Title
Validated analysis of mortality rates demonstrates distinct genetic mechanisms that influence lifespan

Authors:
Kelvin Yen
Department of Neuroscience
Mount Sinai School of Medicine
New York, NY, USA

David Steinsaltz
Department of Statistics
University of Oxford
Oxford, UK

Charles Vernon Mobbs
Department of Neuroscience
Mount Sinai School of Medicine
New York, NY, USA

Corresponding Author and Address:
Charles Mobbs
Mount Sinai School of Medicine
One Gustave Levy Place
Box 1065
New York, NY 10029
Phone# 212-659-59299
Charles.mobbs@mssm.edu

Running Title:
Genetic effects on mortality rates

Keywords:
Gompertz
Aging
Lifespan
Senescence
Mortality

Abstract
A key goal of gerontology is to discover the factors that influence the rate of senescence, which in this context refers to the age-dependent acceleration of mortality, inversely related to the morality rate doubling time. In contrast factors that influence only initial mortality rate are thought to be less relevant to the fundamental processes of aging. To resolve these two determinants of mortality rate and lifespan, initial morality rate and rate of senescence are calculated using the Gompertz equation. Despite theoretical and empirical evidence that the Gompertz parameters are most consistently and reliably estimated by maximum likelihood techniques, and somewhat less so by nonlinear regression, many researchers continue to use linear regression on the log-transformed hazard rate. The present study compares these three methods in the analysis of several published studies. Estimates of the mortality rate parameters were then used to compare the theoretical values to the actual values of the following parameters: maximal lifespan, 50% survival times, variance in control groups, and agreement with the distribution of deaths. These comparisons indicate that maximum likelihood and non-linear regression estimates provide better estimates of mortality rate parameters than log-linear regression. Of particular interest, the improved estimates indicate that most genetic manipulations in mice that increase lifespan do so by decreasing initial mortality rate, not rate of senescence, whereas most genetic manipulations that decrease lifespan surprisingly do so by increasing the rate of senescence, not initial mortality rate.

Introduction

Although many manipulations (such as decreasing environmental toxicity) can increase average or median lifespan, the rate of senescence, as measured by the age-associated exponential increase in mortality rate, is generally thought to be approximately invariant across groups within a species and is thus thought to reflect fundamental processes driving the aging process itself. For example, mortality rate in US men is at least 50% higher than in women at all ages, yielding a life expectancy lower by five years. However, since this mortality rate difference is present at all ages, it would be incorrect to conclude that men senesce more quickly than women (Arking 2006). In fact, male and female mortality rates converge with age, demonstrating that males exhibit increased initial mortality rate but reduced age-dependent acceleration of morality rate, compared to females, which partly accounts for the relatively small effect of the consistently elevated rate of senescence in men. To calculate the rate of senescence, it is useful to fit a family of model life tables with one parameter that can reasonably be identified with senescence rate, while other parameters may be identified with extrinsic hazards or general physiologic robustness that determine initial mortality rate.

The Gompertz equation has been used to model mortality rate since its introduction in 1825 (Olshansky & Carnes 1997). This equation relates the arithmetic progression of time to the geometric increase of mortality rate and is written mathematically as (Pletcher 1999):

$$h(t) = Ae^{Gt}$$

In the above equation, A is the initial mortality rate and G quantifies the age-dependent acceleration of mortality rate and is sometimes derived by log-transformation, in which case G is referred to as the "Gompertz slope". G is inversely related to the mortality rate doubling time (MRDT), by the formula MRDT = ln2/G. Studies of human populations have suggested that the MRDT is stable under extrinsic mortality influences that drastically change the initial mortality. For instance, Finch (1990) cites studies by Jones (1959) and Bergman (2007), who found that G in Australian prisoners of war during World War II was indistinguishable from that in civilian Australian females from 1944-1945, although the absolute mortality rates differed drastically between these groups. (Finch also points out that this value is virtually identical to that of U.S. females in 1980.) Finch (1990) confirmed this approximate invariance of G for other human populations. Similarly, values of G in seven different wild-caught strains of *Drosophila* were statistically indistinguishable even though they exhibited statistically distinguishable average lifespans (Spencer & Promislow 2005). In general, the rate of senescence, as operationally defined by the value of G in the Gompertz equation, is stable across populations of the same species even when average lifespan, all-cause mortality rate, and maximal lifespan (when studying a small cohort) vary dramatically. Thus senescence is considered to reflect a more fundamental general process that drives the age-dependent acceleration in disease and mortality.

Parameters in the Gompertz equation can be estimated by several different methods. The present paper compares how well empirically measured parameters are predicted by linear and non-linear regression, both of which use an ordinary least-squares method to fit the data, as well as maximum-likelihood estimates, which finds the values of the variables that would make the observed data most probable.

The limitations of using a least-squares linear regression on the empirical log hazard rate to determine mortality rate values have been enumerated in many reports (Eakin *et al.* 1995; Promislow *et al.* 1999; Garg *et al.* 1970; Mueller *et al.* 1995; Wilson 1993). Linear regression depends upon computing a log-transformed estimate of the mortality rate. Several papers have demonstrated that deriving mortality parameters from linear regression introduces a substantial bias when reconstructing the survival curves (Eakin *et al.* 1995; Garg *et al.* 1970; Wilson 1993), which we confirm here. Mueller (1995) also showed that linear regression is particularly problematic, yielding standard errors vastly larger than those of other methods when used to estimating mortality parameters for relatively small populations, in the range of population sizes for most published studies and all of the studies under consideration here. In summary, they wrote, "Our first conclusion is that the linear regression methods typically used to estimate parameters of the Gompertz do not work well and should probably never be used." As discussed by Promislow *et al*.(1999), estimates of the log mortality rate during early ages are highly variable and inaccurate; during late ages, when few survivors remain, the estimates again become highly variable. The common unweighted least-squares approach to regression strongly violates the general statistical principle that highly-variable data be given less weight in an estimate than less variable data. In addition, the log transformation becomes meaningless for time intervals with no deaths, a common occurrence when a small numbers of animals are used or the observations are frequent. For such an interval, the naïve mortality rate estimate would be 0 and the log of

0 is undefined or negative infinity. Methods that try to correct for this problem have been shown to further increase the bias (Promislow *et al.* 1996,1999).

A superior method for fitting mortality parameters is maximum-likelihood estimation (MLE). This approach, defines for each possible value of the Gompertz parameters the likelihood of observing the data – the particular times of death, which occurred – if the mortality rate were given exactly by a Gompertz curve with those parameters. The MLE is then the particular choice of parameters that maximizes this likelihood – that is, the parameters that would make the observations as likely as possible. Unlike linear regression, MLE does not require any transformation of the primary data, avoiding one source of bias. MLEs are also asymptotically efficient (under fairly general conditions), with asymptotically normal sampling distribution and computable asymptotic variance. The probability density function (PDF) based on the Gompertz hazard equation is written mathematically as (Lindgren 1993):

$$f(t) = A*\exp[\, Gt - (A/G)(e^{Gt} - 1)]$$

This equation gives the probability that an individual will die at time t and can be derived from the Gompertz hazard equation (Wang & Lee 2003).

A third method that can be used to calculate mortality rate parameters is non-linear regression (Wilson 1993). As with linear regression, non-linear regression relies on a least-squares method for determining the best fit and estimates for the mortality variables. It has been previously shown that this method provides reasonable estimates for the mortality variables and the present study confirms that finding (Wilson 1993). The Gompertz survival function can be derived from the hazard function and is written mathematically as (Garg *et al.* 1970):

$$S(t) = \exp[(A/G)*(1-e^{G*t})]$$

The Gompertz survival function gives the probability of surviving to time t. It can be derived from the Gompertz hazard equation (see Wang & Lee 2003 for the derivation).

De Magalhaes *et al.* (2005) recently reported that log-linear regression provides similar or superior estimates of Gompertz parameters than MLE. The present study, based on a larger set of data and several additional criteria, as well as boot-strap methods, indicates the opposite, that MLE methods are in fact robustly superior to log-linear regression methods.

Material and Methods

Survival data were obtained from supplementary data provided by de Magalhaes *et al*. (2005), other published papers (Miskin & Masos 1997; Schriner *et al.* 2005; Turturro *et al.* 2002; Weindruch & Walford 1982), or directly from the laboratory of Dr. Andrzej Bartke. Survival curve data from de Magalhaes *et al*.was double-checked to ensure that the values were approximately correct and two of the survival curves (PRDX1 and BubR1) were re-analyzed because the values obtained by de Magalhaes et al.did not fit as closely to the actual survival curve as the other data (Baker *et al.* 2004; de Magalhaes *et al.* 2005; Neumann *et al.* 2003). When survival curves were the only data available, values for analysis were estimated using either Adobe Photoshop CS (Adobe Systems Inc., San Jose CA, USA) or ImageJ (NIH, Bethesda Maryland, USA). MLE for

G and A variables were obtained with WinModest (Promislow *et al.* 1999) and re-confirmed with two separate scripts written for the freely available R statistical programming environment (David Steinsaltz and Kelvin Yen wrote independent MLE scripts). Hypothesis testing for the MLE values to test for significant changes in Gompertz variables utilized a likelihood ratio test, which compared the likelihood of obtaining G and A values for each group independently to the likelihood of obtaining either variable when one variable was forced to be the same between the groups (Eliason 1993). If the null hypothesis were true, that there is no difference between the constrained variable, the distribution of two times the difference between the two log-likelihoods will follow approximately a chi-squared distribution with one degree of freedom. If the observed value of the statistic is sufficiently exceptional for the chi-squared statistic, the null hypothesis would be rejected and the difference in the variable would be considered significant (Eliason 1993). In some cases, control data were unusable, so control data from the other 21 experiments were used to determine the average G and ln(A) values as well as the confidence intervals (2.90± .27 SE and -5.60± 0.56 SE respectively by MLE and 2.93± .25 and -5.43± .52 by NLR). While compromised controls raises complexities in the interpretation of these data , the ready availability of reasonably comparable control data led us to analyze these data as well. Because approximately normal confidence intervals are also generated when using Winmodest (Promislow *et al.* 1999) or Prism, the confidence intervals between the averaged control and experimental groups were compared to determine significance in these cases. Whenever control groups were from the same study, a likelihood ratio test was employed instead.

The Gompertz-Makeham (GM) equation is an extended form of the Gompertz equation (Finch 1990), with one additional parameter. The utility of the GM equation compared to the Gompertz equation was also assessed using maximum-likelihood methods in WinModest (Promislow *et al.* 1999) and was only found to be a statistically better fit in 4 out of 51 cases. This result supports recent studies that have used the simpler Gompertz equation (de Magalhaes *et al.* 2005; Harper *et al.* 2006), thereby simplifying comparisons between the studies.

G and A were estimated by non-linear regression (NLR) on survival curves by Prism (Graphpad Software Inc., San Diego CA, USA). An extra sum of squares F test was used to test if G and A, derived by NLR, were statistically distinguishable across groups. Runs test and D'Agostino and Pearson omnibus normality tests on the residuals were performed using Prism.

Ages at 50% survival ($S_{.5}$) were calculated using the following equation:

$$S_{.5} = (\ln(1-(G/A)*\ln(0.5)))/G$$

Maximum lifespan is also considered to constitute a key demographic parameter that, like senescence, is thought to be relatively stable across populations of the same species regardless of average lifespan (though influenced by manipulations that influence senescence such as dietary restriction). However, while this has an obvious meaning for a set of observed organisms, it is not so clear what "maximum lifespan" for an abstract model lifetable should be. For example, unlike fifty-percent survival, maximum lifespan depends on the number of individuals under observations. Finch and Pike (Finch & Pike

1996) defined maximum lifespan for a Gompertz lifetable applied to a population of size N by analogy with fifty-percent survival, as the time when the expected fraction of survivors is 1/N:

$T_{max} = \ln[1+G\ln(N)/A]/G.$

However, this ignores the randomness in the time of death of the last few individuals, thus producing a life-table parameter for which the last time of death in the population is not an unbiased estimator. (That is, if we average all the times of last death in the many population this would be expected to be larger than the average of the T_{max}.) Therefore the present study also implemented a more appropriate comparison between the maximum observed lifespan and the *expected maximum lifespan*. Since life expectancy may be computed as the integral of the survival function from 0 to infinity (see page 30 of Keyfitz and Caswell 2005), the expected time of the last death may be computed by applying the standard formula for the survival function of the maximum of N independent observations to the Gompertz survival function (see section 3.7 of Rice 1995), and then integrating to obtain

Kel..., 18/8/08 11:18
Deleted: is simply

Kel..., 18/8/08 11:23
Formatted: Font color: Red

$$\mathbb{E}[T_{\max}] = \int_0^\infty \left(1 - \left(1 - \exp\left\{\frac{A}{G}\left(e^{Gx} - 1\right)\right\}\right)^N\right) dx.$$

Bootstrap hypothesis testing was performed using scripts written by David Steinsaltz and Kelvin Yen. Samples were created using combinations of individual day of death. For NLR, the R function nls was used with the n2sol algorithm and the start values were the MLE of the dataset. If nls could not find an appropriate value, which occurred in an average of 2.2% of the time, a new random sample was selected to replace the set.

C. elegans studies

The N2 strain of *C. elegans* was used for all experiments and maintained on nematode growth media (NGM) plates seeded with OP50 strain of *E. coli* or axenic media supplemented with cholesterol. OP50 is a uracil-auxotrophic strain that has limited growth ability to facilitate the counting of the worms on the plate. Axenic media is a liquid media devoid of any bacteria and is made of 3% w/v soy peptone, 3% w/v yeast extract, 0.5 mg/ml hemoglobin, and 5 µg/mL cholesterol. This medium was supplemented with 50 µg/ml ampicillin and 25 µg/ml tetracycline to prevent bacterial growth.

Lifespan Assays

Lifespan studies were performed as in Adachi *et al.* 2000 (Adachi & Ishii 2000). Briefly, eggs were collected from gravid nematodes by standard hypochlorite treatment and grown on standard NGM plates until L4-adult stage. The worms were then transferred to media supplemented with 5-fluorodeoxyuridine (FUDR) to inhibit growth of progeny and transferred to fresh media every week or month for monoxenic or axenic cultures respectively. Worms were scored at least every 3 days to check for dead worms. Worms that were not moving or did not respond to gentle prodding from a platinum wire were scored as dead. Any worms that died of internal hatching or crawled off the plate

were censored on the date that they were last observed. All worms were maintained at 25°C for control temperature or 16°C for cold/hypothermia-induced longevity (CHIL).

Results

Validation of linear regression, non-linear regression, and maximum likelihood estimates

To compare the accuracy of each estimation procedure, several different actuarial parameters were computed from the estimated mortality parameters (Table 1). Observed 50% survival time was similarly predicted by MLE or NLR and less accurately by linear regression. The average relative difference between the predicted and actual 50% survival times differed by 7.0% or 5.0% when using either MLE or NLR respectively compared to 13.3% when using linear regression. Thus, both MLE and NLR were statistically superior* ($p < .001$) to linear regression at predicting this parameter. Comparing estimates of MLE to NLR, NLR is slightly, but significantly better than MLE.

Observed maximal lifespan was also similarly predicted by MLE and NLR, but less accurately by linear regression. The estimates of maximum lifespan produced by MLE and NLR differed from the maximal lifespan derived from the more empirically-based equation of Finch and Pike (see above) by 5.2% and 7.5% respectively, whereas the estimate produced by linear regression differed by 11.1% (use of the alternate equation provided very similar numbers). For either equation, both MLE and NLR were statistically better at predicting maximal lifespan ($p < .001$) than linear-regression. Comparing the results of MLE to NLR demonstrated that MLE was statistically better at predicting this parameter.

Variability in estimates of mortality rate doubling times (MRDT) between control groups from separate experiments were also similarly reduced by MLE and NLR compared to linear regression. Thus the coefficient of variation between control groups was significantly lower when derived by MLE (24.0%) or NLR (26.8%) than reported values estimated by linear-regression (42.9%) (de Magalhaes *et al.* 2005).

Similarly, MLE and NLR improved the estimate of the actual distribution of deaths (i.e. the integrated difference between the predicted probability distribution function and the actual distribution), by about 30% compared to LR. NLR was also significantly better than MLE, with a small reduction of 3%. Reconstruction of the original survival curves visually demonstrated the results of the quantitative analysis. As seen in figure 1, MLE and NLR re-create the survival curve with greater accuracy than LR.

Bootstrap methods are an alternative approach to generating 95% confidence intervals and testing hypotheses, which though computationally more demanding than parametric statistics do not assume certain constraints such as asymptotic normality (Carpenter & Bithell 2000; Efron & Tibshirani 1994). Therefore statistical comparisons by MLE and NLR, using likelihood ratio tests and F-tests, were compared to bootstrap methods (described in the Appendix). NLR statistical comparisons using the F-test frequently disagreed with the bootstrap tests (52% of the time), and always in the same

* Our use of the term "statistically superior" refers to a confidence test (two-sample T-test) that treats the studies (not entirely properly) as though they were a random sample from a superpopulation of possible studies, to which these methods might be applied.

direction: Every difference judged significant by the bootstrap test was also significant by the F-test, but about as many comparisons were found significant by the F-test but not by the bootstrap method. The bootstrap test based on the NLR estimates was also more conservative than the likelihood ratio test for the MLE. These two methods disagreed in about 15% of the comparisons, with the likelihood ratio test detecting significant differences in each case. It is perhaps worth noting that in all the differences, the p-value given by the bootstrap method was between 0.9-0.6. When comparing the bootstrap test for MLE to the likelihood ratio test for MLE, the two methods also disagreed about 15% of the time. In contrast to NLR, the MLE bootstrap hypothesis test detected one significant change that the likelihood ratio test did not (a significant increase in G for p53 KO mice). As with the NLR bootstrap results, 6 out of 8 differences yielded p-values between .9 and .6. Comparing the results of NLR bootstrap to the MLE bootstrap, there was a 94.4% agreement indicating that the use of bootstrap methods with either method for estimating G and A values may be valid.

Analysis of life-extending and life-reducing manipulations

To further validate the use of MLE and likelihood ratio test for hypothesis testing, lifespan studies in *C. elegans* were conducted and analyzed. As with studies in rats, dietary restriction (DR) significantly decreased the age-associated acceleration of mortality rate (Table 2). Cold/hypothermic induced longevity (CHIL) also significantly decreased the rate of senescence, in agreement with mortality data in *Drosophila* under cold conditions (Mair *et al.* 2003) (Table 2). Dietary restriction has also been shown to decrease core body temperature in rodents and decreased body temperature has also been shown to extend lifespan in mice (Rikke *et al.* 2004; Conti *et al.* 2006). Mutations in the insulin/IGF pathway have also been shown to extend lifespan and reduce age-associated acceleration of mortality rate (Johnson, 1990); in the present study, MLE and the likelihood ratio test applied to worms with the *daf-2* mutation led to the same conclusion (Table 2).

MLE

Hypothesis testing using the likelihood ratio test on the maximum likelihood estimates show that of the life-extending manipulations examined, MLE only detected reduced G in *prop1* mice in the larger of the two experiments, dietary restriction in Ames mice, and dietary restriction in rats. Other life-extending manipulations were found to only decrease the initial mortality rate: fat-specific insulin receptor KO (firko) mice, urokinase-type plasminogen activator transgenic mice, and female *prop1* KO mice.

Of the lifespan-reducing manipulations, several significantly increased G. This list includes mouse strains *Gh, Klotho*, *Lmna, PolgA*, and *Sam*. NZB/W mice exhibited a significant increase in senescence and this increase was ameliorated by addition of fish oil to the diet. Unusually, manipulation of Prdx1 and Top3b reduced decreased rates of senescence but still shortened lifespans due to increased initial mortality rates. Significance tests on Klotho, Lmna, NZB/W, PolgA, Prdx1, and Top3b mice must be interpreted with caution because of the small number of individuals in the population studied. Judgments of significance were based on the asymptotic normality of the MLE, which is valid only for sufficiently large samples (perhaps around 60 animals, according to Eliason (1993)). The studies in mice with altered p53 expression had usable control

mice and these were analyzed using a likelihood ratio test. Either increased or decreased expression of p53 led to an increase in the initial mortality rate although was only significant for the decreased expression.

NLR

The sum-of-squares F test is widely used in conjunction with regression analysis. While basic assumptions of the F test are not satisfied for these data, it was nevertheless of interest to compare the results of such analyses to the results from MLE and NLR. In particular, for the many cases of differing lifespan but no significant change in parameters according to the likelihood ratio test, it was of interest to assess if a significant change in A or G would be detected by the F test. In the present study, the F test detected significance more than twice as often as the likelihood ratio test. All the significant changes in G detected by the likelihood ratio test were similarly detected by the F test. In addition, the F-test detected significant effects on G in 2 of the 4 dietary restriction studies, and in studies in mice involving the manipulation of several genes (C/EBP mice, pit1, Ghrhr, Igf1r, prop1, MsrA, and mitochondrially targeted catalase). As with the likelihood ratio test, the F-test detected significant reduction in G yet shortened lifespans with genetic manipulations of Atm+Terc, Prdx1, and Top3b. Table 1 contains a summary of the results from MLE and NLR.

Discussion

Although with larger datasets linear regression has been shown to give reasonable estimates of the variables of the Gompertz hazard function (Finch & Pike 1996), the varying methods by which researchers must try to compensate for the inherent problems of linear regression, leads to less consistent results with smaller datasets (de Magalhaes *et al*. 2005; Promislow *et al.* 1999). Comparison of linear regression, maximum likelihood, and non-linear regression methods demonstrate the superiority of maximum likelihood methods (ML) and non-linear regression (NLR) over linear regression (LR) in a number of parameters. Both methods are easily implemented with currently available software and similar in recovering the true 50% and maximal survival times, as well as minimizing the variability in MRDT between control groups. ML methods may nevertheless be favored because of its statistical properties, which are important for hypothesis testing. MLEs are asymptotically (i.e., as the number of samples tend to infinity) unbiased, efficient, and normally distributed (Eliason 1993), thus constituting the preferred method for estimating Gompertz parameters. The F-test with NLR detected a larger number of changes, but most of these are evidently spurious, as indicated by comparison with the bootstrap test. While the residuals are normal (a common test for appropriateness of the F test), the basic assumption of homoscedasticity was not satisfied. In fact, simulation tests readily show that the standard test statistic does not have anything close to the F distribution.

In the present study, MLE and the likelihood ratio test confirmed the positive results by de Magalhaes *et al*., with a number of exceptions. We did not find that C/EBP, GHR, pit-1, p66, or MSRA had an effect on G. MLE and the likelihood ratio test confirmed that dietary restriction decreases the rate of senescence in rats and worms. Similarly, dietary restriction also decreased the rate of senescence in one of the strains of mice analyzed here and Harper *et al*. (2006) also reported a decrease in the rate of

senescence in a heterogeneous mouse colony recently derived from wild mice subject to dietary restriction. MLE and the likelihood ratio test also confirmed that cold temperature and defective insulin/IGF signaling significantly reduces G in nematodes (Johnson, 1990).

Based on 4 criteria, ability to recover maximal lifespan, recover 50% survival times, match the overall distribution of deaths, and minimize the variation in MRDT between control groups, ML and NLR are both superior to LR, and only slightly different from each other. While there are some differences in the hypothesis tests, the results show a high degree of agreement, as long as bootstrap tests are used to avoid relying on unfounded assumptions. The similarity in results from the likelihood ratio test and bootstrap method for hypothesis testing (as well as the results of our simulations) suggests that the likelihood ratio test is also a valid method of hypothesis testing for these kinds of data. Therefore, the use of MLE and the likelihood ratio test seems to be the best method for analyzing lifespan data. In approximately 40% of the analyzed data, the MLE combined with likelihood ratio test was able to detect significant effects. In fact, in two of the four cases where the total number of animals analyzed was less than 30, significant changes in either A or G were detected. This suggests that even with a relatively small number of animals Gompertz analysis is possible and should be conducted.

Acknowledgements
We would like to thank Andrzej Bartke and Mike Bonkowski for providing us with their survival data. We would also like to thank Joao de Magalhaes for help with his data set. These studies were funded in part by the Ellison Medical Foundation. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR).

Appendix: Bootstrap Methods

We have used two different simulation approaches to confirm our analyses. First, we tested the appropriateness of the chi-squared approximation for the likelihood ratio test. Second, we performed a non-parametric bootstrap test complementary to the likelihood ratio test. We describe both of these briefly here. The original software, written in the R programming language, is available on request.

1) The likelihood ratio test is based on the logic that if the null hypothesis is true, the transformed likelihood ratio will converge to a chi-squared distribution as the number of samples goes to infinity. In this case, the null hypothesis states that the deaths came from exactly Gompertz hazard rates, and that the Gompertz parameter G (or Gompertz parameter A) is identical between the two populations. Two potential concerns that need to be addressed are: First, if the number of samples is small, the asymptotic result may not be relevant; Second, if the hazard rates are not do not conform to the Gompertz equation, the mathematical theory will tell us nothing about the behavior of the maximum-likelihood fit to the misspecified model. It is obvious that a simulation approach can inform about the behavior of small samples drawn from the correct model, thus addressing the first point, to a reasonable extent. The second point is more problematic, because the null hypothesis is not even defined if the data are not drawn from a Gompertz model.

To enhance the validity of our analysis, at least with respect to age-dependent acceleration of mortality, we redefine the null hypothesis as: The hazard rates of the two populations differ by a fixed number of days. In the case of Gompertz hazard, this is simply the original null hypothesis, but the new version is meaningful for any hazard rates. We can now ask the question: Suppose we have two sets of mortality data, satisfying the generalized null hypothesis. They are not drawn from a Gompertz curve, but are matched to our actual observations; following the bootstrap principle, we resample one group from the empirical observed mortality in one of our experiments, and another from the same mortality shifted backward or forward by a fixed number of days. We resample the same number of deaths that were actually observed. We fit them to the Gompertz model with and without the constraint that the Gompertz slopes are the same, and compute -2 times the log likelihood ratio. This process was repeated 1000 times, and the collection of outcomes was compared with the appropriate chi-squared distribution. Somewhat surprisingly, in all of the data sets we tried, even with as few as 15 simulated individuals, the fit was extremely good. More investigation is needed, but this preliminary result suggests that we should have some confidence in the results of the likelihood ratio test.

2) As an alternative, we measure the variability in the A and G parameters directly with a nonparametric bootstrap simulation. We resample from the observed deaths in the experimental and control data, and to each resampled set of deaths we fit the Gompertz parameters. Both parameters were found to be fairly close to log normal, so we performed a standard normal hypothesis test on the difference in the log parameters, using the means and variances estimated from the bootstrap samples. As would be

expected (because of the optimality of the likelihood ratio) this test proved generally more conservative than the likelihood ratio test, but the outcomes were quite similar.

References

Adachi, H., Ishii, N., 2000. Effects of Tocotrienols on Life Span and Protein Carbonylation in Caenorhabditis elegans. J. Gerontol. A Biol. Sci. Med. Sci. 55, B280-285.

Arking, R., 2006. The Biology of Aging: Observations and Principles. Oxford University Press, New York.

Baker, D.J., Jeganathan, K.B., Cameron, J.D., Thompson, M., Juneja, S., Kopecka, A., Kumar, R., Jenkins, R.B., de Groen, P.C., Roche, P., van Deursen, J.M., 2004. BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. Nat. Genet. 36, 744-749.

Bartke, A., 2003. Can growth hormone (GH) accelerate aging? Evidence from GH-transgenic mice. Neuroendocrinology 78, 210-216.

Bartke, A., Wright, J.C., Mattison, J.A., Ingram, D.K., Miller, R.A., Roth, G.S., 2001. Longevity: Extending the lifespan of long-lived mice. Nature 414, 412.

Bergman, A., Atzmon, G., Ye, K., MacCarthy, T., Barzilai, N., 2007. Buffering Mechanisms in Aging: A Systems Approach Toward Uncovering the Genetic Component of Aging. PLoS Comput. Biol. 3, e170.

Bluher, M., Kahn, B.B., Kahn, C.R., 2003. Extended longevity in mice lacking the insulin receptor in adipose tissue. Science 299, 572-574.

Brown-Borg, H.M., Borg, K.E., Meliska, C.J., Bartke, A., 1996. Dwarf mice and the ageing process. Nature 384, 33.

Carpenter, J., Bithell, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 19, 1141-1164.

Chang, S., Multani, A.S., Cabrera, N.G., Naylor, M.L., Laud, P., Lombard, D., Pathak, S., Guarente, L., DePinho, R.A., 2004. Essential role of limiting telomeres in the pathogenesis of Werner syndrome. Nat. Genet. 36, 877-882.

Chiu, C.-H., Lin, W.-D., Huang, S.-Y., Lee, Y.-H., 2004. Effect of a C/EBP gene replacement on mitochondrial biogenesis in fat cells. Genes Dev. 18, 1970-1975.

Conde, C., Weller, S., Gilfillan, S., Marcellin, L., Martin, T., Pasquali, J.-L., 1998. Terminal Deoxynucleotidyl Transferase Deficiency Reduces the Incidence of Autoimmune Nephritis in (New Zealand Black New Zealand White)F1 Mice. J. Immunol. 161, 7023-7030.

Conti, B., Sanchez-Alavez, M., Winsky-Sommerer, R., Morale, M.C., Lucero, J., Brownell, S., Fabre, V., Huitron-Resendiz, S., Henriksen, S., Zorrilla, E.P., de Lecea, L., Bartfai, T., 2006. Transgenic Mice with a Reduced Core Body Temperature Have an Increased Life Span. Science 314, 825-828.

Coschigano, K.T., Holland, A.N., Riders, M.E., List, E.O., Flyvbjerg, A., Kopchick, J.J., 2003. Deletion, But Not Antagonism, of the Mouse Growth Hormone Receptor Results in Severely Decreased Body Weights, Insulin, and Insulin-Like Growth Factor I Levels and Increased Life Span. Endocrinology 144, 3799-3810.

de Magalhaes, J.P., Cabral, J.A.S., Magalhaes, D., 2005. The Influence of Genes on the Aging Process of Mice: A Statistical Assessment of the Genetics of Aging. Genetics 169, 265-274.

Eakin, T., Shouman, R., Qi, Y., Liu, G., Witten, M., 1995. Estimating parametric survival model parameters in gerontological aging studies: methodological problems and insights. J. Gerontol. A Biol. Sci. Med. Sci. 50, B166-176.
Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap Chapman & Hall/CRC
Eliason, S.R., 1993. Maximum Likelihood Estimation Logic and Practice. Sage Publications, Newbury Park.
Finch, C.E., 1990. Longevity, senescence, and the genome. The University of Chicago Press, Chicago.
Finch, C.E., Pike, M.C., 1996. Maximum life span predictions from the Gompertz mortality model. J. Gerontol. A Biol. Sci. Med. Sci. 51, B183-794.
Flurkey, K., Papaconstantinou, J., Miller, R.A., Harrison, D.E., 2001. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proc. Natl. Acad. Sci. U. S. A. 98, 6736-6741.
Garg, M.L., Rao, B.R., Redmond, C.K., 1970. Maximum-Likelihood Estimation of the Parameters of the Gompertz Survival Function. Appl. Stat. 19, 152-159.
Harper, J.M., Leathers, C.W., Austad, S.N., 2006. Does caloric restriction extend life in wild mice? Aging Cell 5, 441-449.
Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloen, A., Even, P.C., Cervera, P., Le Bouc, Y., 2003. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature 421, 182-187.
Jolly, C.A., Muthukumar, A., Avula, C.P.R., Troyer, D., Fernandes, G., 2001. Life Span Is Prolonged in Food-Restricted Autoimmune-Prone (NZB NZW)F(1) Mice Fed a Diet Enriched with (n-3) Fatty Acids. The Journal of Nutrition 131, 2753-2760.
Johnson, T.E., 1990. Increased life-span of age-1 mutants in Caenorhabditis elegans and lower Gompertz rate of aging. Science 249, 908-12.
Jones, J.B., 1959. The Relation of the Human Health to Age, Place, and Time, in: Birren, J.E. (Ed.), Handbook of Aging and the Individual. The University of Chicago Press, Chicago, pp. 336-337.
Keyfitz, N., Caswell, H. Applied Mathematical Demography. Springer, Cambridge, MA.
Kuro-o, M., Matsumura, Y., Aizawa, H., Kawaguchi, H., Suga, T., Utsugi, T., Ohyama, Y., Kurabayashi, M., Kaname, T., Kume, E., Iwasaki, H., Iida, A., Shiraki-Iida, T., Nishikawa, S., Nagai, R., Nabeshima, Y.-i., 1997. Mutation of the mouse klotho gene leads to a syndrome resembling ageing. Nature 390, 45-51.
Kwan, K.Y., Wang, J.C., 2001. Mice lacking DNA topoisomerase IIIbeta develop to maturity but show a reduced mean lifespan. Proc. Natl. Acad. Sci. U. S. A. 98, 5717-5721.
Lee, E.T., Wang, J.W., 2003. Statistical Methods for Survival Data Analysis. Wiley, John & Sons, Incorporated, New York, NY.
Lindgren, B.W., 1993. Statistical Theory, Fourth ed. Chapman & Hall /CRC, Boca Raton, Florida.
Mair, W., Goymer, P., Pletcher, S.D., Partridge, L., 2003. Demography of Dietary Restriction and Death in Drosophila. Science 301, 1731-1733.
Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboldi, P., Pandolfi, P.P., Lanfrancone, L., Pelicci, P.G., 1999. The p66shc adaptor protein controls oxidative stress response and life span in mammals. Nature 402, 309-13.
Miskin, R., Masos, T., 1997. Transgenic mice overexpressing urokinase-type

Kel..., 18/8/08 11:27
Formatted: Font color: Red

plasminogen activator in the brain exhibit reduced food consumption, body weight and size, and increased longevity. J. Gerontol. A Biol. Sci. Med. Sci. 52, B118-24.

Mitsui, A., Hamuro, J., Nakamura, H., Kondo, N., Hirabayashi, Y., Ishizaki-Koizumi, S., Hirakawa, T., Inoue, T., Yodoi, J., 2002. Overexpression of human thioredoxin in transgenic mice controls oxidative stress and life span. Antioxidants & Redox Signaling 4, 693-6.

Moskovitz, J., Bar-Noy, S., Williams, W.M., Requena, J., Berlett, B.S., Stadtman, E.R., 2001. Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. Proc. Natl. Acad. Sci. U. S. A. 98, 12920-12925.

Mounkes, L.C., Kozlov, S., Hernandez, L., Sullivan, T., Stewart, C.L., 2003. A progeroid syndrome in mice is caused by defects in A-type lamins. Nature 423, 298-301.

Mueller, L.D., Nusbaum, T.J., Rose, M.R., 1995. The Gompertz equation as a predictive tool in demography. Exp. Gerontol. 30, 553-569.

Neumann, C.A., Krause, D.S., Carman, C.V., Das, S., Dubey, D.P., Abraham, J.L., Bronson, R.T., Fujiwara, Y., Orkin, S.H., Van Etten, R.A., 2003. Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression. Nature 424, 561-565.

Olshansky, S.J., Carnes, B.A., 1997. Ever since Gompertz. Demography 34, 1-15.

Promislow, Tatar, Pletcher, Carey, 1999. Below-threshold mortality: implications for studies in evolution, ecology and demography. J. Evol. Biol. 12, 314-328.

Promislow, D.E.L., Tatar, M., Khazaeli, A.A., Curtsinger, J.W., 1996. Age-Specific Patterns of Genetic Variance in Drosophila melanogaster. I. Mortality. Genetics 143, 839-848.

Quick, K.L., Ali, S.S., Arch, R., Xiong, C., Wozniak, D., Dugan, L.L., 2008. A carboxyfullerene SOD mimetic improves cognition and extends the lifespan of mice. Neurobiol. Aging 29, 117-128.

Rice, J.A., 1995, Mathematical Statistics and Data Analysis, 2nd edition, Duxbury Press, Belmont CA.

Rikke, B.A., Yerg, J.E., III, Battaglia, M.E., Nagy, T.R., Allison, D.B., Johnson, T.E., 2004. Quantitative Trait Loci Specifying the Response of Body Temperature to Dietary Restriction. J. Gerontol. A Biol. Sci. Med. Sci. 59, B118-125.

Schriner, S.E., Linford, N.J., Martin, G.M., Treuting, P., Ogburn, C.E., Emond, M., Coskun, P.E., Ladiges, W., Wolf, N., Van Remmen, H., Wallace, D.C., Rabinovitch, P.S., 2005. Extension of Murine Life Span by Overexpression of Catalase Targeted to Mitochondria. Science 308, 1909-1911.

Spencer, C.C., Promislow, D.E.L., 2005. Age-Specific Changes in Epistatic Effects on Mortality Rate in Drosophila melanogaster. J. Hered. 96, 513-21.

Takeda, T., Hosokawa, M., Takeshita, S., Irino, M., Higuchi, K., Matsushita, T., Tomita, Y., Yasuhira, K., Hamamoto, H., Shimizu, K., 1981. A new murine model of accelerated senescence. Mech. Ageing Dev. 17, 183-194.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T., Larsson, N.-G., 2004. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature 429, 417-423.

Turturro, A., Duffy, P., Hass, B., Kodell, R., Hart, R., 2002. Survival Characteristics and

Age-Adjusted Disease Incidences in C57BL/6 Mice Fed a Commonly Used Cereal-Based Diet Modulated by Dietary Restriction. J. Gerontol. A Biol. Sci. Med. Sci. 57, B379-389.

Tyner, S.D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., Lu, X., Soron, G., Cooper, B., Brayton, C., Hee Park, S., Thompson, T., Karsenty, G., Bradley, A., Donehower, L.A., 2002. p53 mutant mice that display early ageing-associated phenotypes. Nature 415, 45-53.

Weindruch, R., Walford, R.L., 1982. Dietary restriction in mice beginning at 1 year of age: effect on life-span and spontaneous cancer incidence. Science 215, 1415-1418.

Wilson, D.L., 1993. A comparison of methods for estimating mortality parameters from survival data. Mech. Ageing Dev. 66, 269-281.

Wong, K.-K., Maser, R.S., Bachoo, R.M., Menon, J., Carrasco, D.R., Gu, Y., Alt, F.W., DePinho, R.A., 2003. Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. Nature 421, 643-648.

Yu, B.P., Masoro, E.J., Murata, I., Bertrand, H.A., Lynd, F.T., 1982. Life span study of SPF Fischer 344 male rats fed ad libitum or restricted diets: longevity, growth, lean body mass and disease. J. Gerontol. 37, 130-41.

Tables

Gene/ Intervention	Type	Strain	Nonlinear Regression		Maximum likelihood estimate		N	Reference
			Ln(A)	G	Ln(A)	G		
Atm, Terc	KO	C57BL/6 x WW6	-1.22+	1.29*	-1.50+	1.60*	51	Wong *et al*. (2003)
Autoimmune Prone (NZB/w mice)	WT	(NZW x NZB)F1	-3.32	6.92+	-3.16	6.25+	22	Conde *et al.* (1998)
Bub1b	H/H	129/Sv x FVB	-1.13+	3.52	0.32+	2.55	212	Baker *et al.* (2004)
C/ebp	b/+	C57BL/6J	-5.71	3.33	-6.02	3.40	30	Chiu *et al*. (2004)
	b/b		-4.85+	2.18*	-5.29	2.33	30	
	ob/ob		-4.19	3.28	-4.81	3.65	40	
Dietary Restriction	AL	C57BL/6J	-5.56	3.04	-6.04	3.25	24	Weindruch and Walford (1982)
	DR		-6.14	2.62*	-5.57	2.38	29	
	AL	C57BL/10Sn x C3H/HeDiSn	-6.46	2.52	-6.95	2.70	68	
	DR		-7.25*	2.52	-7.44	2.55	67	
Dietary Restriction	AL	male, C57BL/6J	-5.55	2.8	-5.21	2.56	56	Turturro *et al.* (2002)
	DR		-6.19	2.45	-6.43	2.48	56	
	AL	female, C57BL/6J	-4.94	2.7	-4.68	2.47	56	
	DR		-5.74*	2.25*	-5.56	2.13	56	
Dietary Restriction	AL	male, Rats	-9.26	5.31	-7.26	4.02	115	Yu *et al*. (1982)
	DR		-4.88+	1.8*	-4.99+	1.80*	115	
Cat	WT	B6(B6C3F1) x C57BL/6J (4403)	-5.72	2.95	-4.95	2.44	58	Schriner *et al*. (2005)
	mitochondrial TG		-5.34	2.26*	-5.95	2.48	42	
Fish Oil (FO)	AL	(NZB x NZW)F1	-6.05	11.56	-5.81	10.62	30	Jolly *et al.* (2001)
	FO	females	-4.14+	5.74*	-4.17	5.54*	30	
bGh (bovine growth hormone)	WT	C56BL/6 x SJL	-5.23	2.4	-4.56	2.14	16	Bartke *et al.* (2003)
	Tg		-7.08	6.24+	-6.73	6.06+	9	
Ghr	WT	OLA-BALB/cJ	-3.56	2.05	-3.48	1.95	15	Coschigano *et al.* (2003)
	KO		-5.4*	2.35	-5.25	2.14	11	
Ghrhr	WT	C57BL/6	-6.35	2.82	-6.77	2.86	31	Flurkey *et al*. (2001)
	KO		-9.34*	3.33+	-8.47	2.90	35	
Igf1r	WT	129/J	-2.58	1.77	-3.36	2.25	17	Holzenberger *et al.* (2003)
	KO		-5.41*	2.85+	-5.62	2.88	20	
Insr (Firko)	WT	FVB x 129S4	-4.02	1.95	-4.75	2.22	67	Bluher *et al.* (2003)
	KO- fat specific		-8.5*	3.51+	-7.56*	2.98	60	
Kl (klotho)	KO	C57BL/6J + C3H/J	-1.67	31.84+	-0.88	22.06+	29	Kuro-o *et al.* (1997)

Lmna	KO	C57BL/6 + 129S1/Sv	-0.97+	62.23+	-1.27	61.06+	25	Mounkes *et al*. (2003)
MsrA	WT	C57BL/6 + 129/SvJ	-2.8	1.86	-4.13	2.58	14	Moskovitz *et al*. (2001)
	KO		-2.64	4.39+	-3.12	4.23	17	
p53	WT	129/SV	-7.31	3.51	-7.22	3.34	56	Tyner *et al.* (2002)
	+/-	+ C57BL/6	-3.96+	3.2	-3.70+	2.75	217	
	+/m		-6.04	3.62	-5.20	2.97	35	
p66	WT	129/Sv	-13.31	7.22	-13.50	7.17	14	Migliaccio *et al.* (1999)
	KO		-7.05+	3.09*	-8.25	3.57	15	
Pit1	WT	C3H/HeJ	-7.6	3.56	-7.66	3.44	34	Flurkey *et al.* (2001)
	KO	+DW/J	-8.86	2.84*	-7.96	2.47	25	
Plau (alpha-MUPA)	WT	FVB/N	-4.89	2.43	-4.87	2.38	33	Miskin and Masos (1997)
	TG		-5.99	2.36	-8.47*	3.32	33	
PolgA	mut	129 + C57BL/6	-10.06*	13.29+	-8.57*	11.27+	38	Trifunovic *et al*. (2004)
Prdx1	KO	B6 x 129SvEv	-3.15+	1.55*	-3.15	1.46*	34	Neumann *et al.* (2003)
Prop1	WT	Ames Stock	-3.99	2.38	-3.80	2.04	13	Brown-Borg *et al.* (1996)
	KO		-10.41	3.27+	-9.57*	2.89	16	
Prop1	WT-AL	Ames	-6.1	3.51	-6.68	3.80	26	Bartke *et al*. (2001)
	WT-DR	Ames	-3.48	1.47*	-4.26	1.81*	27	
	KO-AL	Ames	-5.33+	2.1*	-6.05	2.33*	24	
	KO-DR	Ames	-5.11+	1.65*	-5.16	1.70	24	
Sam	WT	AKR	-2.74	3.17	-2.22	2.36	377	Takeda *et al*. (1981)
	SAM		-2.88	4.87+	-2.48	3.94+	493	
SOD mimetic C3	WT	C57BL/6	-5.54	3.01	-6.02	3.17	123	Quick *et al*. (2006)
	C3		-6.41*	3.17	-6.07	2.89	115	
Top3B	KO	C57BL/6J +129/svEv	-0.81+	0.57*	-1.10+	0.89*	30	Kwan and Wang (2001)
Trx	WT	C57BL/6	-3.1	2.39	-3.73	2.67	82	Mitsui *et al*. (2002)
	TG		-4.62*	2.86	-4.26	2.52	94	
Wrn, Terc	KO	C57BL/6 + 129Sv + BALB/c + SLJ	-0.54+	2.2	-0.67+	2.08	39	Chang *et al*. (2004)

Table 1: Nonlinear regression values and maximum likelihood estimates (MLE) for Gompertz equation variables. Significance was determined by F-test for NLR values or likelihood ratio test for MLEs unless shaded. Shaded boxes indicate that the data have an unusable experimental control and therefore a generalized control group was used (see methods for more information). Tests for significance in these groups may not be accurate. +significant increase *significant decrease

Gene/ Intervention	Type	Maximum likelihood estimate		N
		Ln(A)	G	
Dietary Restriction	WT	-7.94	0.49	57
	Dietary Restricted	-6.45	0.18*	42
Cold Temperature	WT	-5.88	0.31	149
	Cold Temperature	-7.61	0.21*	160
Insulin/IGF Mutation	WT	-9.14	0.37	39
	Daf-2	-7.55	0.17*	57

Table 2: Maximum likelihood estimates for Gompertz parameters in C. elegans. Comparisons in A values were not performed. * significant decrease

Fig. 1. Reconstruction of survival curves based on values obtained from linear-regression (LR), maximum likelihood (MLE), or non-linear-regression (NLR) methods. Four example reconstructions are shown for the following experiments: (A) replacement of CCAAT/enhancer-binding protein alpha (C/EBP) with C/EBP-beta, (B) bovine growth hormone (bGH) transgenic mice, (C) p66 mutant mice, and (D) senescence accelerated mice (Sam). Values for LR were obtained from de Magalhaes *et al.* (de Magalhaes *et al.* 2005).