



THE GOMPERTZ EQUATION AS A PREDICTIVE TOOL IN DEMOGRAPHY

LAURENCE D. MUELLER, THEODORE J. NUSBAUM and MICHAEL R. ROSE

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717

Abstract—The Gompertz demographic model describes rates of aging and age-independent mortality with the parameters α and A , respectively. Estimates of these parameters have traditionally been based on the assumption that mortality rates are constant over short to moderate time periods. This assumption is questionable even for very large samples assayed over short time intervals. In this article, we compare several methods for estimating the Gompertz parameters, including some that do not assume constant mortality rates. A maximum likelihood method that does not assume constant mortality rates is shown to be best, based on the bias and variance of the Gompertz parameter estimates. Moreover, we show how the Gompertz equation can then be used to predict mean longevity and the time of the n_{th} percentile of mortality. Methods are also developed that assign confidence intervals to such estimates. In some cases, these statistics may be estimated accurately from only the early deaths of a large cohort, thus providing an opportunity to estimate longevity on long-lived organisms quickly.

Key Words: Gompertz equation, estimating longevity, jackknife, demography, bootstrap, maximum likelihood

INTRODUCTION

THERE ARE compelling reasons to believe that a single phenomenon, the decline in the force of natural selection, is the cause of aging in plants and animals (Rose, 1991). But prior to the application of evolutionary theory to the problem of aging, phenomenological models were developed that attempted to capture the observation of increasing rates of mortality with age. One such model is that due to Benjamin Gompertz (1825), the "Gompertz model," which assumes that mortality rates increase exponentially with age. Although evolutionary theory shows that rates of mortality will increase with age in most multicellular organisms (Charlesworth, 1980), there is no deductive demonstration for the exponential form of the Gompertz model. Nevertheless, the Gompertz

Correspondence to: L.D. Mueller
(Received 27 May 1994; Accepted 20 April 1995)

model has proven popular as a general description of many observed patterns of mortality (e.g., Finch, 1990).

Recently, several large studies with medflies (Carey *et al.*, 1992) and *Drosophila* (Curtsinger *et al.*, 1992) have concluded that rates of mortality may not increase exponentially in the oldest age classes, contrary to the Gompertz model. Brooks, Lithgow, and Johnson (1994) collected data that suggest that this phenomenon may be due, in part, to the existence of genetic or phenotypic subpopulations with very different rates of aging. Despite this uncertainty, there are still potentially useful applications of the Gompertz equation.

In this article, we explore some practical issues concerning the use of the Gompertz equation to infer demographic parameters. We test these methods using a large array of data we have collected from populations of *Drosophila melanogaster*. For many species, it may be quite difficult and time consuming to follow a cohort through their entire life span in order to estimate the mean longevity, maximum longevity, and the parameters of the Gompertz equation. We show how one may use information from deaths early in life to estimate the mean longevity and the time to which the last 5 or 1% of the population is expected to survive. In addition, we develop techniques for placing confidence intervals on these estimates, and we test their accuracy using computer situations.

STATISTICAL ANALYSIS

Gompertz equation

Mortality data is typically collected by following an initial cohort of N_0 individuals over time and recording the number of survivors, N_t , at fixed intervals. The Gompertz equation that describes the instantaneous rate of mortality in these types of experiments is,

$$u(t) = \frac{-1}{N} \frac{dN}{dt} = A \exp(\alpha t). \quad (1)$$

Consequently, the number of survivors at any time in the future can be described with eq. (1) as,

$$N_t = N_0 p_t \quad (2)$$

where,

$$p_t = \exp\left\{\frac{A(1 - \exp(\alpha t))}{\alpha}\right\} \quad (3)$$

p_t , then, is the probability of any single individual surviving to age t . Let the observation vector of a single experiment be $\mathbf{T} = (t_1, t_2, \dots, t_{N_0})$, where t_i is the time of death of the i th individual.

Methods of estimation

There are several approaches that have been taken to estimating the values of A and α from eq. (1). Here, we briefly review four of these methods.

Linear regression. These methods used by Brooks *et al.* (1994), Hughes and Charlesworth (1994), Promislow (1991), Tatar and Carey (1994), and others works with the log transformed version of eq. (1),

$$\ln(u(t)) = \ln(A) + \alpha t. \quad (4)$$

To estimate the dependent variables in (4), observations of mortality are divided into discrete intervals. The number of deaths in each interval is used as an estimate of $u(t)$. For instance, to estimate $u(t)$ from the number of survivors the following approximation is generally used,

$$u(t) \cong -\ln(P_t) = -\ln\left(\frac{N_{t+1}}{N_t}\right). \quad (5)$$

The length of the time intervals used must be large enough to include at least one observed death. This technique implicitly assumes that $u(t)$ does not change in the interval t to $t + 1$ which, of course, is contrary to the form of eq. (1). Such an approximation can only be expected to work if the time interval is very small.

Nonlinear regression. This method will find values of A and α , as the parameters that minimize the residual sum of squares below (Gavrilov and Gavrilova, 1991; Hirsh *et al.*, 1994),

$$\sum_{i=1}^{N_0-1} \left[\ln\left(\frac{N_0 - i}{N_0}\right) - \ln p_{t_i} \right]^2 \quad (6)$$

which can be found using standard numerical search algorithms (Marquardt, 1963). The last death is not utilized in the estimation process, because eq. (6) is undefined in that single case. Using natural logarithms reduces the chance of obtaining numerical overflows during the search of the $A - \alpha$ space for the least-squares solutions because the logarithm of p_t is invariably a smaller number (either in absolute value or in its exponent) than p_t .

Maximum likelihood with approximation. To describe the likelihood function (Fukui *et al.*, 1993) we first note that the last death among a cohort of N_0 individuals is on day t_{N_0} . Thus, we can also record the number of deaths on each day as,

$$d_1, d_2, \dots, d_{t_{N_0}}.$$

Let $q(t)$ be the probability that an individual that lived to day t , dies by time $t + 1$, then the likelihood function can then be defined as,

$$L = \prod_{i=1}^{i=t_{N_0}} \binom{N_i}{d_i} q(i)^{d_i} (1 - q(i))^{(N_i - d_i)} \tag{7}$$

Equation (7) is then evaluated assuming that $-\ln(1 - q(t)) = A \exp(\alpha t)$, which is formally the same as eq. (5). The likelihood function follows from the assumption that the number of deaths (d_i) on day i is a binomial random variable with probability $q(i)$.

Maximum likelihood. The most accurate application of the maximum likelihood technique would be to use eq. (7), estimating $q(t)$ as $1 - p_{t+1} (p_t)^{-1}$ from eq. (3).

Important results derived from the Gompertz equation

Because p_t is the chance of surviving to time t , the chance of an individual dying before time t is simply,

$$\text{Prob}(t_i \leq t) = 1 - p_t \tag{8}$$

If we view the time of death as a random process, then $1 - p_t$ is the distribution function (which by definition is $\text{Prob}(t_i \leq t)$) of this process. Further, the density function can be derived as,

$$f(t) = \frac{\partial(1 - p_t)}{\partial t} = A \exp \left\{ \frac{A(1 - \exp(\alpha t))}{\alpha} + \alpha t \right\} \tag{9}$$

The Gompertz equation then yields as an estimate of the mean longevity,

$$\int_0^\infty t f(t) dt. \tag{10}$$

In addition to the mean longevity, the time (t_β) by which some critical fraction (β) of the population is expected to be dead may be calculated from eqs. (3) and (8) as

$$t_\beta = \frac{\ln \left[1 - \frac{\alpha \ln(1 - \beta)}{A} \right]}{\alpha}. \tag{11}$$

Bootstrap

The bootstrap is a numerically intensive technique for making statistical inferences (for a recent review, see Efron and Tibshirani, 1993). It is usually reserved for problems that are not readily solved by traditional statistical theory. The success of the bootstrap depends on the recognition that the actual sample may be used as an empirical estimate of the true distribution of the underlying random variables. From this empirical distribution, multiple samples may be drawn and the behavior of the statistics of interest

observed in these samples. A crucial, but nontrivial, component of the bootstrap is the use of a proper sampling scheme from the original observations. For the mortality data used in this article, there are at least three different ways of sampling: sampling individuals, sampling residuals, and sampling the stochastic process. Because preliminary work suggested that only the last of these methods worked well, we discuss it in more detail below.

Sampling the stochastic process. This process involves using the estimated values of A and α from the original data to estimate the probability of surviving to each census point. For the data analyzed here, populations are checked at daily intervals for deaths. The chance of an individual dying between day x and day $x + 1$ is given by,

$$\hat{u}_x = 1 - \left(\frac{\hat{p}_{x+1}}{\hat{p}_x} \right), \quad x = 1, 2, \dots, t_{N_0}$$

and the “hat” indicates that estimates of A and α have been substituted in eq. (3). The original N_0 individuals are then followed through time. On day 1, a random number with a uniform distribution on (0,1) is chosen for each of the N_0 individuals. If the number is less than u_1 , that individual dies; otherwise, it lives. The process is repeated for the survivors at day 2, 3, . . . until N_0 deaths are recorded. The whole process of generating N_0 deaths is repeated m times. The behavior of these m replicate data sets can then be used to make statistical inferences such as confidence intervals.

Confidence intervals. The “BC_a” method (Efron and Tibshirani, 1993) is used for generating confidence intervals. With this method, the m values of the statistic of interest are ordered from smallest to largest. The lower $100\alpha_1$ th percentile of these order statistics is used for the lower confidence limit and the $100\alpha_2$ th percentile for the upper limit. Where,

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right), \quad \alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})} \right)$$

$\Phi(\cdot)$ is the standard-normal cumulative distribution function, and $z^{(\alpha)}$ is the 100α th percentile of a standard normal distribution. If χ is the fraction of all bootstrap estimates of the statistic that are less than the original estimate, then,

$$\hat{z}_0 = \Phi^{-1}(\chi).$$

Efron and Tibshirani call \hat{a} the acceleration parameter, which is estimated from the jackknife. Under the assumption that the statistic of interest has a normal distribution, the variance of that statistic will be the same no matter what the true value is. In practice, the variance may, in fact, change for different population values of the statistic and require correction. The acceleration parameter helps to achieve this correction.

The acceleration parameter is computed from the following relationship,

$$\hat{\alpha} = \frac{\sum (\bar{s} - s_i)^3}{6[\sum (\bar{s} - s_i)^2]^{3/2}},$$

where s_i is the i th pseudo-value computed by deleting the time of death for the i th individual from \mathbf{T} and finding the maximum likelihood estimates of A and α from eq. (7). Because there are N_0 points in \mathbf{T} , there are N_0 pseudo-values whose mean is \bar{s} .

Simulated data sets

To test the accuracy of the estimates provided by eqs. (10) and (11), as well as the bootstrap confidence intervals, computer simulations were conducted. The first step in these simulations was to generate samples from a simulated population with Gompertz mortality rates. These samples consisted of the recorded day of death for each of N individuals (where $N = 50, 100, \text{ or } 150$). We assumed that the population was checked at daily intervals for new deaths. For a fixed combination of A and α , we computed the probability of surviving to time t as p_t , from (3). Suppose there are N_t surviving adults at this time. The chance of surviving to time $t + 1$ is p_{t+1} . Thus, the probability of dying, u_t , between day t and day $t + 1$ is $1 - (p_{t+1}/p_t)$. For each of the N_t survivors, a uniform random number from (0,1) was chosen. If the number was less than u_t , then the individual died on day $t + 1$; otherwise, the individual lived. This process was repeated until all N adults had died. For each of the three sample sizes, 1000 artificial data sets were created by following the sampling scheme just described.

With each of the 1000 data sets, the techniques described in the previous section were used to estimate A and α . From these estimates of A and α , the mean longevity, $t_{0.5}$, $t_{0.7}$, $t_{0.9}$, and $t_{0.95}$ were calculated.

In general, we would like to minimize the variance and bias of estimators. We are also interested in investigating the ability of the Gompertz model to predict these statistical quantities, when only a fraction of the data set has been used. As a means of studying this problem, only portions of the artificial data sets were used to carry out the estimation process. In particular, we utilized the first 20, 35, and 50% of each of the data sets to estimate A and α .

Empirical data sets

To illustrate the methods described here and determine if there are differences in the ability of each demographic model to describe real data, we have applied these methods to 20 cohorts of *Drosophila melanogaster*. These populations have been subjected to age-specific selection (Rose, 1984) and are of two treatment types: O, selected for late life reproduction; and B, selected for early life reproduction. There are five replicates of each type (e.g., B_1, B_2, \dots, B_5) and data has been collected separately for males and females. Previously we have shown that the Bs and Os differ in demographic parameters, including their "rates of aging," α (Nusbaum *et al.*, submitted).

Numerical methods

Numerical integration of eq. (10) was performed by a modified Romberg integration procedure, using the Pascal procedure QROMO described in Press *et al.* (1986). The upper limit to this integral must be set to reasonable limits; otherwise, calls to the

Gompertz density functions may result in numerical overflows. One convenient way to estimate a reasonable upper limit is to use eq. (11) and set β to 0.9999. In general, whatever limit is chosen must also be used in any bootstrap procedure to avoid systematic bias. The maximum of the likelihood function was found numerically with a downhill simplex method (Press *et al.*, 1986). This method does not require finding the derivatives of eq. (7), which would be an onerous task for the full Gompertz equation.

Random numbers were calculated with the RAN3 function (Press *et al.*, 1986). This function uses two calls to the random number generator for each number produced. The first call picks a random cell in an array of random numbers, while the second call replaces the random number chosen by the first call. Although slower than a procedure that simply generates a single number per call, RAN3 effectively increases its cycle time by the square of the cycle time of the source generator. Of course, the longer it takes the random number generator to cycle, the less likely it is that cycles will induce correlations in the observations. The normal distribution function was approximated by the polynomial expansion in Johnson and Kotz (1970).

RESULTS

The instantaneous death rate approximation

As mentioned earlier, a number of studies (Tatar *et al.*, 1993; Brooks *et al.*, 1994; Tatar and Carey 1994) utilize the approximation in eq. (5). Using the notation developed in this article, it follows that,

$$P_t = \frac{N_0 p_{t+1}}{N_0 p_t} = \exp\left\{\frac{A}{\alpha} [(1 - \exp(\alpha(t + 1))) - (1 - \exp(\alpha t))]\right\}.$$

After some algebra, one can show that,

$$-\ln(P_t) = \frac{A}{\alpha} [\exp(\alpha t)(\exp(\alpha) - 1)]. \quad (12)$$

The right side of eq. (12) will equal the Gompertz equation (1) only if we use the approximation that $e^\alpha \cong 1 + \alpha$.

More generally the probability of surviving over a single time interval is sometimes estimated from the geometric average of observations over many time intervals, e.g., $\bar{P} = (P_t P_{t+1} \dots P_{t+k-1})^{1/k}$. In this case, it can be shown that the approximation $u(t) = -\ln(\bar{P})$ depends on the accuracy of the approximation $e^{k\alpha} \cong 1 + k\alpha$.

Clearly, the accuracy of these approximations will depend on the time interval used and the value of α .

For instance, many of the *Drosophila* populations examined in this article have an α close to 0.1. In this case, when eq. (5) is used on daily observations of mortality, the difference between the approximation and the true value is only 0.5% of the true value. However, if eq. (5) had been used on weekly data, the difference between the true value and the approximation would be 16% of the true value. The reason this problem arises in the first place is that the Gompertz model assumes that mortality rates change

continuously with time, while methods for estimating mortality from survival data can only do so over finite time intervals. The larger the time interval, the worse the approximation. Because there are essentially no problems that can't be solved with the full Gompertz model without this approximation from eq. (5), we see little value in using it given the potentially large errors that can result.

Methods of estimation

Four different methods were described above for estimating the two parameters of the Gompertz equation. Here, we examine the ability of each method to provide accurate estimates of A and α . To do this, we created 1000 computer-generated data sets from the Gompertz equation with $A = 0.00113$ and $\alpha = 0.0928$. These values were chosen to be close to values we have observed in our *Drosophila* populations. For each of these data sets, the values of A and α were estimated from each method. For each technique we then determined the bias, variance, and mean square error (= variance + bias²) (Bickel and Doksum, 1977).

The results (Tables 1–2) show that in general the maximum likelihood method was best, and the linear regression technique the worst. Thus, the maximum likelihood method without approximation had smaller variance and bias (and, thus, smaller mean squared error) relative to all other methods. Using eq. (5) with maximum likelihood, even when the time intervals were only one day, increased both the variance and bias of the estimator relative to the maximum likelihood without this approximation.

TABLE 1. THE STATISTICAL PROPERTIES OF SEVERAL METHODS FOR ESTIMATING THE PARAMETER A OF THE GOMPERTZ EQUATION AT THREE DIFFERENT SAMPLE SIZES

Method	Sample size	Percent bias	Variance ($\times 10^9$)	Mean square error ($\times 10^9$)
LR	50	991	28500	153000
	100	885	14800	116000
	150	843	11600	102000
NR	50	42	1010	1240
	100	31	460	584
	150	21	296	355
MLA	50	2.9	3.3	4.3
	100	2.5	1.5	2.2
	150	2.4	1.0	1.8
ML	50	1.4	3.1	3.3
	100	1.0	1.4	1.5
	150	0.9	1.0	1.1

Methods studied include: linear regression (LR), nonlinear regression (NR), maximum likelihood utilizing an approximation to the Gompertz function (MLA), maximum likelihood utilizing the exact Gompertz (ML). Mortality rates were measured at weekly intervals in the linear regression analysis, while daily intervals were used for the other three methods.

TABLE 2. THE STATISTICAL PROPERTIES OF SEVERAL METHODS FOR ESTIMATING THE PARAMETER α OF THE GOMPERTZ EQUATION AT THREE DIFFERENT SAMPLE SIZES

<i>Method</i>	<i>Sample size</i>	<i>Percent bias</i>	<i>Variance</i> ($\times 10^6$)	<i>Mean square</i> <i>error</i> ($\times 10^6$)
LR	50	19	96	419
	100	19	60	357
	150	18	47	322
NR	50	5.9	251	282
	100	5.9	126	155
	150	4.4	93	109
MLA	50	2.9	22	29
	100	2.5	9.9	15
	150	2.4	6.9	12
ML	50	1.4	21	22
	100	1.0	9.4	10
	150	0.9	6.5	7.3

Methods studied include: linear regression (LR), nonlinear regression (NR), maximum likelihood utilizing an approximation to the Gompertz function (MLA), maximum likelihood utilizing the exact Gompertz (ML).

Bootstrap confidence intervals

Utilizing the data sets described in the previous section we have tested the accuracy of 95% confidence intervals produced from the bootstrap and maximum likelihood techniques. For each sample, a bootstrap confidence interval was created for mean longevity, t_{50} , t_{70} , t_{90} , and t_{95} . We then noted if the interval included the true value or not. The results of these simulations (Table 3) show that the confidence levels of these intervals are very close to the nominal value of 95%.

Simulation results

We also simulated a population with A and α values similar to those in B populations. With the simulated data sets, we estimated the statistics of interest with only the first 20, 35, and 50% of the data. The effects of using different fractions of these data sets and

TABLE 3. THE EMPIRICAL CONFIDENCE LEVEL FOR THE 95% BOOTSTRAP CONFIDENCE INTERVALS GENERATED BY THE STOCHASTIC SAMPLING TECHNIQUE

<i>Longevity</i>	t_{50}	t_{70}	t_{90}	t_{95}
96.5	96.5	96.6	96.8	96.5

The sample size was 150 and only 50% of the data were used. Preliminary work with jackknife confidence intervals gave very poor results with this example; hence, it has been used as the test case for the bootstrap.

different sample sizes of individuals are shown in Fig. 1 for the variance in mean longevity, $t_{0.5}$, $t_{0.7}$, $t_{0.9}$, and $t_{0.95}$.

The variance of these estimators clearly decreases with increasing sample size and increasing fractions of the data (Fig. 1). But the shape of these curves is, in fact, very similar for each statistic, suggesting a common numerical response to the size of the data sets used. Empirically, we can estimate from Fig. 1 that the variance of these estimators is decreasing in proportion to N , if N is the total sample size.

The effects of the fraction of the data used are more complicated. For instance,

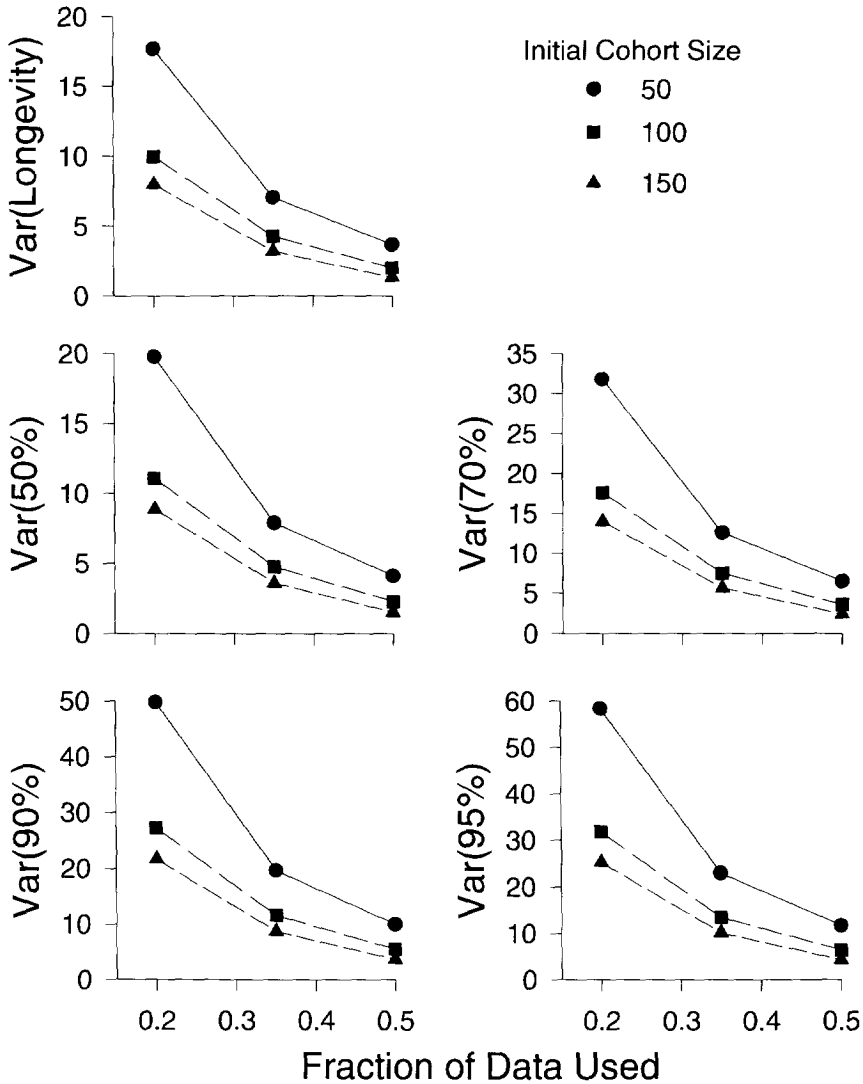


FIG. 1. The variance of several demographic statistics as a function of (i) the initial size of the cohort and (ii) the fraction of all deaths that were observed prior to estimation of the Gompertz parameters.

inspection of Fig. 1 shows that the drop in variance in mean longevity from 20 to 35% is about the same for each sample size (50, 100, 150). In addition, the decrease in variance from 35 to 50% of the data is about the same as the decrease seen from 20 to 35%. In general, the variance depends on both the fraction of data used and the sample size.

Application to B and O data

The Gompertz equation was used with mortality data (Figs. 2–3) from the B and O populations to predict the mean longevity and $t_{0.95}$ (Tables 4–5). In addition, bootstrap confidence intervals were calculated utilizing the stochastic sampling technique. For each population in Tables 4–5, these estimates are made with various fractions of the initial data set: 100, 50, 35, or 20%.

Examination of the size of the confidence intervals in Tables 4–5 shows that often there are only small increases in interval size with decreasing fractions of the data used. This is especially so when moving from 100 to 50% or 50 to 35%. Although we do not know the true mean longevity of these populations, we can use the observations to estimate the mean and a confidence interval. The confidence interval on these observed longevities almost always overlaps the confidence interval of the predicted longevities.

We have also enumerated the number of deaths that exceed $t_{0.95}$ (Table 6). Because we have different estimates of A and α for each fraction of the data used, we also get different values of $t_{0.95}$ in each case. Accordingly, we have counted the number of deaths that exceed each of the different $t_{0.95}$. Because the sample sizes were small, we pooled all counts from like sex and population. In 15 out of 20 comparisons the observed number of deaths in this tail of the distribution is within the expected 95% confidence interval (these are exact confidence intervals computed from the binomial distribution). To evaluate these results, it is important to remember that the value of $t_{0.95}$ is, itself, an estimate. Had we looked at the performance of the Gompertz using the range provided by the 95% confidence intervals (Tables 4–5) on $t_{0.95}$, the level of agreement would have been much better.

These results suggest that considerable savings in time and energy may be achieved by computing the expected size of these confidence intervals ahead of time for various combinations of N and fraction of data used. This can be done with the techniques outlined here if one can make a reasonable guess at A or α . If this cannot be done, a range of A and α values might be examined. Estimates of A and α can be obtained if one has estimates of two values of t_{β} . If the values of β used are not very close (e.g., 50 and 90%), then estimates may be obtained by numerically solving the systems of nonlinear equations that result from eq. (11), using techniques like the method of bisection or Newton's method (Philips and Taylor, 1973). This technique is also useful for estimating the starting values of A and α when doing maximum likelihood.

DISCUSSION

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. The estimates obtained by linear regression depend on unstable features of the data set, which make them useless for comparison across populations. These problems can be easily avoided by using the maximum likelihood techniques described here. The max-

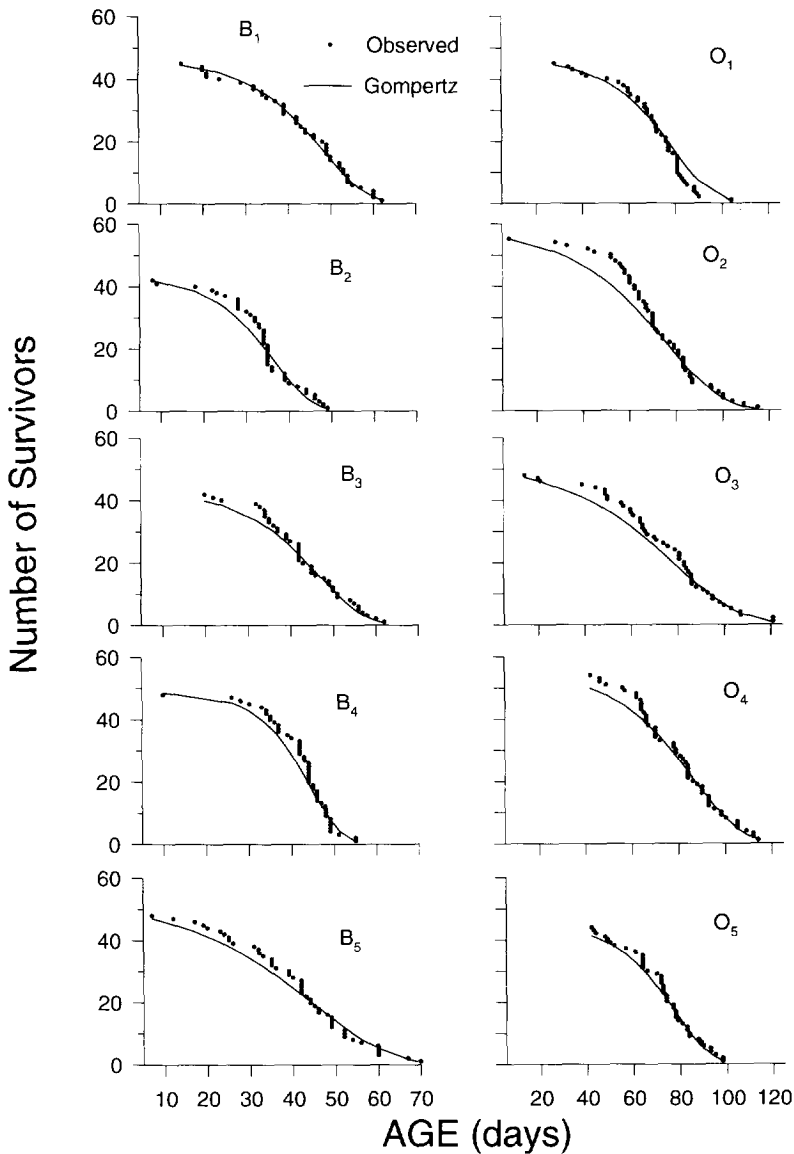


FIG. 2. The survival of females in the B and O populations.

imum likelihood method, however, assumes sampling is binomial. This is a reasonable assumption for the laboratory populations studied here, but may not be justified in other circumstances. Clearly, appropriateness of this assumption must be evaluated on a case-by-case basis.

For populations in which the Gompertz equation appears to work well, estimates of mean longevity, $t_{0.95}$, $t_{0.99}$, etc., can be made by observing the first 20% or so of all deaths. Confidence intervals on these estimates may be made as small as desired by simply increasing the initial cohort size. The design of experiments can be aided by first

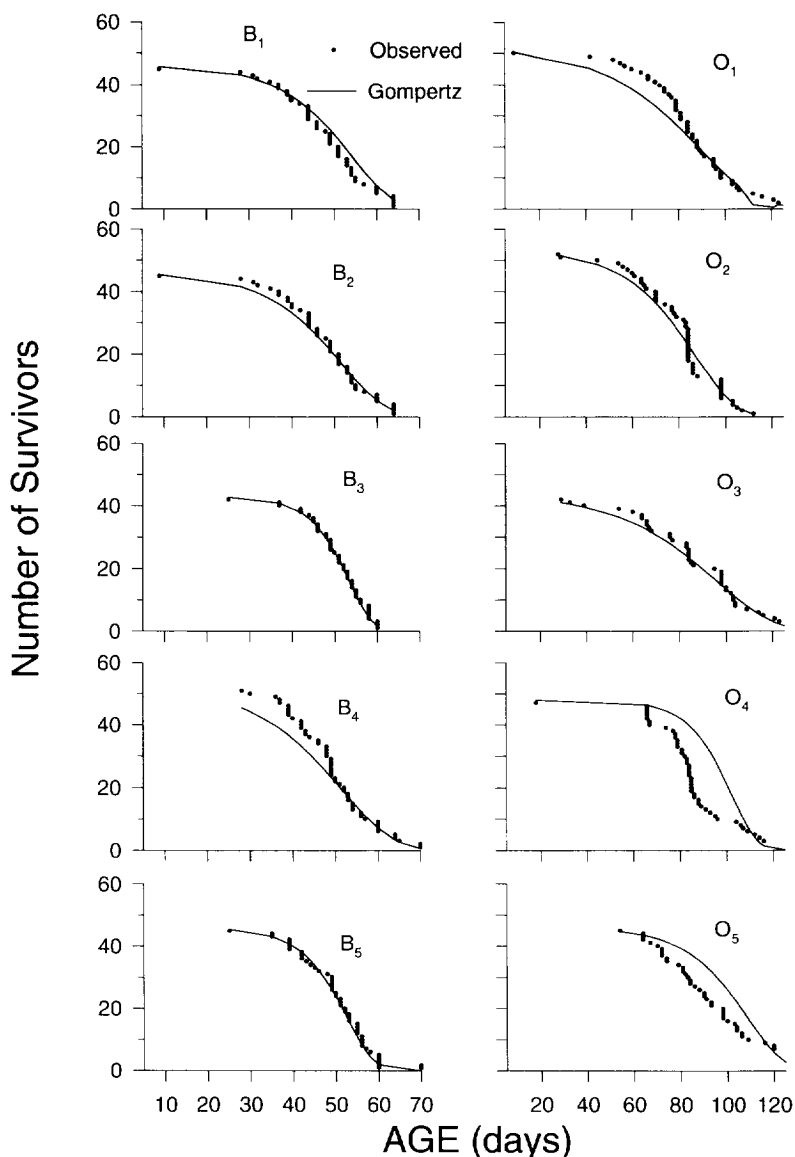


FIG. 3. The survival of males in the B and O populations.

making a guess at A and α for the population of interest. Following the bootstrap process outlined here, estimates of the size of confidence intervals for any of the important statistics that follow from the Gompertz equation can be made. The effects of initial cohort size and fraction of the data examined can be easily evaluated with these techniques to ascertain the most efficient manner to collect mortality data.

For populations that do not obey Gompertz mortality kinetics, the techniques in this article can be easily modified. For instance, Carey *et al.* (1993) suggest that medflies violate crucial assumption of the Gompertz equation. Hirsh *et al.* (1944) describe sev-

TABLE 4. THE MEAN LONGEVITY AND $t_{0.95}$ FOR B AND O FEMALES (WITH BOOTSTRAP 95% CONFIDENCE INTERVALS CALCULATED WITH THE STOCHASTIC SAMPLING TECHNIQUE)

<i>Population</i>	<i>Initial cohort size</i>	<i>Percent data</i>	<i>Observed longevity ($\pm 95\%$ C.I.)</i>	<i>Mean longevity</i>	$t_{0.95}$
B ₁	46	100	43 \pm 4	43 (41,44)	60 (57,62)
		50		43 (39,46)	60 (53,66)
		35		43 (37,48)	60 (50,68)
		20		42 (36,52)	59 (46,75)
B ₂	43	100	34 \pm 3	32 (31,33)	46 (44,48)
		50		33 (31,37)	49 (43,57)
		35		36 (33,43)	55 (45,69)
		20		38 (34,53)	58 (47,91)
B ₃	43	100	44 \pm 3	41 (40,43)	59 (56,61)
		50		40 (36,43)	56 (49,61)
		35		41 (37,48)	59 (51,71)
		20		44 (38,57)	64 (50,88)
B ₄	49	100	42 \pm 4	40 (39,42)	53 (51,55)
		50		43 (40,47)	56 (52,62)
		35		42 (38,46)	55 (49,60)
		20		44 (38,50)	58 (49,67)
B ₅	49	100	42 \pm 4	39 (38,42)	65 (62,74)
		50		39 (37,45)	64 (56,90)
		35		39 (35,45)	56 (47,67)
		20		38 (33,46)	56 (44,72)
O ₁	46	100	71 \pm 5	71 (68,74)	98 (93,102)
		50		72 (66,77)	96 (87,104)
		35		73 (64,81)	96 (83,109)
		20		78 (68,102)	114 (91,161)
O ₂	56	100	72 \pm 6	66 (64,70)	102 (97,112)
		50		69 (64,74)	92 (85,99)
		35		64 (58,77)	99 (83,127)
		20		85 (46,143)	148 (83,447)
O ₃	49	100	74 \pm 7	69 (66,72)	110 (104,119)
		50		70 (66,82)	113 (97,148)
		35		71 (63,80)	98 (85,113)
		20		71 (61,85)	98 (81,121)
O ₄	55	100	80 \pm 5	76 (74,79)	109 (104,114)
		50		77 (71,85)	110 (99,125)
		35		73 (65,83)	103 (89,120)
		20		79 (68,98)	115 (91,150)
O ₅	45	100	73 \pm 5	70 (67,73)	95 (90,99)
		50		71 (65,76)	96 (86,105)
		35		74 (67,87)	102 (90,124)
		20		76 (65,95)	105 (87,136)

eral two parameter models that appear to provide much better descriptions of medfly mortality kinetics. All the techniques described here can be used easily with these other models. In fact, one simply has to replace two lines of computer code in all the programs used in this article to compare equivalent results for the JMA and gamma models described by Hirsh *et al.* (these programs are available from the authors upon request).

TABLE 5. THE MEAN LONGEVITY AND $t_{0.95}$ FOR B AND O MALES (WITH BOOTSTRAP 95% CONFIDENCE INTERVALS CALCULATED WITH THE STOCHASTIC SAMPLING TECHNIQUE)

<i>Population</i>	<i>Initial cohort size</i>	<i>Percent data</i>	<i>Observed longevity</i> ($\pm 95\%$ C.I.)	<i>Mean longevity</i>	$t_{0.95}$
B ₁	51	100	50 \pm 3	49 (47,51)	65 (62,68)
		50		49 (45,52)	65 (59,71)
		35		47 (41,50)	61 (54,67)
		20		47 (40,54)	63 (51,73)
B ₂	46	100	48 \pm 3	46 (44,48)	63 (61,66)
		50		47 (44,53)	65 (60,75)
		35		47 (43,55)	65 (58,79)
		20		50 (42,63)	70 (55,93)
B ₃	43	100	51 \pm 2	50 (49,52)	60 (58,61)
		50		51 (49,54)	61 (58,64)
		35		51 (48,54)	61 (56,64)
		20		52 (47,56)	62 (55,67)
B ₄	52	100	50 \pm 3	45 (44,47)	65 (62,68)
		50		46 (42,51)	67 (59,75)
		35		48 (43,55)	71 (57,83)
		20		48 (42,69)	70 (56,111)
B ₅	46	100	50 \pm 2	49 (47,51)	60 (58,62)
		50		50 (48,53)	61 (58,65)
		35		52 (48,55)	63 (58,68)
		20		49 (43,54)	59 (52,66)
O ₁	51	100	85 \pm 6	78 (75,80)	117 (110,121)
		50		78 (72,87)	116 (103,136)
		35		83 (76,102)	128 (109,170)
		20		97 (91,189)	157 (132,366)
O ₂	53	100	80 \pm 5	78 (75,81)	106 (102,111)
		50		81 (75,88)	112 (101,124)
		35		82 (74,92)	113 (101,131)
		20		80 (69,97)	110 (91,139)
O ₃	43	100	88 \pm 8	84 (81,86)	123 (117,128)
		50		80 (73,89)	117 (103,133)
		35		90 (81,109)	135 (114,175)
		20		82 (71,112)	119 (96,177)
O ₄	48	100	87 \pm 6	96 (94,99)	114 (111,118)
		50		83 (78,86)	98 (93,102)
		35		84 (80,89)	100 (95,105)
		20		87 (81,94)	103 (95,112)
O ₅	46	100	94 \pm 6	100 (97,105)	126 (122,133)
		50		88 (84,95)	110 (104,119)
		35		87 (81,96)	108 (100,119)
		20		83 (74,93)	104 (92,116)

Not all populations will be expected to obey the Gompertz equation. Nor do we expect the Gompertz to work well in the very old age classes, in which natural selection has not been an effective force at molding rates of mortality (Nusbaum, Mueller, and Rose, submitted). Nevertheless, for populations that are relatively long-lived and follow Gompertz mortality curves, or some other simple mortality kinetic model, sub-

TABLE 6. THE FREQUENCY OF B AND O DEATHS THAT EXCEED $t_{0.95}$ COMPUTED FROM EQ. (11)

Population	Sex	Total number of observed deaths	95% C.I. on the observed number of deaths $> t_{0.95}$	Percent of the data used to estimate $t_{0.95}$	Observed number of deaths $> t_{0.95}$
B	M	238	(5,19)	100	11
				50	8
				35	9
				20	13
				100	15
B	F	230	(5,18)	50	10
				35	13
				20	12
				100	19
				50	30
O	M	241	(5,19)	35	23
				20	30
				100	16
				50	21
				35	23
O	F	251	(6,19)	20	9

Although all populations have been pooled the particular $t_{0.95}$ from each population was used in these calculations.

stantial savings in estimating longevity could be achieved by applying the methods presented here.

Acknowledgments—We thank J. Curtsinger, A. Joshi, and an anonymous referee for comments on the manuscript, and NIH Grant AG09970 for support.

REFERENCES

- BICKEL, P.J. and DOKSUM, K.A. *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.
- BROOKS, A., LITHGOW, G.J., and JOHNSON, T.E. Mortality rates in a genetically heterogeneous population of *Caenorhabditis elegans*. *Science* **263**, 668–671, 1994.
- CAREY, J.R., LIEDO, P., OROZCO, D., and VAUPEL, J.W. Slowing of mortality rates at older ages in large medfly cohorts. *Science* **258**, 457–461, 1992.
- CHARLESWORTH, B. *Evolution in Age-Structured Populations*, Cambridge University Press, London, 1980.
- CURTSINGER, J.W., FUKUI, H.H., TOWNSEND, D.R., and VAUPEL, J.W. Demography of genotypes: Failure of the limited life-span paradigm in *Drosophila melanogaster*. *Science* **258**, 461–463, 1992.
- EFRON, B. and TIBSHIRANI, R.J. *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.
- FINCH, C.E. *Longevity, Senescence and the Genome*, University of Chicago Press, Chicago, 1990.
- FUKUI, H.H., XIU, L., and CURTSINGER, J.W. Slowing of age-specific mortality rates in *Drosophila melanogaster*. *Exp. Gerontol.* **28**, 585–599, 1993.
- GAVRILOV, L.A. and GAVRILOVA, N.S. *The Biology of Life Span: A Quantitative Approach*, Harwood Academic Publishers, Chur, Switzerland, 1991.
- GOMPERTZ, B. On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies. *Philos. Trans. R. Soc. Lond. Ser. A* **115**, 513–585, 1825.
- HIRSCH, A.G., WILLIAMS, R.J., and MEHL, P. Kinetics of medfly mortality. *Exp. Gerontol.* **29**, 197–204, 1994.

- HUGHES, K.A. and CHARLESWORTH, B. A genetic analysis of senescence in *Drosophila*. *Nature* **367**, 64–66, 1994.
- JOHNSON, N.L. and KOTZ, S. *Continuous Univariate Distributions*. 1, Wiley, New York, 1970.
- MARQUARDT, D.W. An algorithm for least-squares estimation of nonlinear parameters. *Soc. Indust. Appl. Math.* **11**, 431–441, 1963.
- PHILLIPS, G.M. and TAYLOR, P.J. *Theory and Applications of Numerical Analysis*, Academic Press, London, 1973.
- PRESS, W.H., FLANNERY, B.P., TEUKOLSKY, S.A., and VETTERLING, W.T. *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.
- PROMISLOW, D.E.L. Senescence in natural populations of mammals: A comparative study. *Evolution* **45**, 1869–1877, 1991.
- ROSE, M.R. Laboratory evolution of postponed senescence in *Drosophila*. *Evolution* **38**, 1004–1010, 1984.
- ROSE, M.R. *Evolutionary Biology of Aging*, Oxford University Press, New York, 1991.
- TATAR, M. and CAREY, J.R. Genetics of mortality in the bean beetle *Callosobruchus maculatus*. *Evolution* **48**, 1371–1376, 1994.
- TATAR, M., CAREY, J.R., and VAUPEL, J.W. Longterm cost of reproduction with and without accelerated senescence in *Callosobruchus maculatus*: Analysis of age-specific mortality. *Evolution* **47**, 1302–1312, 1993.