedicate themselves into it. However, since the death process is so complicated that an organism may die in many ways, it is extremely difficult to find an appropriate but relatively simple way to describe it. Thus, most scholars tend to model population survival in terms of the instantaneous rate of mortality ever since Gompertz (1825) first proposes that mortality rates increase exponentially with age. Although he suggests at the same time that an unspecified force might destroy the material of organization necessary for life, the connection between the force and mortality rates is so vague that makes the concept of an instantaneous mortality rate elusive and biologically tenuous (Li and Anderson 2009). Most importantly, the focus on the mortality event itself implicitly disregards the fact that, as Aalen and Gjessing (2001) note, apart from pure accidents, mortality events do not happen out of the blue, they are the endpoint of some process that develops with age. Therefore, it seems logical and potentially enlightening to consider mortality more in a context of biological mechanisms.

A prominent attempt to explain the Gompertz model through a biological theory is made by Strehler and Mildvan (1960). It is known as the famous Strehler-Mildvan (SM) general theory of mortality and aging. They define the term "vitality" as the capacity of an individual organism to stay alive which linearly declines at each age. Death occurs when the Maxwell-Boltzmann distributed challenge magnitudes exceed the vitality level. Although this is a successful illumination to link the instantaneous mortality rate with biologically traceable kinetics, there are still several important issues the theory fails to

handle. Firstly, because the vitality framework created by SM theory is done in a deterministic format, i.e. there is no heterogeneity allowed in a population, it makes assessing how the variability among individuals might affect the population survival impossible. Secondly, rooted on the Gompertz law, the SM theory lacks power to distinguish the different processes they are trying to unveil with limited parameter space. Thus, artificial constraints must be put on the coefficients to estimate other underlining parameters e.g. the attrition coefficient *B*. Finally, the famous negative relationship between Gompertz parameters found by SM has been challenged with a significant sign of instability (Yashin et al. 2001, 2002a).

An alternative approach, sometimes called first passage or Markov mortality models, describes the stochastic rate of loss of survival capacity, vitality, to a killing boundary at zero vitality. This concept was first proposed half a century ago (Sacher 1956), developed by Anderson (Anderson 1992, 2000; Anderson et al. 2008), Weitz and Frazer (2001), Aalen and Gessing (2001) and has recently been advanced by Li and Anderson (2009) called the vitality model. The model characterizes the stochastic rate of vitality loss by a Wiener process in which the time to death is determined by the first passage time of vitality to the zero boundary. Because of the stochastic property, the model inherently incorporates heterogeneity among a population in terms of the vitality. The Li-Anderson model even specifies the initial and evolving variation of a population separately and explored how they might shape the population's survival curves. However, under the framework of the vitality model, the exterior killing other than the boundary killing is modeled as a constant which is too simple particularly for capturing human mortality patterns, since human beings usually experience a much more complex life style. And

intuitively, the chance of surviving from an external stress may highly depend on the internal biological system of an organism, which is the basic concept of the SM theory.

Therefore, the idea of further developing the SM theory based on the stochastic vitality process emerges naturally. Nevertheless, it will not be simply a stochastic version of the SM theory but a multi- process view to understand and quantify the aging process. In this work, we formulate a two-process model: the intrinsic vitality process and the extrinsic challenge process, which work simultaneously to shape the population survival. On the other hand, mortality rate can be partitioned into the intrinsic absorption rate and the extrinsic killing rate as well as two sources of heterogeneities. Through the lens of partition, the model provides a unique and informative perspective into the aging process. It allows analysis and comparison on mortality patterns based on the intrinsic and extrinsic death processes and also the variation structures. Most of all, we will show that this new two-process model well characterizes the patterns of human mortality which the Gompertz-type models fail to capture because of their inadequate framework.

2. The two-process model

To make the process model biologically meaningful and valuable, the definition of each process is better to be associated with the classification of mortality as what scholars usually consider. Although there is no consensus about the partition criterion, it is generally agreed to divide mortality into the senescence relevant part indentified as "actual death" and the accidental part indentified as "avoidable death" (McGlnnis and Foege 1993). Following Carnes and Olshansky's (1997) terms, we refer the former one as the intrinsic mortality and the latter one as the extrinsic mortality.

2.1 Intrinsic process

In the terminology of Carnes et al. (2006), intrinsic mortality is described as arising from the inside of an organism. However, it is still vague that how intrinsic death occurs. In our consideration, intrinsic mortality is the consequence of senescence that it is the end point of an organism's gradual loss of survival capacity. Thus, the intrinsic mortality should reflect the internal aging process. According to this definition, we construct the intrinsic mortality process on the framework of the vitality survival model initially published in a paper by Anderson (1992) and further expanded by Li and Anderson (2009). The vitality model for survival summaries varied mechanisms by a single quantity called "vitality" which denotes the remaining survival capacity of an organism that is similar to the definition in SM theory. Each individual begins with an initial vitality, v_0, and intrinsic death occurs when its vitality reaches zero (Fig. 1). The random trajectory of vitality, *v*, between v_0 and 0 is described by the Wiener process:

$$dv / dt = -\rho + \sigma\varepsilon_t \tag{1}$$

where ρ is the mean value of the rate of vitality loss, σ is the magnitude of the stochastic component and ε_t is a white noise process that spreads the distribution. The two parameters ρ and σ are constant at both individual and population level. That means a population shares common parameters that have been averaged across each individual's lifespan. To derive the mortality due to the loss of vitality we require the initial vitality distribution $p(v_0)$ and the conditional probability distribution of its first passage time to the zero-boundary $f(t \mid v_0)$. By definition, the fraction of total population that has not died from intrinsic causes at time *t*, is equivalent to the probability that the individual's

vitality has not reached zero by time *t*. Since the cumulative density function of the first arrival time $F(t)$ (failure rate) gives the absorption probability, the vitality based survival rate $l_v(t)$ (not dying from intrinsic process) is expressed as

$$\begin{aligned} l_v(t) &= 1 - F(t) \\ &= 1 - \int_0^t f(t)dt \\ &= 1 - \int_0^t \int_0^\infty f(t \mid v_0)\, p(v_0) dv_0 dt \end{aligned} \tag{2}$$

The Weiner processes first passage time is given by the inverse Gaussian distribution as (Cox and Miller 1965)

$$f(t \mid v_0) = \frac{v_0}{\sigma\sqrt{2\pi}} t^{-2/3} \exp\left(-\frac{(v_0 - \rho t)^2}{2\sigma^2 t} \right) \tag{3}$$

We represent the initial vitality distribution with a Gaussian distribution with mean μ and variance τ^2 :

$$p(v_0) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(v_0 - \mu)^2}{2\tau^2} \right) \tag{4}$$

Using Eqs. (3) and (4), we approximate the marginal distribution of first arrival time with the integration

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} f(t \mid v_0)\, p(v_0)\, dv_0 = \frac{\tau^2\rho + \sigma^2\mu}{\sqrt{2\pi\left(\tau^2 + \sigma^2 t\right)^3}} \exp\left(-\frac{(\mu - \rho t)^2}{2\left(\tau^2 + \sigma^2 t\right)} \right) \\ &= \frac{u^2 r + s^2 r}{\sqrt{2\pi\left(u^2 + s^2 t\right)^3}} \exp\left(-\frac{(1 - rt)^2}{2\left(u^2 + s^2 t\right)} \right) \end{aligned} \tag{5}$$

where

$$r = \rho / \overline{v}_0 \tag{6}$$

is the normalized mean rate of vitality loss, or drift rate,

$$s = \sigma / \overline{v}_0 \tag{7}$$

is the normalized variability in the rate of loss of vitality, or spread rate, and

$$u = \tau / \overline{v}_0 \tag{8}$$

is the coefficient of variation of the initial vitality distribution. Note that in Eq.(5), we integrate over the range $(-\infty, \infty)$ which violates the allowable range of vitality $v \geq 0$, but in most situation, it does not become a problem when $\mu / \tau > 2$ or $u < 0.5$. Then the survival associated with intrinsic process becomes

$$\begin{aligned} l_i(t) &= 1 - \int_0^t f(t)\,dt = \Phi\left(\frac{\mu - \rho t}{\sqrt{\tau^2 + \sigma^2 t}}\right) - \exp\left(\frac{2\tau^2\rho^2}{\sigma^4} + \frac{2\mu\rho}{\sigma^2}\right)\Phi\left(-\frac{\frac{2\tau^2\rho}{\sigma^2} + \mu + \rho t}{\sqrt{\tau^2 + \sigma^2 t}}\right) \\ &= \Phi\left(\frac{1 - rt}{\sqrt{u^2 + s^2 t}}\right) - \exp\left(\frac{2u^2 r^2}{s^4} + \frac{2r}{s^2}\right)\Phi\left(-\frac{\frac{2u^2 r}{s^2} + 1 + rt}{\sqrt{u^2 + s^2 t}}\right) \end{aligned} \tag{9}$$

where Φ is a cumulative normal distribution. Therefore, the intrinsic mortality rate is expressed as

$$\mu_i(t) = \frac{f(t)}{l_i(t)} \tag{10}$$

The conditional vitality distribution evolving with time can be formulated analytically (Li and Anderson 2009)

$$f_v(v_t \mid v_0) = \frac{p_v(v_t \mid v_0)}{\int_0^\infty p_v(v_t \mid v_0)dv} = \frac{p_v(v_t \mid v_0)}{l_i(t)} \tag{11}$$

where $l_i(t)$ is eq. (9) the intrinsic survival rate at time *t* and

$p_v(v_t \mid v_0) = \frac{1}{\sqrt{2\pi(\tau^2 + t\sigma^2)}} \exp\left(-\frac{(v_t - v_0 + \rho t)^2}{2(\tau^2 + t\sigma^2)}\right)\left[1 - \exp\left(-\frac{2v_t v_0(\rho\tau^2 + \sigma^2)}{\sigma^2(\tau^2 + t\sigma^2)}\right)\right]$ is the absolute density function of the vitality which evolves from a Gaussian distribution into a quasistationary gamma-like distribution that is finally absorbed into the zero-vitality boundary (Li and Anderson, 2009).

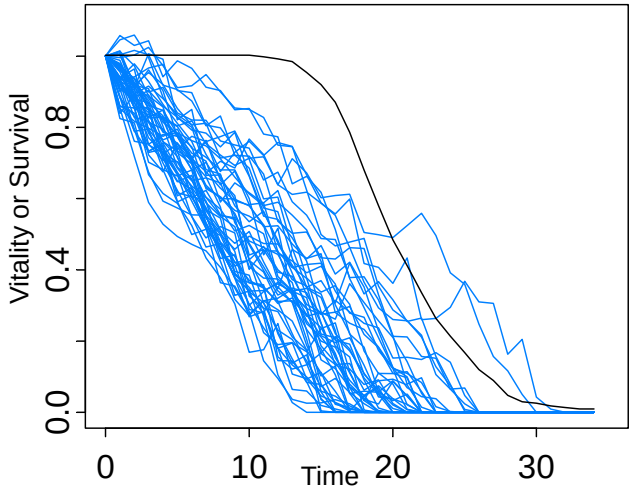


Figure1. Depicts individual vitality trajectories (eq.(1)) and survival (eq.(9)) with $\bar{v}_0$ = 1 and τ = 0.

2.2 Extrinsic process

Comparing to intrinsic mortality, extrinsic mortality is usually considered as death that is "relatively preventable and treatable", including "mortality mainly from infections and accidents." (Shryock and Siegel 1975). Some scholars, such as Makeham (1867) have suggested using the age-independence as the criterion to distinguish extrinsic mortality

from intrinsic mortality. In an early version of the vitality model (Anderson et al. 2008; Li and Anderson 2009), extrinsic mortality was independent with the vitality process thus as a constant with age. Although it is difficult to determine theoretically whether extrinsic mortality exhibits age-dependent properties, there is evidence that the extrinsic process may be associated with age especially for human beings. For instance, the partition conducted by Carnes et. al. (2006) demonstrates an increasing trend in extrinsic mortality with age for the U.S. 1996 mortality data according to their criterion. Assuming a constant extrinsic mortality rate, the early version vitality model doesn't fit the human data as well as it does to other animals' like fish, insects and mammals (Li and Anderson 2009). However, by adding a time-dependent term to the extrinsic part, the goodness of fit is significantly improved. This might be explained in terms of the longer life spans and more complex lifestyles of humans compared to animals necessitating expressing time dependency in the extrinsic elements.

In our framework, let $x_t, t \geq 0$ be a random point process with rate λ to represent the occurrence of instantaneous extrinsic challenges (stresses) such as a natural disaster or a contact with infectious diseases. Then λ measures the frequency of challenges. For each extrinsic event, a magnitude variable y_t with a cumulative distribution function $\varphi(y)$ denotes the intensity of the challenge. We assume that only when the magnitude y_t exceeds the current vitality level v_t, that the extrinsic challenge results in death, i.e. death occurs when $\Pr(Y_t > v_t)$. This assumption couples the risk of death from external forces to the intrinsic vitality level of the individual. For instance, twenty and eighty year old men may experience the same accident, such as falling down the stairs, but the outcomes

are likely to be very different. Note that the model also implies that the external challenges will not change the vitality trajectories unless extrinsic death happens. Thus, based on the model two conditions are necessary to develop extrinsic mortality: the occurrence of the challenge (stress) and the magnitude of the challenge surpassing the intrinsic resistance described by vitality. If we assume a specific case that x_t is a Poisson process, the point process becomes history-independent and the extrinsic mortality rate for each individual can be expressed as

$$m(t)dt = \Pr(Y_t > v_t)\lambda dt = (1-\varphi(v_t))\lambda dt \tag{12}$$

If we further assume that the magnitude of the event is exponentially distributed such that most external events are small and the probability of large events declines relative to their magnitude, then the cumulative distribution function is $\varphi(y) = 1 - e^{-y/D}$ where D is the scale parameter. Now the rate of extrinsic mortality for an individual with absolute vitality v_t

$$m(t)dt = (1-(1-e^{-v_t/D}))\lambda dt = \lambda e^{-v_t/D}dt = \lambda e^{-(\overline{v}_0/D)\frac{v_t}{\overline{v}_0}} = \lambda e^{-v_t^*/\beta} \tag{13}$$

The vitality is normalized to its initial mean, $v_t^* = v_t / \overline{v}_0$, so $\beta = D/\overline{v}_0$ becomes an parameter indicating the normalized challenge intensity. It characterizes the environmental deleteriousness in a sense that a larger β implies that high intensity challenges occur more frequently. It is also worth noting that eq. (13) implies the susceptibility to extrinsic death is dependent on individual's survival capacity and β severs as the scale, which expresses our intuition that the effect of stressors varies with age and vitality. However, we can derive it from a more biologically robust process.

The aggregated extrinsic mortality rates at a population level is

$$\mu_e(t) = \int_0^\infty \lambda e^{-v_t/D} f'(v_t \mid \overline{v}_0) dv = \int_0^\infty \lambda e^{-v_t^*/\beta} f'(v_t^* \mid \overline{v}_0^* = 1) dv^* \quad (14)$$

where $f'(v_t^* \mid \overline{v}_0^* = 1)$ is the conditional vitality distribution at time *t* when starting with a normal distribution $N(1, u^2)$ at *t*=0. Noted it is different from eq.(11). Because the extrinsic process unequally removes individuals from the population, it changes the original vitality distribution. In addition, the change of vitality distribution will also influence the intrinsic mortality by altering the absorbing probability. Here we finish the theoretical definition of the extrinsic death process, which could be considered as a stochastic version of the SM theory incorporating population heterogeneity.

2.3 Complete model

The new vitality model assumes two sources of death: an extrinsic killing and an intrinsic killing expressed as a boundary absorption. Thus, the total mortality rate is the sum of eq. (10) and (14) :

$$\mu(t) = \mu_i(t) + \mu_e(t) \quad (15)$$

and the corresponding survival function equals $l(t) = l_i(t) l_e(t)$ with five parameters *r*, *s*, *u*, λ and β. The intrinsic process is characterized by the vitality parameters r, s and u, and the extrinsic process is characterized by a challenge frequency parameter λ and a challenge intensity parameter β. Both extrinsic parameters are determined by the environment conditions. Note *r*, *s*, *u* and β are normalized parameters and are modified by the initial vitality while λ is independent of the initial vitality level. In contrast to

Gompertz-type model, the parameters are all process-based and completely specify both the population vitality and mortality trajectories.

3. Fitting to human mortality data

A good fit to real mortality data is essential for evaluating any model and here we compare the two-process vitality model with the classical models. The Gompertz model characterizes mortality rate by an exponential function. However, this simple structure cannot capture the early-age "hook" and old-age "plateau" in mortality patterns, which respectively refer to the anomalously high infant death rate and leveling off or decline in the death rate at old age (Carey et al. 1992; Vaupel 2004; Vaupel et al. 1998). Neither the Gompertz nor the SM models fit these anomalies in the mortality trajectory discrepancies and so the classical models are typically used to fit human mortality between the ages of 35 and 85. These discrepancies are well known, but because the classical models fit the majority of the age structure that the models are considered good enough. In contrast, the vitality model is able to describe the entire survival curve in a biologically realistic framework.

The vitality model spontaneously captures the old age plateaus due to the construction of intrinsic vitality as Markov process. Early as 2001, Weitz and Evans (2001) explained mortality plateaus as a generic consequence of considering death in terms of first passage times for processes undergoing a random walk with drifts, which was equivalent to Anderson's vitality model. Li and Anderson (2009) even quantified the observation of plateaus under the vitality parameters. As stated by Steinsaltz and Evans (2004, 2007), the convergence to a mortality plateau—is, in fact, the natural property of Markove-

process models convergence to a quasistationary distribution that is the time the shape of the probability mass is stable, and the level of distribution sinks proportionately at every location. In other word, despite the sink of total population, the absorption rate conditioned on the survival population is a constant. Although explicit solutions would be difficult, Steinsaltz and Evans (2004) proved a general case of Brownian motions including both extrinsic killing and boundary killing to approach a quasistationary distribution.

3.1 Specifying the challenge frequency term λ to account early-age "hook"

It is usually believed λ partially measures environment deleteriousness by specifying the challenge frequency from the environment. However, Yashin et. al. (2001, 2002a) suggested that this term might actually reflect the interaction between environment and organism's self-protection system by representing the frequency of challenges that arrive at the internal system after passing through the biological defense. In other words, suppose the total challenge frequency initialed by the environment equals λ_0 but only a proportion of 1-*p* are able to get through the protection system that $\lambda = (1-p)\lambda_0$ where *p* characterizes the efficiency of defensive mechanisms.

Having an independent parameter λ in the vitality model increases the flexibility in characterizing the effective frequency of stressful events, i.e. an age-dependent effect of λ. For human beings, experiencing the same environment, infants with a smaller *p* would be more likely to be threatened from external challenges, as their biological protection systems are still developing whereas teenagers are apt to raise the challenge frequency directly from environment λ_0 and thus increase λ due to their immature behaviors.

Despite such distinct causations, it is reasonable to assume a time-depend challenge frequency term when considering entire survival curve. We express this pattern through an exponential form $\lambda(t) = K + ce^{-dt}$ where K represents the base or minimum frequency of challenges without age effects; c indicates the maximal frequency added to the base rate at age 0 and d measures how quickly the early high frequency degenerates with time. An example is shown in Appendix A.

More elaborated structure can be established for the frequency term to capture complex changes in stresses with aging. Thus, including a flexible λ has more benefits than just fitting the lifetime survival data of human beings, but may help understand age-variant effects of both environment stresses and body defensive systems on specific subpopulations, e.g. smoking group.

3.3 Model approximation and fit

Eq. (14) gives the expected extrinsic mortality rate at age *t*. However, there is no closed form for $f'(v_t^* \mid \overline{v}_0^* = 1)$ making it difficult to quantify both extrinsic and intrinsic mortality. Therefore, we resort to approximation.

Suppose each individual in the population at age *t* has the same rate of extrinsic mortality $\tilde{\mu}_e(t) = \lambda(t)e^{-\overline{v}_t^*/\beta}$, where $\overline{v}_t$ is the mean vitality at age *t*. In effect to characterize the extrinsic mortality, we assume each individual of a given age has the same vitality. By doing this, the extrinsic process essentially ignores individual heterogeneity in the population. With this assumption the extrinsic mortality does not affect the rate of change

of the intrinsic mortality, i.e. it does not alter the vitality distribution in eq.(11). From Anderson et al. (2008) we can further approximate $\overline{v}_t^*$ as a linear function of time *t*:

$$\overline{v}_t^* \approx 1 - rt \tag{16}$$

Eq.(16) holds when the variance term *s* is small. Under a normalized model, the initial mean vitality equals 1 and then vitality declines almost linearly with time at a rate of *r* if the variation of the population is small implied by the Wiener process. For human mortality data, *s* is usually less than 0.01which guarantees eq.(16) is a good approximation. There are two advantages to use eq.(16). Firstly, using age term instead of the average vitality simplifies the equation, and thus, estimation algorithms converge quickly to a solution. Secondly, since most classical models tend to express mortality rates as a function of age, it provides a way to interpret the traditional results through a view of the vitality process. The averaged extrinsic mortality rate becomes $\hat{\mu}_e(t) = \lambda(t)e^{-(1-rt)/\beta}$ and the exact intrinsic rate remains as in eq.(10). The total approximated mortality rate is expressed as

$$\mu(t) = \mu_i(t) + \hat{\mu}_e(t) = \mu_i(t) + \lambda e^{-(1-rt)/\beta} \tag{17}$$

The eq. (17) has 5 basic parameters *r*, *s*, *u,* λ and β. Adding a time-variant challenge frequency term $\lambda(t) = K + ce^{-dt}$ expands the model to 7 parameters: *r*, *s*, *u, K,* β, *c* and *d*. The parameter estimation problem is cast as a maximal likelihood optimization as developed by Salinger et al. (2003). Simulations are conducted to assess the approximation. The approximation underestimates *r* while overestimates λ, but a correction can be applied to both parameters according to the strong relationships

between the true parameters and estimated parameters. The detailed method of parameter estimation, the procedure of simulation and correction are all listed in Appendix B and C.

Figure 4(A) demonstrates the vitality model fitting to period survival data of Swedish females at year 1890 and 2006 separately (data source: Human mortality database (Wilmoth and Shkolnikov 2010)). It yields good fitting for entire survival curves including the early-age "hook". Even restricting mortality data between ages of 30 and 100, the vitality model still outperforms both Gompertz (1825) and Makeham-Gompertz model (1867) indicated by Figure 4(B), which depicts an example of mortality rate in log form against age.

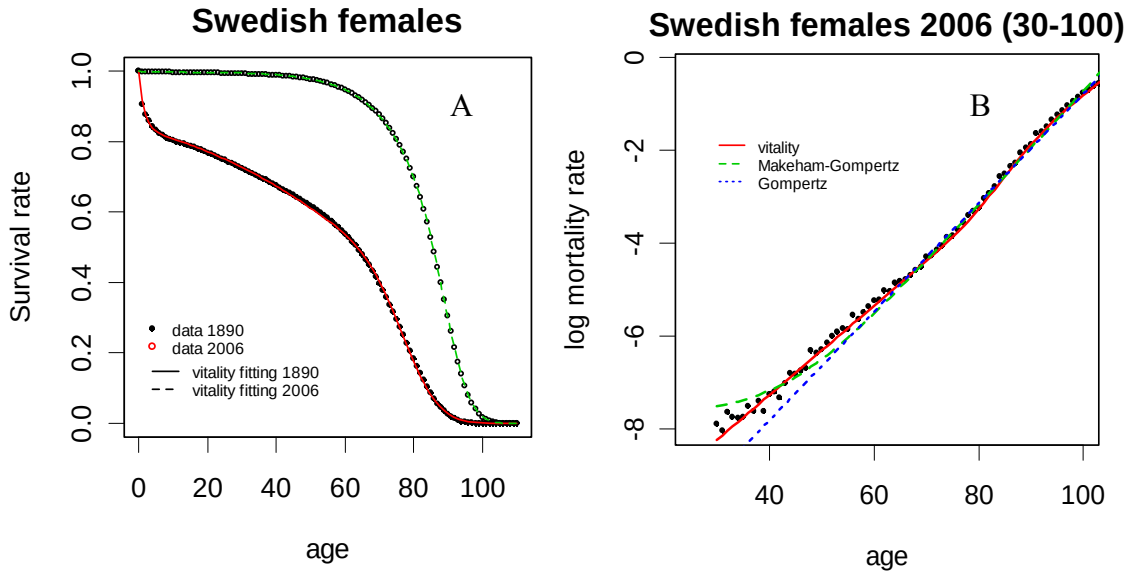


Figure 4: (A). The vitality model fits to period survival data of Swedish females at year 1890 and 2006. Dots and circles indicate real survival data for year 1890 and 2006 respectively. The two lines demonstrate survival curves generated from the vitality model.

(B). The vitality, Gompertz and Makeham-Gompertz model fit to mortality rate in log transform from period data of Swedish females at year 2006.

Though fitting to mortality data is not the ultimate goal of the model, it does provide support for the utility and biological basis of the model.

4. New explanations to survival patterns in human data

We suggest the vitality model has value for understanding aging process and patterns of survival. The model also provides a reinterpretation of patterns of parameters found in other models, in particular from the SM theory, since the two have similar structures.

An important theoretical finding of SM theory is the negative relationship between Gompertz coefficients *a* and *b* (the Gompertz law $\mu(t) = ae^{bt}$) which is expressed in the Gompertz notation as $\ln a = \ln K - b/B$. We can reexpess this in terms of the vitality notation as

$$\ln a = \ln \lambda - \frac{b}{r} \tag{18}$$

where *K* is our challenge frequency λ and *B* is the vitality loss rate coefficient r. The environmental deleteriousness $D_\varepsilon = B/b = r/b$ represented by the SM theory also shares a common meaning with our challenge intensity term β. These analogues result because the extrinsic process in the vitality model is similar to that in the SM theory. However, there are two fundamental differences between the two: 1) the stochastic structure of the vitality model allows heterogeneity in mortality within a population; 2) the killing process demonstrated in SM theory accounts for all the death whereas the similar extrinsic process in vitality model is only responsible for part of the death. We believe

these distinctions make the vitality model incrementally more realistic that the SM theory and thus provided a better explanation and fit of survival data.

4.1 Explaining the paradox in estimating fraction of vitality loss (B) in SM theory

Because of this expanded model framework, the vitality model can directly estimate all the parameters while the SM theory can only estimate parameter r ,λ and β after some assumptions. The vitality loss rate r is a focal term because $1/r$ is a measure of the expected life span or the age of expected zero vitality (Zheng et al., 2010). Without direct estimation in the SM theory r can be estimated with two methods. The first one is to set λ = 1, such that eq.(18) becomes $\ln a = -b / r$ and r can be derived from $-b / \ln a$ (Strehler and Midvan, 1960). Zheng et al. (2010) recently calculate r from 42 countries following this approach. Perplexingly, the analysis revealed that Central American and South-East Asian countries had lower r, i.e., higher age of expected zero vitality, than most developed countries. This begs the question, do presumably harsher environmental conditions of these countries produce advantages in survival, or is the assumption on λ unrealistic?

Rearranging eq.(18) $-\frac{\ln a}{b} = \frac{1}{r} - \frac{\ln \lambda}{b}$ and combining $\hat{r} = -b / \ln a$ from the restricted estimation, yields

$$\frac{1}{\hat{r}} = \frac{1}{r} - \frac{\ln \lambda}{b} \tag{19}$$

The difference between the estimated value of age at zero vitality and the true value depends on the ratio of the challenge frequency term $\ln\lambda$ to the Gompertz coefficient b. Simulations were conducted to access whether the bias exists or not. Survival curves

were generated according to the vitality model death process, because the vitality structure is flexible enough to assign any values to each of the 5 parameters. The method for simulating survival trajectories is described in Appendix B. All the parameter values chosen for simulation were within a reasonable range for human mortality. Then Gompertz model ($\mu(t) = ae^{bt}$) was fit the simulated survival curves giving $1/\hat{r}$ from $-\frac{\ln a}{b}$. The results are summarized in Figure 5.

The plot depicts the estimated age at zero vitality against the challenge intensity term β with fixed challenge frequency λ. A pattern emerges where the estimated age of zero vitality ($1/\hat{r}$) tends to have a larger value for populations experiencing higher challenge intensity, where in actuality they have the same age of zero vitality ($1/r$), which counters the intuition. Thus, it is likely to be the case that the Central American and South-East Asian countries do not have lower senescence rates, *r*, relative to developed countries. We therefore conclude that the actual difference is due to a higher intensity in environmental stresses, β. That is $1/\hat{r}$ is more exaggerated for the developing countries compared to developed countries. Although we could not exclude possibilities that there were truly genetic and physical advantages in these Central American and South-East Asian countries, the vitality framework provides a more plausible explanation that the estimated longer life expectance may be a mathematic misrepresentation because of unrealistic restriction on parameters in SM theory.

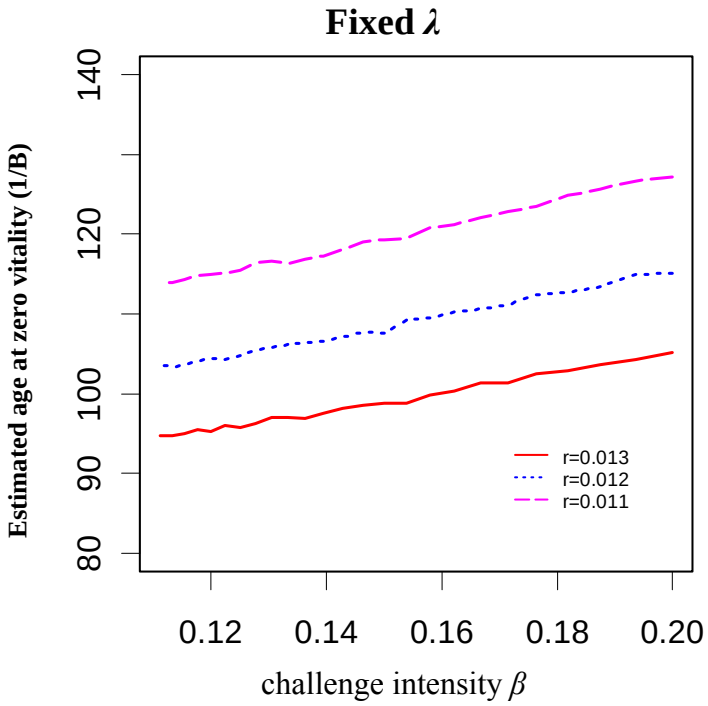


Figure 5: Estimated age at zero vitality against challenge intensity under fixed challenge frequency $\lambda = 0.12$. Each line represents a single true vitality loss rate that was used for generating the survival curves. All curves have the same background variance structure in vitality: $s = 0.001$ and $u = 0.08$.

4.2 Explaining SM correlation patterns

The other method for estimating the vitality loss rate r from SM theory also relies on the regular relationship of Gompertz coefficients expressed by eq.(18). To estimate, r from the Gompertz coefficients requires a linear relationship between $\ln a$ and b, in that the slope of the relationship is $1/r$ according to eq.(18). Thus, r actually represents the averaged vitality loss rate over a series of longitudinal survival curves constructed from a sequential series of years. Thus, the success of the method depends on a stable linear relationship, which would imply that eq.(18) is valid. Early studies confirm a stable

pattern for adult mortality from the year 1900 to 1986 in the US (Riggs 1990) and other developed countries, including overall mortality trends in industrialized countries (Prieto, Llorca and Delgado-Rodriguez 1996; Riggs and Millecchia 1992). However, in recent period and cohort mortality data the stable linear relationship in the pattern is not evident in countries like France, Japan, Sweden and the US (Yashin et al. 2001, 2002a; Yashin, Iachine and Begun 2000; Yashin et al. 2002b). To be specific, for period data, "hooks" emerge in the $\ln a$ vs. b relationship for those countries in the second half of the 20th century. The approximately constant negative slopes (1859-1960) reverse sign and flatten for France, Sweden and the US starting around 1960 and for Japan round 1980. For cohort data, only Sweden exhibits a linear relationship over the length of data. The other countries have complex patterns in which the slope changes sign multiple times over the years of data. The important point here is that the SM theory assumptions break down for the years where the curves change slope and flatten (Figure 6).

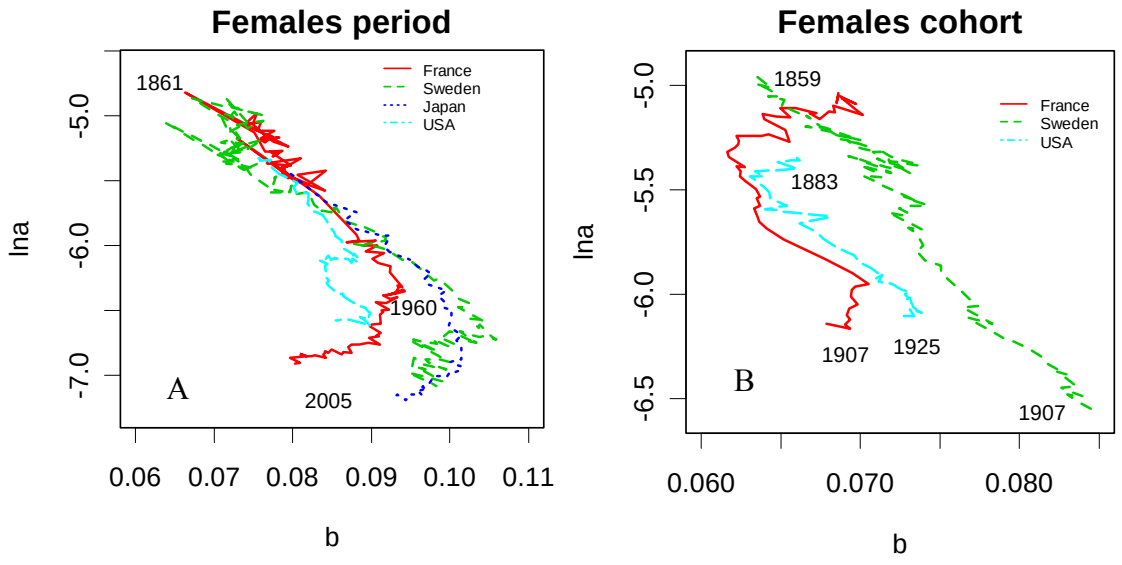


Figure 6: (A) period patterns of SM correlation for females in France (1861~2005), Sweden (1861~2005), Japan (1950~2000) and the US (1938~2005); (B) cohort patterns of SM correlation for females in France (1859~1917), Sweden (1821~1915) and the US (1883~1927). All mortality data are from the human mortality data base (Wilmoth and Shkolnikov 2010) and we use mortality for ages between 40 and 80.

The unstable patterns shake the root of the SM theory. As suggested by Yashin et al. (2001), new concepts need to be developed. To consider the two-process vitality as a replacement, it is important that the vitality framework is able to explain the irregular patterns in SM coefficients. We again use the vitality model to generate survival curves with the 5 process parameters and fit the Gompertz model to get SM coefficients. The estimates of $\ln a$ and b on are shown in Figure 7A under a fixed challenge frequency λ and Figure 7B under a fixed challenge intensity β. In both plots, points on a single line from upper left to lower right represent pairs of ($\ln a$, b) estimated from simulated survival curves in which background r is fixed and λ and β change. A decline in either λ (plot A) or β (plot B) moves the points further down (i.e. upper left to lower right). Each horizontal line denotes points of equal λ (plot A) or β (plot B) in real process, but varies in senescence rate r. As r decreases, the points move to the left.

We can now use the patterns in Figure 7 to explain the patterns in Figure 6. A stable negative linear pattern of the SM coefficients (Figure 6) can be interpreted in terms of changes in the environmental parameters λ or β dominating changes in r. Correspondingly, a reversal and flattening in the $\ln a$ vs b curve reflects the intrinsic senesces parameter dominating the environmental parameters λ or β. Thus we surmise, that over the first half of the 20th century, the improvement in the developed countries

period survival was the result of improvement in the environment, either by reducing the challenge frequency or intensity or both. The "hook" then represents a gradual shift to the dominance of intrinsic improvements, i.e. *r*, over environmental improvements. For cohort data, our model suggests the environment improved steadily for Swedish females but in the US and France changes in longevity involved both environmental and intrinsic factors. Differences in the cohort and period results reflect differences in how the intrinsic and environmental factors change over time but disentangling these issues is difficult and beyond the scope of our paper.

Theoretically, the failure of SM theory to explain the correlation patterns is because it attributes all deaths to one killing process. Although it tries to connect death with an intrinsic process, artificial restrictions on parameters are required. Thus, it lacks power to resolve the differences between improvement in true senescence and improvement in the environment. From this aspect, including two death processes is of great necessity.

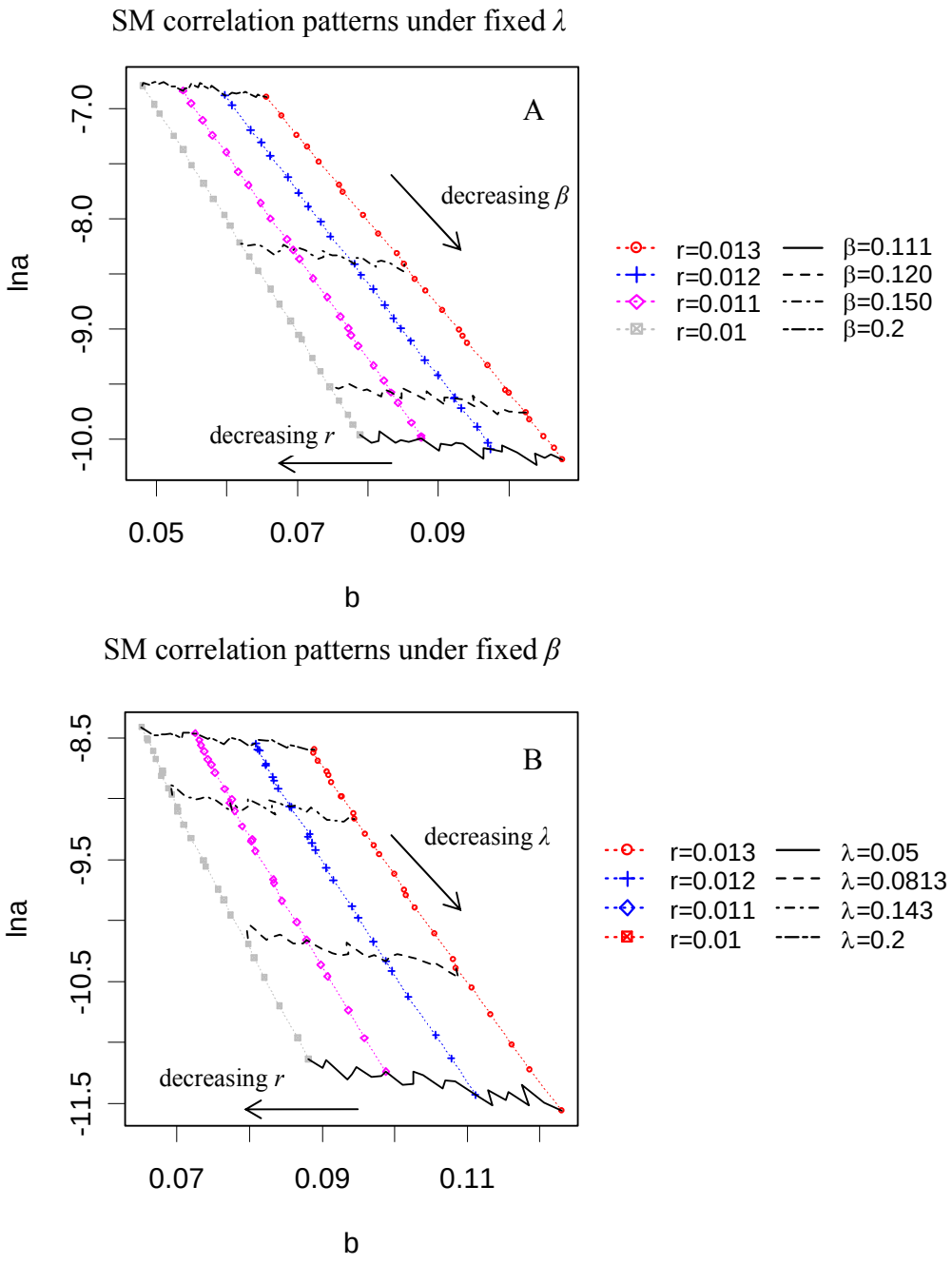


Figure 7: Simulated SM correlation patterns: (A) Survival curves are all simulated under fixed challenge frequency term $\lambda = 0.12$; (B) Survival curves are all simulated under fixed challenge intensity term $\beta = 0.125$. For both (A) and (B), curves all have the same background variance structure in vitality: $s = 0.001$ and $u = 0.08$.

5. Conclusion

The vitality model, which describes death through the loss of vitally, or survival capacity, and an extrinsic killing process is expanded by representing the extrinsic mortality through more a biologically realistic process. The model characterizes mortality through 5 processes: the vitality loss rate indicating aging rate, the initial and evolving variation in vitality indicating population heterogeneity, and the external challenge frequency and intensity indicating environmental deleteriousness. The model fits mortality curves over the entire human lifespan from infants onward. Here we use the model to interpret a puzzling pattern often found in the relationships of parameters in the SM model and suggest a biologically plausible basis for the pattern.

The model has many advantages over the SM theory, among which the two-process structure is essential. The idea of expressing mortality in two processes is not new (Anderson 1992, 2000; Makeham 1867), but a biologically meaningful construction to both processes seldom found in the literature. Two centuries years ago, scientists began thinking about an informative way to partition and standardize mortality measures in the scientific literature (Carnes and Olshansky 1997). Although there is no consensus on the criterion of classification, it is generally agreed to divide mortality into a senescence relevant part indentified as “actual death” and an accidental part indentified as “avoidable death” (McGlnnis and Foege 1993). However, considering extrinsic mortality as a constant with age (Makeham 1867) is far away from satisfactory. Other approach using the registration of death (Carnes et al. 2006) is also problematic due to incomplete knowledge in etiology, missing information and arbitrary judgment. The vitality model

itself quantifies deaths as from two sources, which provides a natural way to partition mortality.

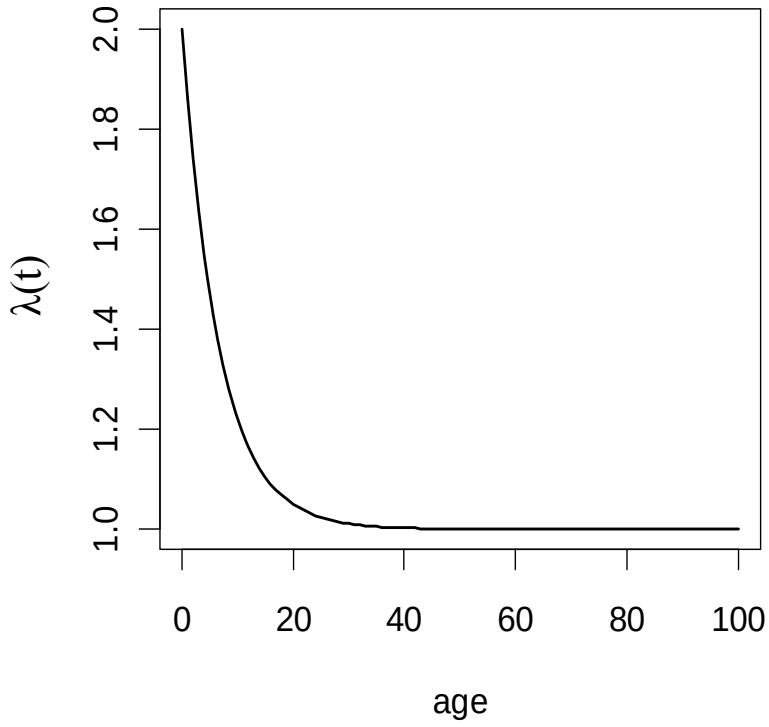
In the past 200 years, a remarkable gain of life expectancy for human beings has been observed in all countries. Average life time has increased from around 30 to 80 years. A fundamental question is whether this mortality decline indicates a slowing of the ageing process, such that people today are biologically younger than in the past (Gurven and Fenelon 2009) or it is because improvement in health care have delayed death from diseases and other stresses. Resolving this question has huge implications in both economic and health policy making. Our initial and speculative conclusion from this analysis is that as a species we are becoming biologically younger.

Reference:

Aalen, O.O.and H.K. Gjessing. 2001. "Understanding the shape of the hazard rate: a process point of view." *Statistical Science* 16(1):1-13.
Anderson, J.J. 1992. "A vitality-based stochastic model for organism survival." *Individual-based models and approaches in ecology: populations, communities and ecosystems. Chapman & Hall, New York*:256-277.
—. 2000. "A vitality-based model relating stressors and environmental properties to organism survival." *Ecological Monographs* 70(3):445-470.
Anderson, J.J., M.C. Gildea, D.W. Williams, and T. Li. 2008. "Linking growth, survival, and heterogeneity through vitality." *Am Nat* 171(1):E20-43.
Carey, J.R., P. Liedo, D. Orozco, and J.W. Vaupel. 1992. "Slowing of mortality rates at older ages in large medfly cohorts." *Science* 258(5081):457.
Carnes, B.A., L.R. Holden, S.J. Olshansky, M.T. Witten, and J.S. Siegel. 2006. "Mortality Partitions and their Relevance to Research on Senescence." *Biogerontology* 7(4):183-198.
Carnes, B.A.and S.J. Olshansky. 1997. "A biologically motivated partitioning of mortality." *Experimental Gerontology* 32(6):615-631.
Cox, D.R.and H.D. Miller. 1965. *The theory of stochastic process*: Wiley.
Gompertz, B. 1825. "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies." *Philosophical Transactions of the Royal Society of London* 115:513-583.
Gurven, M.and A. Fenelon. 2009. "Has Actuarial Aging "slowed?Over the Past 250 Years? A Comparison of Small-Scale Subsistence Populations and European Cohorts." *Evolution* 63(4):1017-1035.
Li, T.and J.J. Anderson. 2009. "The vitality model: A way to understand population survival and demographic heterogeneity." *Theoretical Population Biology*.
Makeham, W.M. 1867. "On the law of mortality." *Journal of the Institute of Actuaries* 13:325-358.
McGlnnis, J.M.and W.H. Foege. 1993. "Actual causes of death in the United States." *JAMA* 270(18):2208.
Prieto, M.D., J. Llorca, and M. Delgado-Rodriguez. 1996. "Longitudinal Gompertzian and Weibull analyses of adult mortality in Spain (Europe), 1900-1992." *Mechanisms of ageing and development* 90(1):35-51.
Riggs, J.E. 1990. "Longitudinal Gompertzian analysis of adult mortality in the US, 1900-1986." *Mech. Ageing Dev* 54:235?247.
Riggs, J.E.and R.J. Millecchia. 1992. "Using the Gompertz-Strehler model of aging and mortality to explain mortality trends in industrialized countries." *Mechanisms of ageing and development* 65(2-3):217.
Sacher, G.A. 1956. "On the statistical nature of mortality, with especial reference to chronic radiation mortality." *Radiology* 67(2):250.
Salinger, D., J. Anderson, and O. Hamel. 2003. "A parameter estimation routine for the vitality-based survival model." *Ecological Modelling* 166(3):287-294.
Shryock, H.S.and J.S. Siegel. 1975. "The Methods and Materials of Demography." *Third Printing (rev.) Washington, DC: Government Printing Office*.

Steinsaltz, D.and S.N. Evans. 2004. "Markov mortality models: implications of quasistationarity and varying initial distributions." *Theoretical Population Biology* 65(4):319-337.
—. 2007. "Quasistationary distributions for one-dimensional diffusions with killing." *Transactions-American Mathematical Society* 359(3):1285.
Strehler, B.L.and A.S. Mildvan. 1960. "General theory of mortality and aging." *Science* 132(3418):14-21.
Vaupel, J.W. 2004. "The biodemography of aging." *Population and Development Review* 30:48-62.
Vaupel, J.W., J.R. Carey, K. Christensen, T.E. Johnson, A.I. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, and P. Liedo. 1998. "Biodemographic trajectories of longevity." *Science* 280(5365):855.
Weitz, J.S.and H.B. Fraser. 2001. "Explaining mortality rate plateaus." *Proceedings of the National Academy of Sciences* 98(26):15383.
Wilmoth, J.R.and V. Shkolnikov. 2010. "The Human Mortality Database."
Yashin, A.I., A.S. Begun, S.I. Boiko, S.V. Ukraintseva, and J. Oeppen. 2001. "The new trends in survival improvement require a revision of traditional gerontological concepts." *Experimental Gerontology* 37(1):157-167.
—. 2002a. "New age patterns of survival improvement in Sweden: do they characterize changes in individual aging?" *Mechanisms of Ageing and Development* 123(6):637-647.
Yashin, A.I., I.A. Iachine, and A.S. Begun. 2000. "Mortality modeling: a review." *Mathematical Population Studies* 8(4):305-332.
Yashin, A.I., S.V. Ukraintseva, S.I. Boiko, and K.G. Arbeev. 2002b. "Individual aging and mortality rate: how are they related?" *Social Biology* 49(3/4):206-217.

Appendix A: The causation term changing with age.



Appendix B: Parameter estimation and model simulation

Since there is no analytical form for the true mortality rate of the vitality model (eq.(15)), simulation is an essential way of exploring the shape of survival curves based on the 5 process parameters. Simulations are needed to assess the model approximation in eq.(17) by comparing the estimated parameters and the true parameters that generate the survival curves. Also, we can fit other mortality models e.g. Gompertz model to the simulated survival curves. Then, the patterns of coefficients from that specific model can be explained in terms of the change in vitality loss rate, initial or evolving population heterogeneity, and the frequency and intensity of the environmental challenges.

The survival curves are simulated from the vitality process. Each population member was assumed to have a vitality of v_0 at time 0. And v_0 follows a Gaussian distribution $N(1,u^2)$. The vitality for each individual was calculated for a single time step by the following equation.

$$v_t = v_{t-1} - r_a + s_a \times W \qquad t=1, 2, 3\ldots \tag{20}$$

where W is the white noise calculated by selecting a random number from a normal distribution. From eq. (20), 10,000 vitality trajectories are generated to represent a population. At each time t, we count death from both intrinsic and extrinsic process. The intrinsic mortality happens when the individual vitality trajectory v_t drops below zero. And the probability of dying from extrinsic process in time interval (t-1,t) is $1-e^{-\lambda(e^{-\beta v_{t-1}}+e^{-\beta v_t})/2} \approx 1-e^{-\int_{t-1}^{t}\lambda e^{-\beta v_m}dm}$. A binomial random variable is generated to determine whether the individual is killed from eternal force according to the probability. When

either intrinsic or extrinsic death occurs, the vitality trajectory is excluded from further calculation. Thus, we generate a survival curve along time from the fraction of vitality trajectories left at each time point. And the mortality trajectory is derived from the survival curve.

The simulated survival curves are controlled by 5 parameters: the vitality loss rate r, the evolving variation s, the initial variation u, challenge frequency λ and challenge intensity β. We call them the background parameters when the simulated curves are fitted to other models. All simulated curves used in Figure 5 and Figure 7 have the same background s=0.001, u=0.08. They other underline parameters vary within a reasonable range for human mortality.

Appendix C: Parameter estimation and evaluation of model approximation

The model parameters r, s, u, λ, β (or the expanded version r, s, u, K, β, c and d) in the approximated mortality equation.(eq.(17)) can be estimated from survival data similarly to the method used in Li and Anderson's paper (2009). The estimation problem is casted as a maximal likelihood optimization as developed by Salinger et al. (2003) to deal with interval-censored data in which mortalities are counted at the end of each time period rather than continuously. The likelihood function is constructed from the multinomial distribution based on the proportion of deaths in each time period. We estimate standard errors by examining the estimated variance matrix. Specifically, standard errors are obtained by taking the square root of the diagonal elements in the inverse of the Hessian of the negative log-likelihood, evaluated at the parameter estimates (Kendall and Stuart, 1997).

We use the simulation approach introduced in appendix B to generate survival curves. We fix the background s=0.001 and u=0.1. Only one of the parameters r (0.01~0.015), λ (0.02~0.15) and β (0.1~1) is changed each time according to a step of 0.0005 for r, 0.005 for u, 0.005 for λ and 0.5 for β. Thus, 5049 different survival curves are generated with all the combination of r, λ and β. The approximated model (eq. (17)) is fitted to those curves to get the estimated parameters. Some results are listed in Table C1 as examples. It seems that the approximated model tends to underestimate r and overestimate λ, whereas the other estimated parameters s, u and β are relatively close to the true values (the true values are within 95% confidence intervals of the estimated parameters). However, there are strong relationships between the estimated parameters and the true ones, which can be used to correct estimation. Those relationships are summarized in

Figure C1. The proportion of the estimated *r* over the true *r* is a linear function with the estimated λ. It yields an equation $\hat{r}/r = 1 - 0.61\hat{\lambda}$ with an r-squared value 0.89. Indicated by plot B in Figure C1, $\hat{\lambda}/\lambda = 0.98 + 0.68\hat{\beta}$. But the uncertainty gets bigger as the increase of estimated β.

Table C1: Simulation results. Parameters are chosen to be close to the real values estimated from the mortality data of Swedish females (Wilmoth and Shkolnikov 2010).

	r	*s*	*u*	λ	β
actual para	0.01200	0.00100	0.10000	0.02500	0.62500
estimated para	0.01178	0.00181	0.09764	0.02622	0.60483
s.e.	0.00004	0.00136	0.00205	0.00126	0.01532
actual para	0.01200	0.00100	0.10000	0.04000	0.60000
estimated para	0.01166	0.00138	0.09855	0.04161	0.59198
s.e.	0.00005	0.00817	0.02321	0.00199	0.02291
actual para	0.01200	0.00100	0.10000	0.06000	0.40000
estimated para	0.01136	0.00145	0.10849	0.06855	0.03832
s.e.	0.00008	0.00778	0.00500	0.00415	0.01153
actual para	0.01200	0.00100	0.10000	0.08000	0.20000
estimated para	0.01095	0.00361	0.09668	0.12130	0.19187
s.e.	0.00007	0.00424	0.00532	0.00275	0.00205
actual para	0.01100	0.00100	0.10000	0.08000	0.25000
estimated para	0.01008	0.00092	0.09906	0.10250	0.24425
s.e.	0.00008	0.00266	0.00563	0.00771	0.00605
actual para	0.01100	0.00100	0.10000	0.06000	0.20000
estimated para	0.01029	0.00429	0.10308	0.08185	0.19440
s.e.	0.00012	0.00807	0.03302	0.01150	0.00709
actual para	0.01100	0.00100	0.10000	0.10000	0.15000
estimated para	0.01254	0.00337	0.09687	0.00044	0.05376
s.e.	0.00004	0.00220	0.01265	0.00004	0.00220

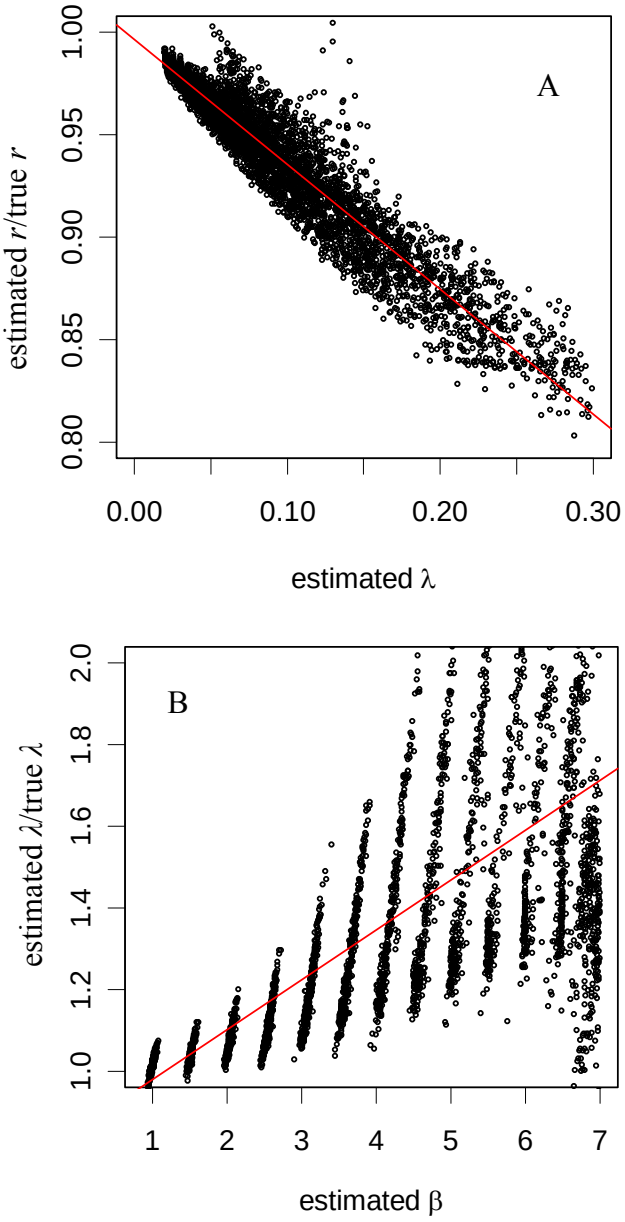


Figure C1: A: The estimated *r* over the true *r* against estimated λ. B: The estimated λ over the true λ against estimated β.